

# 1 **Cross-ancestry analysis identifies genes associated with obesity risk** 2 **and protection**

3

4 Deepro Banerjee<sup>1,2</sup> and Santhosh Girirajan<sup>1,2,3</sup>

5

6 1. Bioinformatics and Genomics Graduate Program, The Huck Institute of the Life

7 Sciences, University Park, PA 16802

8 2. Department of Biochemistry and Molecular Biology, Pennsylvania State University,

9 University Park, PA 16802

10 3. Department of Anthropology, Pennsylvania State University, University Park, PA 16802

11

12

13

14

15

16

17 Correspondence:

18 Santhosh Girirajan

19 205A Life Sciences Building

20 Pennsylvania State University

21 University Park, PA 16802

22 E-mail: [sxg47@psu.edu](mailto:sxg47@psu.edu)

23 **ABSTRACT**

24 Gene discoveries in obesity have largely been based on European cohorts, leading to an ancestral  
25 bias, that limits their generalizability across populations. We performed a gene-based rare variant  
26 association study of 721,941 individuals and identified 116 novel BMI-associated genes with  
27 consistent effects across ancestries, including 50 risk-conferring and 66 protective genes against  
28 obesity. Protective genes such as *DCUNID3* and *NEUROD6* had effect sizes comparable to  
29 high-risk genes such as *MC4R* and *BSN*, and nearly twice that of known protective genes such as  
30 *GPR75*, which, along with five other genes, showed strong European bias. Notably, 82 of the  
31 116 genes showed functional relevance to obesity including adiposity, energy homeostasis, and  
32 glucose metabolism. While polygenic risks or an obesogenic lifestyle amplified the effect of 15  
33 genes on BMI, including the combination of low physical activity and *MACROD1*, 23 genes  
34 including *VIRMA*, *AQP3*, and *PML* retained protective effects even at high polygenic scores. Our  
35 findings provide further insights into the genetic basis of obesity that is conserved across  
36 ancestries and their interactions with obesogenic factors.

37

38

39 **MAIN**

40 Obesity is a complex, heritable disorder with significant global impact, contributing to numerous  
41 comorbidities and public health challenges<sup>1</sup>. It is influenced by a combination of genetic and  
42 lifestyle factors<sup>2,3</sup>, yet the genetic component underlying the etiology of obesity remains a key  
43 area of investigation. While large-scale studies have identified roles for common variants<sup>4</sup>,  
44 including significant effects of polygenic risk scores (PGS)<sup>5</sup>, and rare protein-truncating variants  
45 in obesity risk genes such as *MC4R* and *BSN*<sup>6-9</sup>, as well as protective genes such as *GPR75*<sup>9</sup>,  
46 these studies have predominantly focused on populations of European ancestry. Non-European  
47 populations are often underrepresented or included only for replication purposes<sup>6-8</sup>. This lack of  
48 diversity in genetic studies has resulted in two major issues: a bias in gene discovery towards a  
49 specific ancestry, and a limited ability to detect associations that may be significant in non-  
50 Europeans but lack statistical power in predominantly European cohorts<sup>10</sup>. *For example*,  
51 clinically significant variants in *APOLI1* and *PCSK9* associated with kidney disease and low LDL  
52 cholesterol, respectively, were both discovered in populations of African ancestry<sup>11,12</sup>. The  
53 consequences of this bias extend beyond gene discovery, affecting the generalizability of  
54 obesity-related findings to the broader global population and potentially limiting the  
55 effectiveness of precision medicine approaches<sup>13</sup>.

56 Here, we conducted a rare variant association analysis of body-mass index (BMI) using  
57 genetic and phenotypic data from 721,941 adults of diverse ancestries, leveraging cohorts from  
58 the UK Biobank (UKB) and the All of Us (AoU) initiative. Unlike previous rare variant  
59 association studies of BMI that prioritized European populations for discovery and non-  
60 Europeans for replication, we adopted a discovery approach in non-European populations as  
61 well. Combining statistics across populations, we identified 121 genes, 116 of which have not  
62 been previously linked to BMI in the context of rare variants, that showed consistent risk-  
63 conferring and protective effects across ancestries. Our findings also revealed a strong European  
64 bias in previously identified obesity genes, with effect sizes varying significantly in non-  
65 European populations. These disparities highlight the critical need for more inclusive genomic  
66 studies to ensure equitable obesity interventions.

67

## 68 RESULTS

### 69 Cross-ancestry analysis to identify BMI-associated genes

70 We analyzed genetic and electronic health record data of 721,941 adults from the UKB and AoU  
71 cohorts representing six ancestries<sup>14,15</sup>. We divided the cohorts into five populations, each with at  
72 least 50,000 individuals (**Table 1**). We used REGENIE v3.3<sup>16</sup> to perform gene-burden  
73 association tests by collapsing ultra-rare (minor allele frequency <0.1%) protein truncating  
74 variants (PTVs, defined as predicted loss of function or deleterious missense variants) for each  
75 gene and measuring their effect on BMI for each independent population. Using Bonferroni  
76 multiple testing correction ( $P < 8.34 \times 10^{-7}$ , 20,000 genes, and three variant collapsing models, see  
77 **Methods**) and sample size ( $N \geq 20$ ) thresholds, we identified a total of 11 BMI-associated genes  
78 from the UKB British (*MC4R*, *PCSK1*, *DIDO1*, *BSN*, *UBR2*, *ATP5PO*, *APBA1*, *SLC12A5*, and  
79 *ATP13A1*) and AoU European (*MC4R*, *HECTD4* and *YLPM1*) populations, while no significant  
80 genes were found in the UKB non-British, AoU African, or AoU mixed populations, potentially  
81 due to insufficient statistical power to detect any association (**Supplementary Table 1**).  
82 Therefore, to identify genetic associations that are consistent across populations, we performed  
83 random-effect inverse variance weighted meta-analysis<sup>17</sup> by aggregating summary statistics from  
84 each population into European (UKB British and AoU European populations), non-European  
85 (UKB non-British, AoU African, and AoU mixed populations), and combined (all five  
86 populations) results. Gene-burden associations that passed multiple testing correction in either  
87 the European or non-European meta-analysis, as well as in the combined result, were considered  
88 to be significant gene-BMI associations across ancestries. In contrast to previous cross-ancestry  
89 studies of BMI where European meta-analysis was typically used for discovery and non-  
90 European analysis for replication, we did not define a single discovery population. Instead, we  
91 applied the same discovery approach in both populations, such that discoveries in Europeans  
92 were replicated in non-Europeans, and vice versa. This strategy allowed us to detect association  
93 signals for BMI in both populations, with consistent effect sizes.

94 Using our cross-ancestry approach, we discovered 116 novel genes (50 associated with  
95 increased BMI and 66 with decreased BMI), along with five genes (*APBA1*, *BSN*, *MC4R*,  
96 *ROBO1*, and *UBR2*) that were previously identified in rare-variant studies on BMI<sup>6,7,9,18</sup> (**Fig. 1**,  
97 **Supplementary Tables 2, 3 and 4**). We found that 28 out of the 33 known BMI-associated  
98 genes did not pass the exome-wide significance threshold, including six genes (*GPR75*, *PTPRG*,

99 *SPARC*, *RAB21*, *ATP13A1*, and *DIDO1*) that were only significant in Europeans but not in the  
 100 combined meta-analysis (**Supplementary Table 5**). This result is consistent with a recent study  
 101 by Zhao and colleagues where the association of *ATP13A1* with BMI did not replicate in a non-  
 102 European cohort<sup>7</sup>. In contrast, we found that *GIPR* showed a significant effect in non-Europeans  
 103 ( $\beta=-1.12$  kg/m<sup>2</sup>, 95% CI: -1.54, -0.70,  $P=1.67 \times 10^{-7}$ ) but not in Europeans ( $\beta=-0.44$  kg/m<sup>2</sup>, 95%  
 104 CI: -0.89, 0.05,  $P=0.09$ ). We did not find significance for some genes such as *KSR2*, *ZFH3*,  
 105 and *RAPGEF3*, where previous associations were driven by specific protein-altering variants<sup>6</sup>,  
 106 while for other genes such as *ANO4*, *DPP9*, and *TOX4* cohort-specific biases may have driven  
 107 previous associations. For instance, Zhao and colleagues failed to replicate the association of  
 108 *TOX4* in a non-European cohort<sup>7</sup> (**Supplementary Fig. 1, Supplementary Table 5**).

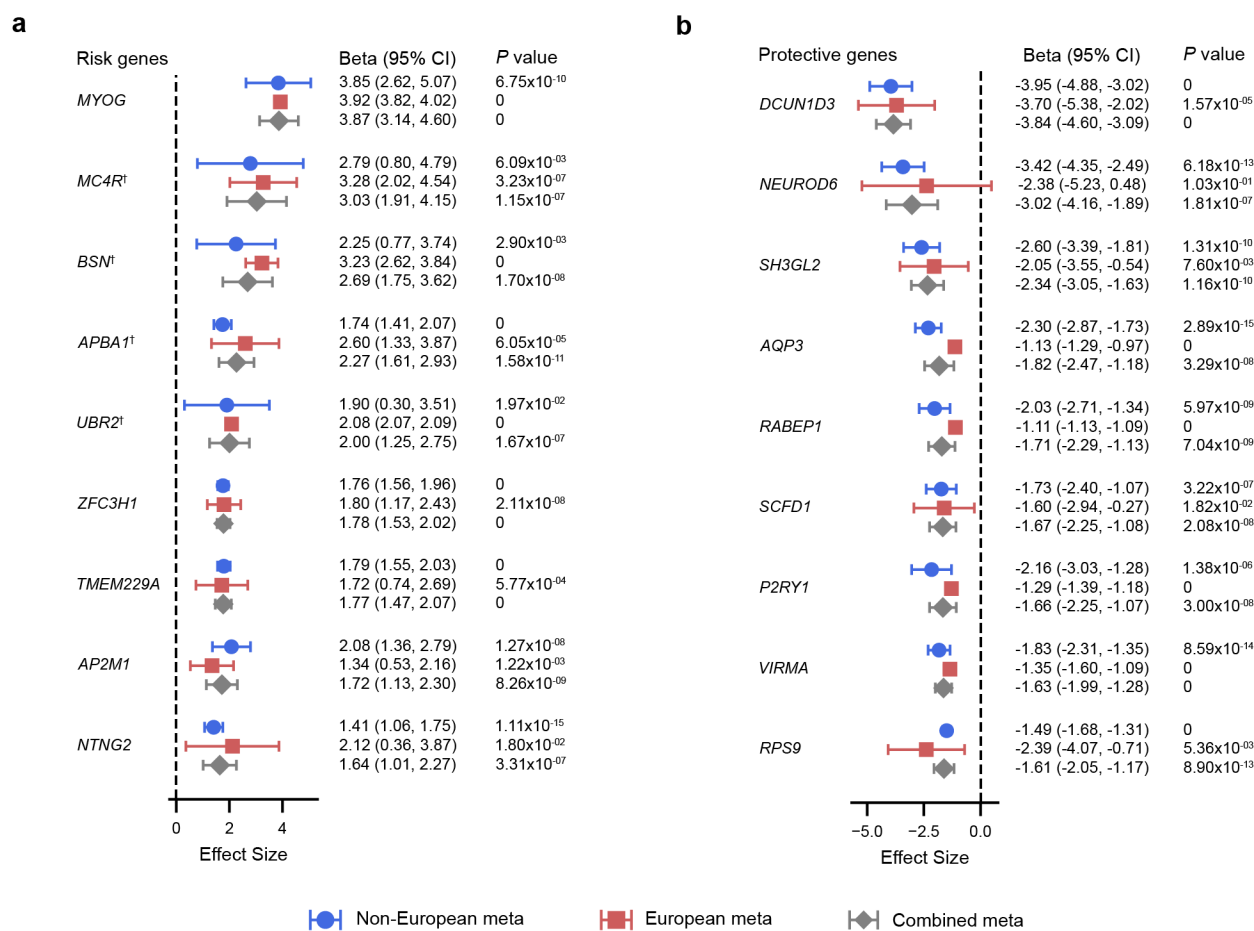
109  
 110

**Table 1:** Cohort and population statistics analyzed in this study.

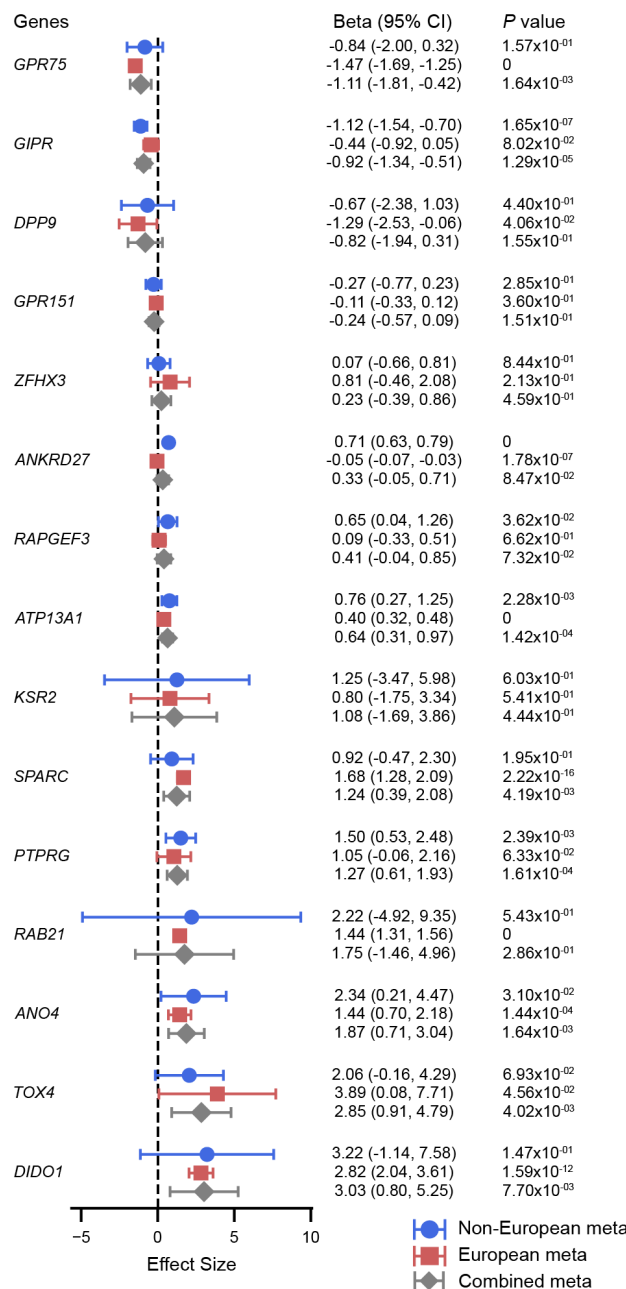
Population	European	Non-European	Combined
UKB British	419,228		419,228
UKB non-British (includes African, East Asian, South Asian ancestries)		68,070	68,070
AoU European	127,644		127,644
AoU African		54,865	54,865
AoU Mixed (includes Latino/admixed American, East Asian, South Asian and Middle Eastern ancestries)		52,134	52,134
<b>Meta population (N)</b>	<b>546,872</b>	<b>175,069</b>	<b>721,941</b>

111  
 112 In our discovered genes, *MYOG* ( $\beta=3.87$  kg/m<sup>2</sup>, 95% CI: 3.14, 4.6,  $P=0$ ), *ZFC3H1*  
 113 ( $\beta=1.78$  kg/m<sup>2</sup>, 95% CI: 1.53, 2.02,  $P=0$ ), and *TMEM229A* ( $\beta=1.77$  kg/m<sup>2</sup>, 95% CI: 1.47, 2.07,  
 114  $P=0$ ) contributed to the highest increase in BMI, with effect sizes comparable to *MC4R*, *BSN*,  
 115 and *UBR2*. In fact, we identified 40 such high-effect genes, including 24 BMI increasing and 16  
 116 BMI decreasing, which contributed to at least 1 kg/m<sup>2</sup> change in BMI (**Fig. 1, Supplementary**  
 117 **Table 4**). The effect sizes of several BMI-increasing genes, including *MACROD1*, *ICE1*, *ADNP*,  
 118 *NACC2*, and *LPCAT4*, were robust and showed minimal variability across ancestries

119 (interpopulation variance,  $V_{\beta} < 0.15$ ) (Fig. 2, Supplementary Table 6). Similarly, BMI-  
 120 decreasing genes such as *B3GNT2* ( $V_{\beta} = 0.06$ ), *SNAP29* ( $V_{\beta} = 0.10$ ), *VIRMA* ( $V_{\beta} = 0.13$ ), and *CCT7*  
 121 ( $V_{\beta} = 0.13$ ) showed more consistent effect sizes than previously reported protective genes,  
 122 including *GPR75* ( $V_{\beta} = 1.06$ ) and *GPR151* ( $V_{\beta} = 0.49$ ) (Fig. 2, Supplementary Table 6). Further,  
 123 the absolute effect sizes of BMI-decreasing genes, *DCUN1D3* ( $\beta = -3.84$  kg/m<sup>2</sup>, 95% CI: -4.6, -  
 124 3.09,  $P = 0$ ), *NEUROD6* ( $\beta = -3.02$  kg/m<sup>2</sup>, 95% CI: -4.16, -1.89,  $P = 1.8 \times 10^{-7}$ ), and *SH3GL2* ( $\beta = -$   
 125 2.34 kg/m<sup>2</sup>, 95% CI: -3.05, -1.63,  $P = 1.16 \times 10^{-10}$ ) were comparable to those of *MC4R* and *BSN*.  
 126 These results highlight the robustness of our cross-ancestry gene discovery approach.  
 127



128  
 129 **Fig. 1: Monogenic genes contributing to at least 1.5 kg/m<sup>2</sup> difference in BMI across**  
 130 **ancestries.** Effect sizes along with 95% confidence intervals (CI) and significance values ( $P$   
 131 value) in European, non-European, and combined meta-analysis for high-effect ( $> 1.5$  kg/m<sup>2</sup>)  
 132 genes associated with (a) increased and (b) decreased BMI are shown. Three variant collapsing  
 133 models were tested for each gene and the one with the lowest  $P$  value in the combined meta-  
 134 analysis is shown here. † indicates known BMI genes. Data for all genes contributing to change  
 135 in BMI (including those contributing to  $> 1$  kg/m<sup>2</sup>) are provided in **Supplementary Table 4**.

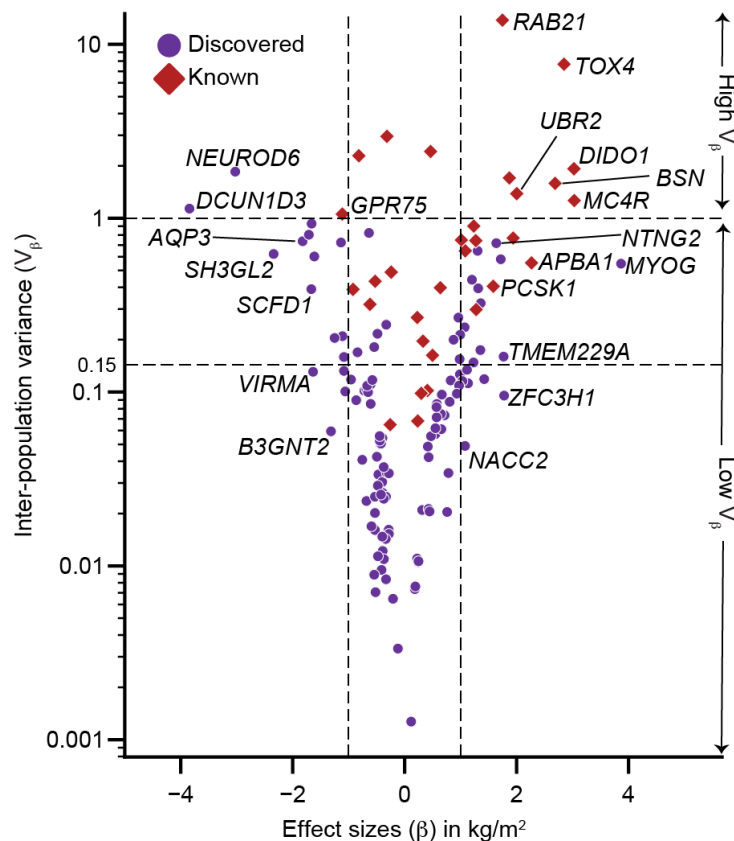


136

137 **Supplementary Fig. 1: Effect of previously identified BMI-associated genes across**  
 138 **ancestries.** Effect sizes along with 95% confidence intervals (CI) and significance values (*P*  
 139 value) in European, non-European, and combined meta-analysis for genes previously associated  
 140 with BMI are shown. Three variant collapsing models were tested for each gene and the one with  
 141 the lowest *P* value in the combined meta-analysis is shown here. Extended data are available in  
 142 **Supplementary Table 5.**  
 143

144

145



146  
147

148 **Fig. 2: Interpopulation variance of known and discovered BMI genes.** Interpopulation  
149 variance plot with effect sizes along x-axis and variance of effect sizes across populations along  
150 y-axis of previously identified (in red diamond) and discovered (in purple dots) BMI genes are  
151 shown. Three variant collapsing models were tested for each gene and the one with the lowest *P*  
152 value in the combined meta-analysis is shown here. Extended data are available in  
153 **Supplementary Table 6.**

154  
155

### 156 **Cardiometabolic profile of carriers of PTVs in BMI-associated genes**

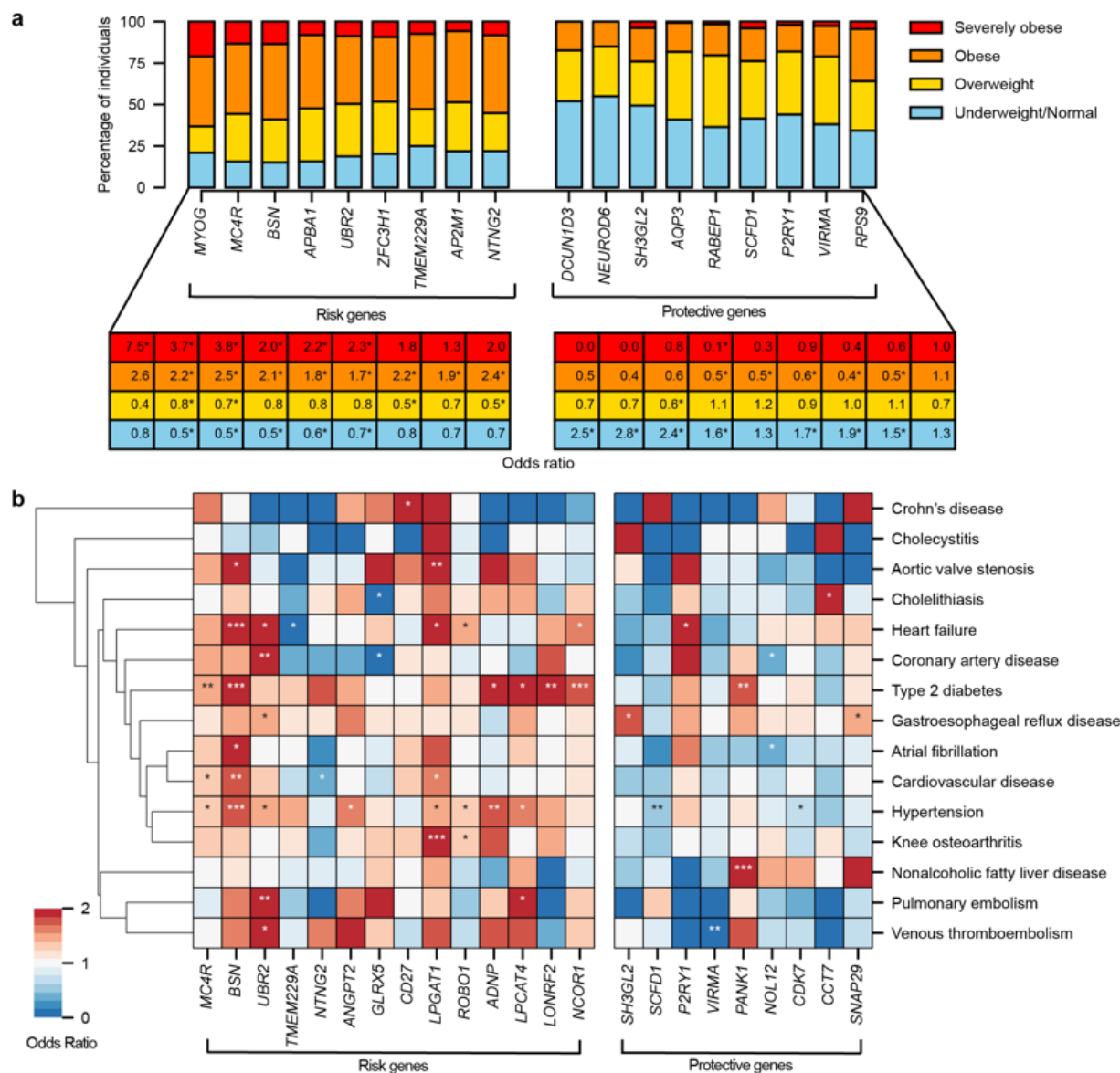
157 We next examined whether PTV carriers in the BMI-associated genes were enriched across  
158 different obesity categories, namely, underweight/normal ( $BMI < 25$ ), overweight ( $25 \leq BMI < 30$ ),  
159 obese ( $30 \leq BMI < 40$ ), and severely obese ( $BMI \geq 40$ ). Of the BMI-increasing genes, PTV carriers  
160 in 35 genes, including *NCOR1* (OR=1.28, 95% CI: 1.04, 1.58,  $P=0.02$ ), *SEC24B* (OR=1.46,  
161 95% CI: 1.04, 2.04,  $P=0.02$ ), *ADNP* (OR=1.77, 95% CI: 1.15, 2.7,  $P=0.007$ ), and *ANGPT2*  
162 (OR=1.62, 95% CI: 1.08, 2.42,  $P=0.01$ ), showed higher odds for obesity, while PTV carriers in  
163 22 genes, including *MACROD1* (OR=2.11, 95% CI: 1.31, 3.25,  $P=0.002$ ), *ICE1* (OR=1.74, 95%  
164 CI: 1.06, 2.73,  $P=0.01$ ), and *MYOG* (OR=7.47, 95% CI: 1.73, 25.21,  $P=0.004$ ), showed higher  
165 odds for severe obesity (**Fig. 3a, Supplementary Fig. 2, Supplementary Table 7**). Of the BMI-



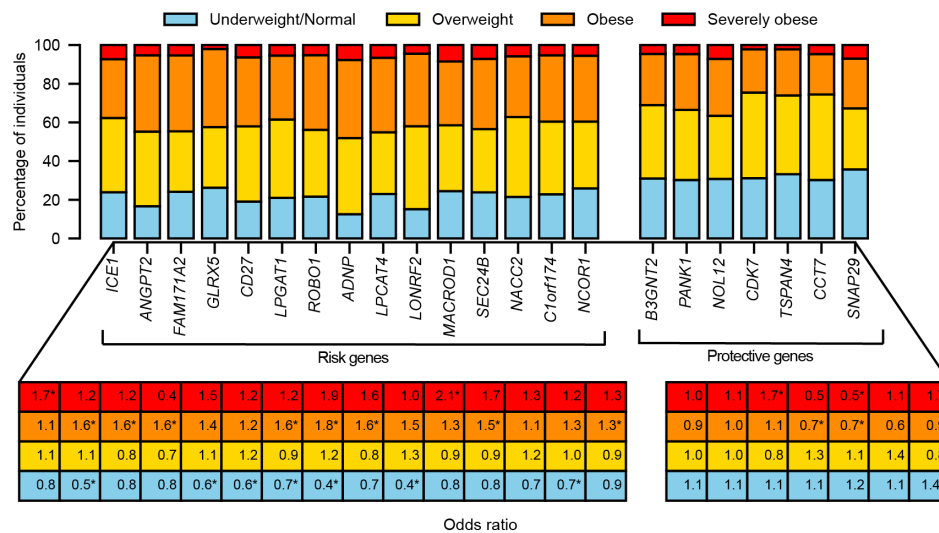
166 decreasing genes, PTV carriers in 15 genes, including *NEUROD6* (OR=2.78, 95% CI: 1.05, 7.61,  
167  $P=0.03$ ), *DCUNID3* (OR=2.49, 95% CI: 1.00, 6.22,  $P=0.04$ ), *SH3GL2* (OR=2.4, 95% CI: 1.49,  
168 3.87,  $P=0.0001$ ), and *VIRMA* (OR=1.47, 95% CI: 1.11, 1.94,  $P=0.006$ ), were more likely to be  
169 underweight or normal than individuals without PTVs in those genes (**Fig. 3a, Supplementary**  
170 **Fig. 2, Supplementary Table 7**). Further, individuals with PTVs in *VIRMA* (OR=0.51, 95% CI:  
171 0.35, 0.73,  $P=7.18 \times 10^{-5}$ ), *RABEP1* (OR=0.53, 95% CI: 0.32, 0.85,  $P=0.005$ ), *CDK7* (OR=0.67,  
172 95% CI: 0.46, 0.96,  $P=0.03$ ), and *P2RY1* (OR=0.44, 95% CI: 0.18, 0.96,  $P=0.03$ ) were less likely  
173 to be obese, while individuals with PTVs in *AQP3* (OR=0.13, 95% CI: 0.003, 0.76,  $P=0.01$ ) and  
174 *TSPAN4* (OR=0.46, 95% CI: 0.22, 0.86,  $P=0.01$ ) were also less likely to be severely obese. This  
175 analysis confirmed obesity risk and protective effects of the BMI-associated genes discovered in  
176 our study.

177 We also found significantly altered risks for 15 obesity-related comorbidities in carriers  
178 of PTVs in both obesity-risk conferring and protective genes compared to non-carriers (**Fig. 3b,**  
179 **Supplementary Table 8**). For instance, individuals with PTVs in *LPGAT1* showed higher risks  
180 for 5 out of 15 comorbidities, including aortic valve stenosis (OR=4.05, 95% CI: 1.46, 9.07,  
181  $P=0.005$ ), heart failure (OR=2.33, 95% CI: 1.17, 4.21,  $P=0.01$ ), and knee osteoarthritis  
182 (OR=2.27, 95% CI: 1.40, 3.55,  $P=0.0006$ ). Notably, *TMEM229A*, despite conferring obesity  
183 risk, showed depletion for heart failure phenotype in carriers (OR=0, 95% CI: 0, 0.93,  $P=0.03$ ).  
184 Among the obesity protective genes, PTV carriers in *VIRMA* and *SCFD1* showed reduced risk  
185 for most of the tested cardiometabolic phenotypes, including significant depletion for venous  
186 thromboembolism (OR=0, 95% CI: 0, 0.65,  $P=0.008$ ) and hypertension (OR=0.49, 95% CI:  
187 0.28, 0.85,  $P=0.007$ ), respectively. Further, PTV carriers in the BMI-decreasing *PANK1* and  
188 *P2RY1* exhibited an increased risk for type 2 diabetes (OR=1.74, 95% CI: 1.16, 2.54,  $P=0.005$ )  
189 and heart failure (OR=2.82, 95% CI: 0.87, 7.11,  $P=0.04$ ), respectively. While obesity risk-  
190 conferring genes typically associated with worse cardiometabolic profiles and protective genes  
191 with better profiles, notable exceptions were observed where obesity protective genes conferred  
192 risk and risk genes were protective for certain cardiometabolic phenotypes.

193



194  
 195 **Fig. 3: Enrichment of obesity and cardiometabolic comorbidities in PTV carriers of the**  
 196 **discovered BMI associated genes. (a) Top:** Proportion of PTV carriers within the obesity  
 197 clinical categories (underweight or normal, overweight, obese, and severely obese) that led to at  
 198 least 1.5 kg/m<sup>2</sup> increase (risk) or decrease (protective) in BMI. **Bottom:** Odds ratio of obesity  
 199 clinical categories in carriers compared to non-carriers. \**P*<0.05. Extended data with the odds  
 200 ratios, 95% confidence intervals and exact *P* values are available in **Supplementary Table 7.** (b)  
 201 Odds ratio for obesity-related comorbidities in PTV carriers of risk or protective genes compared  
 202 to non-carriers. \**P*<0.05, \*\**P*<0.01, \*\*\**P*<0.001. Extended data with the odds ratios, 95%  
 203 confidence intervals and exact *P* values are available in **Supplementary Table 8.**  
 204



205

206 **Supplementary Fig. 2: Enrichment in obesity clinical categories for BMI-associated genes.**

207 Proportion of PTV carriers in obesity clinical categories that led to between 1 to 1.5 kg/m<sup>2</sup>  
 208 increase (risk) or decrease (protective) in BMI. The odds ratio of being underweight or normal,  
 209 overweight, obese, and severely obese among carriers compared to non-carriers is also shown in  
 210 the table below (\**P*<0.05). All data with the odds ratio, 95% confidence intervals, and exact *P*  
 211 values are available in **Supplementary Table 7**.

212

213 **Functional analysis of obesity-associated genes**

214 We evaluated the impact of natural variation in plasma protein levels of the discovered genes on  
 215 BMI of individuals from the UK Biobank, irrespective of their PTV carrier status. Among the  
 216 discovered BMI-associated genes, six had plasma protein measurements from approximately  
 217 50,000 individuals. We constructed linear models using protein expression as the independent  
 218 variable and BMI, corrected for covariates, as the dependent variable. All six proteins had  
 219 significant model coefficients, with *DNER* having the highest absolute effect on BMI (**Fig. 4a**,  
 220 **Supplementary Table 9**), consistent with the association of *DNER* PTVs with severe obesity.  
 221 Protein levels of *FGF2* ( $\beta=0.21$ ,  $P=3.87 \times 10^{-14}$ ) and *GLOD4* ( $\beta=0.79$ ,  $P=7.02 \times 10^{-53}$ )  
 222 associated with increased BMI, supporting the observation that disruption of these genes leads to  
 223 a decrease in BMI (**Supplementary Fig. 3**). Interestingly, increased *CD27* and *ROBO1* levels  
 224 associated with increased BMI, contrary to the expected effect from PTVs in these genes, which  
 225 resulted in increased BMI. This suggests a complex relationship between protein expression of  
 226 these genes and their effect on BMI. Our results show that natural variation in protein expression  
 227 levels, potentially due to other common or rare non-coding variants, of the genes identified in  
 228 our study are associated with BMI.

229 We next assessed whether the 116 novel genes identified in our study had been  
230 previously linked to obesity or related pathways using genetic and functional approaches beyond  
231 rare variant association studies. Nineteen genes, including *MACROD1*, *VIRMA*, and *RABEP1*,  
232 had common variant signals associated with BMI recorded in the GWAS catalog<sup>19</sup>  
233 (**Supplementary Table 10 and 11**). We also queried the Common Metabolic Diseases  
234 Knowledge Portal<sup>20</sup>, which aggregates and scores genotype-phenotype relationships (Human  
235 Genetic Evidence, or HuGE scores<sup>21</sup>) based on evidence from both rare and common variant  
236 association studies. This analysis identified 16 more genes, including *LAMB2*, *FARP2*, *CELSR3*,  
237 and *ANGPT2*, that showed strong associations (HuGE score $\geq$ 10) with BMI or related phenotypes  
238 such as waist-hip ratio adjusted BMI or fasting insulin-adjusted BMI (see **Supplementary Table**  
239 **10 and 11**).

240 We further investigated the functional effects of knocking out obesity-associated genes  
241 using data from the International Mouse Phenotyping Consortium<sup>22</sup> and previous studies in  
242 mouse models. We found that 15 of our novel genes (five associated with risk and ten protective  
243 against obesity) exhibited directionally consistent obesity-related phenotypes in mice, such as  
244 changes in total body fat, lean mass, and glucose tolerance (**Table 2**). Among the obesity risk-  
245 conferring genes, we found impaired glucose tolerance, and increased weight gain in mouse  
246 models of *Lonrf2*<sup>22</sup>, *Ncor1*<sup>23</sup>, and *Gabra5*<sup>24</sup>, respectively. Mouse studies further highlighted the  
247 involvement of these genes in obesity-related processes such as thermogenesis (*Fgf2*<sup>25</sup> and  
248 *Gabra5*<sup>24</sup>), adipogenesis (*Ncor1*<sup>23</sup>), glucose metabolism (*Abca1*<sup>26</sup>), fatty acid metabolism  
249 (*Pml*<sup>27</sup>), and skeletal muscle metabolism (*Mipep*<sup>28</sup>). We also found direct evidence of protective  
250 effects from homozygous knockouts of *Aqp3*, *Kctd7*, and *Fgf2* in mice, resulting in increased  
251 lean mass, reduced total body fat, and resistance to obesity with a high-fat diet, respectively  
252 (**Table 2**). In addition, *Dcun1d3* homozygous knockout mice showed a significant reduction in  
253 total body fat, increased lean mass, and improved glucose tolerance. Finally, we performed a  
254 systematic literature review and found evidence for 71 genes with functional roles in obesity,  
255 including *MMP3*, *SLC25A1*, *CYP3A5*, *APMAP*, and *UCP3* (**Fig. 4b, Supplementary Table 12**).  
256 Overall, we found evidence linking 82 out of the 116 discovered genes to obesity or related  
257 phenotypes, based on previous association studies, plasma proteomics, and mouse models (**Fig.**  
258 **4b, Supplementary Table 11**).

259

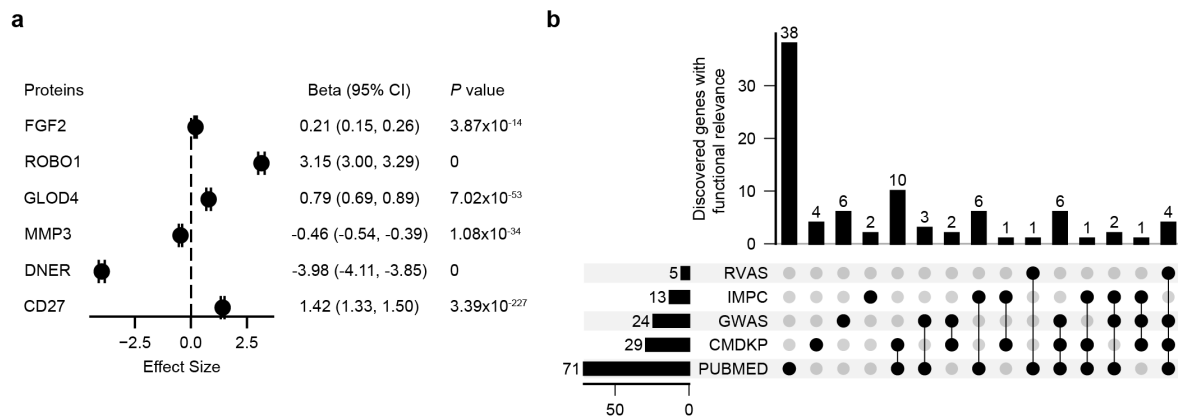
260 **Table 2:** List of mouse models of BMI-associated genes showing obesity-related phenotypes.

<b>Risk gene</b>	<b>Evidence from mouse models</b>
<i>ABCA1</i>	Increased weight gain in adipocyte specific <i>Abca1<sup>-ad/-ad</sup></i> KO mouse when fed a high fat diet <sup>26</sup>
<i>GABRA5</i>	Increased weight gain in chemogenetically inhibited (by clozapine N-oxide) mouse when fed a high fat diet <sup>24</sup>
<i>NCOR1</i>	Increased weight gain in adipocyte specific <i>NCoR<sup>fl/fl</sup>aP2-Cre<sup>+/-</sup></i> KO mouse when fed a high fat diet <sup>23</sup>
<i>LONRF2</i>	Impaired glucose tolerance observed in <i>Lonrf2<sup>em2(IMPC)H</sup></i> homozygous KO mouse model
<i>ANGPT2</i>	Protection from high fat diet induced obesity in adipose tissue specific overexpressed (by doxycycline) mouse <sup>29</sup>
<b>Protective gene</b>	<b>Evidence from mouse models</b>
<i>DCUNID3</i>	Decreased total body fat mass, increased lean body mass and improved glucose tolerance observed in <i>Dcun1d3<sup>em1(IMPC)Wtsi</sup></i> homozygous KO mouse
<i>AQP3</i>	Increased lean body mass observed in <i>Aqp3<sup>tm2b(EUCOMM)Wtsi</sup></i> homozygous KO mouse
<i>SNAP29</i>	Increased lean body mass observed in <i>Snap29<sup>tm1a(EUCOMM)Wtsi</sup></i> heterozygous KO mouse
<i>KCTD7</i>	Decreased total body fat amount observed in <i>Kctd7<sup>em2(IMPC)Bay</sup></i> homozygous KO mouse model
<i>FGF2</i>	Protection from high fat diet induced obesity in <i>Fgf2<sup>-/-</sup></i> homozygous KO mouse <sup>25</sup>
<i>MTFP1</i>	Protection from high fat diet induced obesity in liver specific <i>Alb-Cre<sup>tg/+</sup>Mtfp1<sup>LoxP/LoxP</sup></i> KO mouse <sup>30</sup>
<i>SETX</i>	Improved glucose tolerance observed in <i>Setx<sup>tm1b(EUCOMM)Wtsi</sup></i> homozygous KO mouse
<i>MIPEP</i>	Protection from high fat diet induced obesity in adipocyte specific <i>miPEP<sup>fllox/fllox</sup>Adiponectin-Cre<sup>+/-</sup></i> KO mouse <sup>28</sup>
<i>AHNAK</i>	Improved glucose tolerance observed in <i>Ahnak<sup>tm1b(KOMP)Mbp</sup></i> heterozygous KO mouse
<i>PML</i>	Protection from western diet induced obesity in <i>Pml<sup>-/-</sup></i> mouse <sup>27</sup>

261

262

263

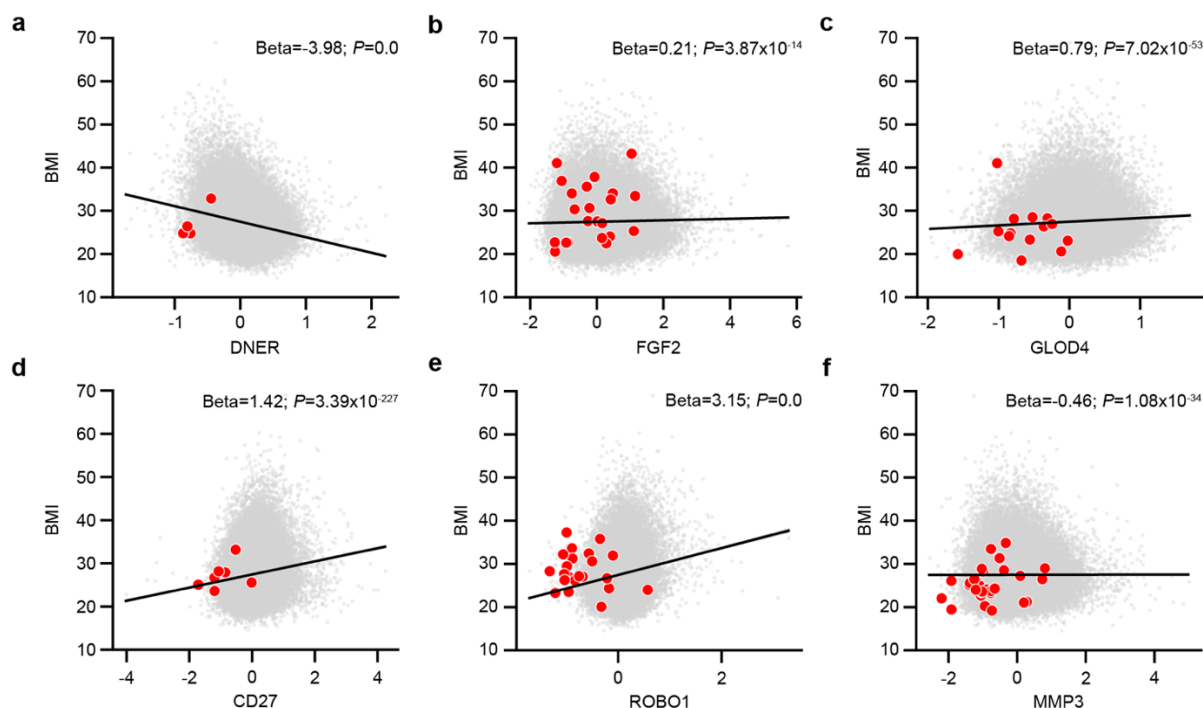


264

265 **Fig. 4: Functional validation of BMI associated genes.** (a) Model coefficients, with 95 % CI  
 266 and *P* value, of a linear model constructed using plasma protein expression values of BMI-  
 267 associated genes as independent variable and BMI as the dependent variable. Data with exact *P*  
 268 values, model coefficients and 95% confidence interval are available in **Supplementary Table**  
 269 **9.** (b) Upset plot of evidence for a role of the identified BMI-associated genes in obesity-related  
 270 functions and phenotypes. The bars represent the number of genes with functional relevance  
 271 found from the respective study, which is represented by the dots. RVAS, Rare variant  
 272 association study; IMPC, International mouse phenotypic consortium; GWAS: Genome wide  
 273 association study; CMDKP: Common metabolic diseases knowledge portal.

274

275



276

277

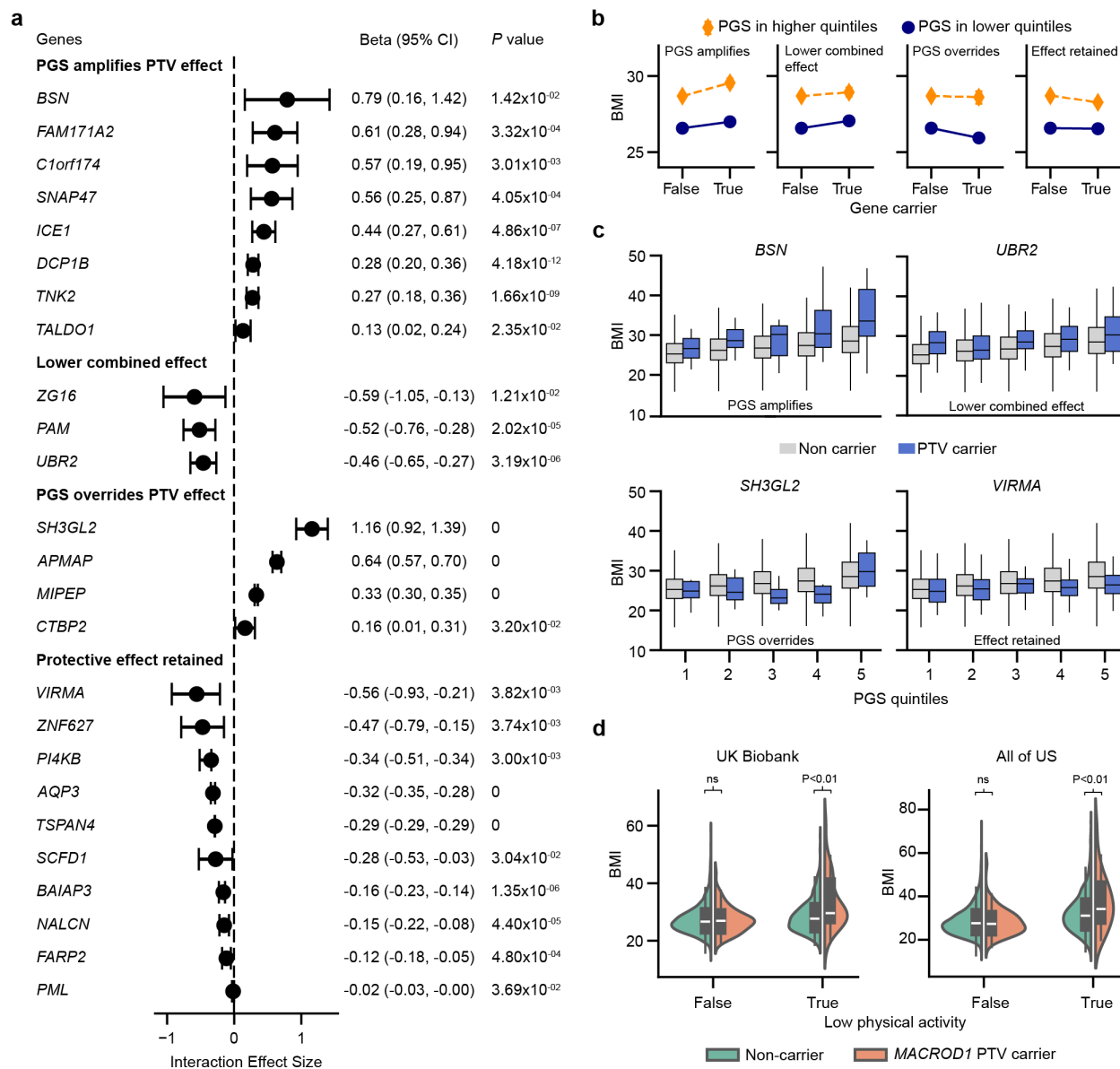
278 **Supplementary Fig. 3: Relationship between plasma protein levels and BMI.** Association of  
 279 the plasma protein levels with BMI in the population (grey scatters) and the PTV carriers (red  
 280 scatters) for (a) DNER, (b) FGF2, (c) GLOD4, (d) CD27, (e) ROBO1, (f) MMP3.

## 281 **Interactions of obesity-associated genes with polygenic risk and lifestyle factors**

282 We assessed how PGS modulated the effect of the identified genes on BMI (**Supplementary**  
283 **Fig. 4a**). We used linear regression to test multiplicative interactions between PGS and each  
284 disrupted gene towards changes in BMI. We identified four scenarios where PGS contributed to  
285 the variability in BMI among individuals with PTVs in the same gene (**Fig. 5a and b**,  
286 **Supplementary Fig. 4b**). In the *first* scenario, BMI was synergistically increased by both PGS  
287 and PTVs in risk genes, leading to more severe obesity. *For example*, individuals with PTVs in  
288 *TSM, LPGAT1, FAMI71A2, ICE1, TNK2*, and *TPRX1* and eight other genes showed a  
289 synergistic increase in BMI in concert with PGS. This pattern was more evident in carriers of  
290 *BSN* PTVs where a non-additive increase in BMI was observed across PGS quintiles (**Fig. 5c**).  
291 *Second*, the combined effect of PGS and PTVs in five risk genes, including *UBR2* and *PAM*, was  
292 much lower than their expected additive effects on BMI (**Fig. 5c**). *Third*, PGS was found to  
293 override the effect of six protective genes such as *SH3GL2* and *APMAP*, and individuals with  
294 high PGS still showed increased BMI despite carrying PTVs in BMI-decreasing genes (**Fig. 5c**).  
295 In the *fourth* scenario, we noted that 23 protective genes override the strong obesity risk-  
296 conferring effect of PGS, and individuals remain protected despite carrying high polygenic risk  
297 for obesity. *For example*, the effect of PGS in the highest quintile among carriers of PTVs in  
298 *VIRMA* was 2.53 kg/m<sup>2</sup> less than in non-carriers (**Fig. 5c**). We did not identify non-additive  
299 effects of PGS on other disrupted genes such as *MC4R*, *APBA1* and *ROBO1*, consistent with  
300 previous reports<sup>7,9</sup> (**Supplementary Table 13**).

301 We next investigated the effect of obesity-inducing lifestyle factors on the BMI of PTV  
302 carriers after controlling for PGS. We found that the effects of PTVs in *MACROD1* and *VIRMA*  
303 was modulated by the degree of physical activity of individuals. In particular, individuals  
304 carrying *MACROD1* PTVs with low physical activity (UKB) or bad physical health (AoU) had a  
305 higher BMI compared to non-carriers, in both UKB and AoU cohorts (**Fig. 5d, Supplementary**  
306 **Table 14**). On the other hand, the protective effect of *VIRMA* was more pronounced in  
307 individuals with low physical activity in the UKB cohort but was diminished in individuals self-  
308 reported to be in bad physical health in the AoU cohort (**Supplementary Fig. 4c**,  
309 **Supplementary Table 14**). This observation could be attributed to differences in how lifestyle  
310 factors are measured across UKB and AoU, or other cohort-specific factors. We also identified  
311 associations for carriers of *MC4R*, *TNK2*, and *GADLI* PTVs with diet, sleep, and sedentary

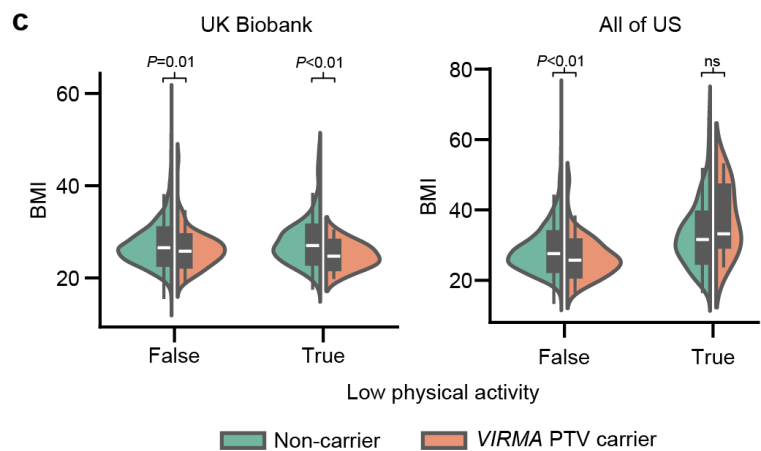
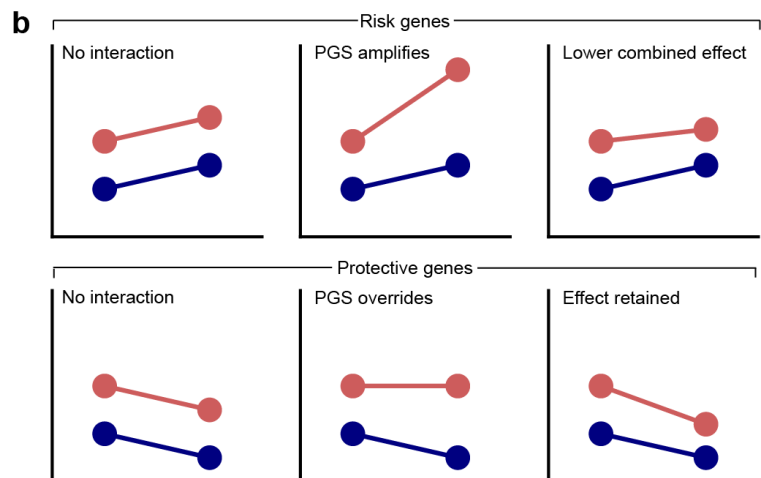
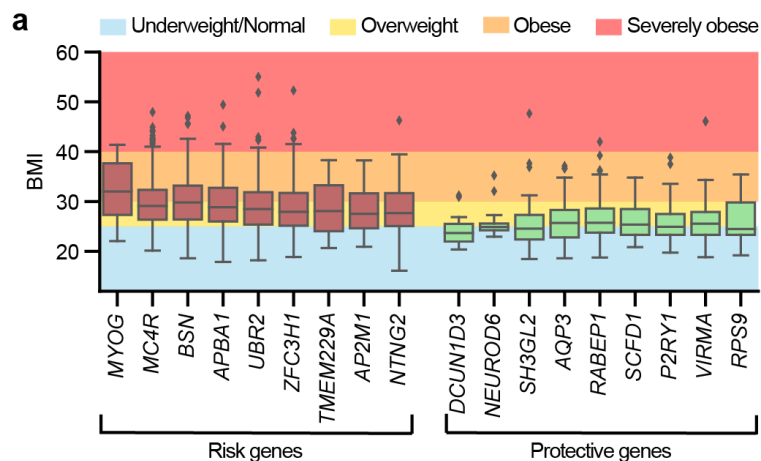
312 lifestyle factors, respectively, in UKB (**Supplementary Table 15**), but these associations could  
 313 not be tested in AoU due to the lack of comparable data. Our results underscore the influence of  
 314 PGS and obesogenic factors towards the effect conferred by obesity risk and protective genes.  
 315



316  
 317 **Fig. 5: Effect of obesity-associated genes modulated by polygenic and obesogenic risk**  
 318 **factors. (a)** Interaction model coefficient plot of combined risk of PTVs in the discovered genes  
 319 and PGS on BMI, categorized by the four different scenarios. Data with exact *P* values and other  
 320 associated statistics are available in **Supplementary Table 13**. **(b)** Interaction plots showing the  
 321 mean BMI of PTV carriers and non-carriers in the higher (fourth and fifth) and lower (first,  
 322 second, and third) PGS quintiles are shown. The PTV carriers of genes were aggregated based on  
 323 their interaction category with PGS. **(c)** BMI distribution across PGS quintiles of PTV carriers  
 324 (blue) compared to non-carriers (grey) of the representative examples for each of the four



325 interaction scenarios. (d) BMI distribution of *MACROD1* PTV carriers compared to non-carriers,  
 326 stratified by their physical activity levels in UKB (left) and AoU (right) cohorts.  
 327



328  
 329 **Supplementary Fig. 4: Effect of obesity-associated genes modulated by polygenic and**  
 330 **obesogenic risk factors.** (a) BMI distribution of PTV carriers in high-effect obesity genes. (b)  
 331 Interaction plots depicting the different scenarios through which PGS modulates the BMI of

332 individuals with PTVs in the discovered genes. (c) BMI distribution of carriers of PTVs in  
333 *VIRMA* compared to non-carriers stratified by their physical activity levels in UKB (left) and  
334 AoU (right) cohorts.

335

## 336 **DISCUSSION**

337 Here, we discovered 116 genes with consistent effects on BMI across ancestries and provide  
338 accurate estimates of effect sizes for known obesity genes. By using genetic data from both  
339 European and non-European populations, we uncovered gene-discovery biases present in  
340 previous studies particularly in genes such as *DIDO1*, *SPARC*, and *RAB21* that showed  
341 significant associations only in Europeans. Typically, protective obesity genes have been more  
342 challenging to identify than risk genes, but three notable ones, *GPR75*, *GPR151*, and *GIPR*, have  
343 emerged more recently through rare variant association studies<sup>6,9,31</sup>. However, our findings  
344 suggest that these genes may also be influenced by ancestral biases, with variable effect sizes  
345 across populations, which could impact their potential as therapeutic targets for obesity. In  
346 contrast to previous studies, we identified genes with consistent effects across ancestries and  
347 functional evidence, such as mouse knockout models for these genes, confirm their role in  
348 obesity-related pathways. These findings emphasize the value of a cross-ancestry analysis, which  
349 enabled the discovery of several novel genes with effect sizes comparable to canonical obesity  
350 genes such as *MC4R* and *BSN*.

351 A significant finding from our study was the identification of 66 protective genes, some  
352 of which exhibit effect sizes more than twice that of known protective genes. Notably,  
353 *DCUNID3* showed the highest effect size, with PTVs in these genes leading to approximately  
354 3.8 kg/m<sup>2</sup> decrease in BMI. This gene is highly expressed in adipocytes, and *Dcun1d3* knockout  
355 mice showed reduced fat body mass, increased lean body mass and improved glucose  
356 tolerance<sup>22</sup>. *DCUNID3* encodes a protein involved in neddylation, a post translation  
357 modification that attaches a small ubiquitin-like molecule, NEDD8, to substrate proteins, altering  
358 their properties<sup>32,33</sup>. Disruption of neddylation has been linked to impaired energy metabolism  
359 and age-related metabolic disorders<sup>34,35</sup>. While the mechanisms by which *DCUNID3* might  
360 influence these pathways or diseases are unclear, our findings suggest that its inhibition could be  
361 a therapeutic target for obesity. Similarly, *VIRMA* consistently led to decreased BMI, is  
362 supported by evidence from nearby common variant signals<sup>19</sup>, and individuals with PTVs in this  
363 gene displayed improved cardiometabolic health. Two other protective genes, *AQP3* and *FGF2*,

364 not only showed evidence from mouse models for obesity protection but also have well-  
365 established functional roles in glycerol metabolism<sup>36</sup> and thermogenesis<sup>25</sup>. Currently,  
366 pharmacological inhibitors or negative modulators, such as Bisacodyl and Pentosan polysulfate,  
367 exist for *AQP3* and *FGF2*; however, they are used as a laxative and for bladder pain,  
368 respectively<sup>37</sup>. We also found that Setmelanotide, the *MC4R* agonist used for weight loss, is also  
369 an agonist of *ANGPT2*, an obesity risk gene discovered in our study, with evidence from mouse  
370 models demonstrating its role in body weight regulation<sup>29,38</sup>. Thus, our discoveries provide  
371 avenues for repurposing existing drugs or devising new ones to alleviate obesity risk.

372 While polygenic risks have been found to additively influence rare variant risk for  
373 obesity<sup>9</sup>, a recent study showed *BSN* and PGS interact non-additively. In addition to  
374 recapitulating the effect between *BSN* and PGS, we found synergistic effects of PGS on other  
375 genes, such as *TSMF*, *LPGATI*, and *FAMI71A2*, as well. Further, synergistic interactions  
376 between obesity causing risk factors can also exacerbate the risk of severe obesity and its  
377 associated morbidities. *For instance*, carriers of PTVs in *MACROD1*, an obesity risk gene,  
378 showed significantly higher BMI when they were physically inactive compared to carriers who  
379 were active. While previous studies have shown the adverse effect of lifestyle factors on  
380 common variants associated with BMI<sup>2</sup>, we provide evidence that rare variant risk can be  
381 modulated by obesogenic factors. Notably, carriers of PTVs in protective genes such as *VIRMA*  
382 and *PI4KB*, remained protected despite having high PGS, highlighting the complex patterns of  
383 interactions between specific obesity genes and PGS. Understanding these genetic and gene-  
384 environment interactions could be key to developing more effective interventions, where lifestyle  
385 changes may mitigate genetic risk in high-risk individuals, and protective variants might offer  
386 resilience against obesogenic factors. Overall, using a multi-ancestry approach, we provide a  
387 more comprehensive view of obesity genetics, identifying novel genes, refining the effects of  
388 known genes, and demonstrating the importance of including diverse populations in genetic  
389 studies.  
390

## 391 **METHODS**

### 392 **Variant quality control, filtering, and annotation**

393 We analyzed whole exome sequencing data of 469,835 individuals from the UK Biobank (UKB)  
394 cohort, available as multi-sample project variant call format (pVCF) files in the UKB Research  
395 Analysis Platform (RAP). The sequencing method and preparation of the pVCF files have been  
396 previously described<sup>39</sup>. Using the Hail<sup>40</sup> platform on DNANexus, we first split multi-allelic  
397 records, and then filtered to retain rare variants with intracohort frequency  $<0.001$ . These  
398 variants were annotated using variant effect predictor<sup>41</sup> (VEP v109) and dbNSFP v4<sup>42</sup> to identify  
399 the predicted variant effects on gene transcripts. All variants within protein coding genes in the  
400 autosomes with a call rate of 50% were used for subsequent analysis. We then annotated each  
401 variant based on their most deleterious functional impact on the transcript into three categories,  
402 from the most harmful to the least harmful, as follows: (1) Loss of function or “lof”, which  
403 included frameshift, stop gained, splice acceptor, and splice donor VEP annotations, (2)  
404 “missense strict”, which included missense variants predicted to be deleterious by nine  
405 deleteriousness prediction tools (SIFT<sup>43</sup>, LRT<sup>44</sup>, FATHMM<sup>45</sup>, PROVEAN<sup>46</sup>, MetaSVM<sup>47</sup>,  
406 MetaLR<sup>47</sup>, PrimateAI<sup>48</sup>, DEOGEN2<sup>49</sup>, and MutationAssessor<sup>50</sup>) available through dbNSFP  
407 database, and (3) “missense lenient”, which included missense variants predicted to be  
408 deleterious by at least seven out of the nine deleteriousness prediction tools. All UKB analyses  
409 were performed in the DNANexus UKB RAP.

410 We also analyzed whole genome sequencing data of 245,388 individuals in the All of Us  
411 (AoU) cohort. Variant call files for regions overlapping with the exomes were available as a Hail  
412 matrix table in the AoU portal. After filtering the variants for intracohort frequency ( $<0.001$ ), we  
413 annotated the variants using Nirvana<sup>51</sup>, available in the AoU research platform as a Hail table,  
414 and determined the functional impact of each variant using the same criteria as defined for the  
415 UKB data. All AoU analyses were performed in the Researcher Workbench of the AoU portal.

416

### 417 **Phenotype, obesogenic lifestyle, polygenic risk analyses**

418 *For the UKB cohort*, BMI, genetic sex (Data Field 22001), age (Data Field 21003), ethnic  
419 background (Data Field 21000), PGS (Data Field 26216), genetic kinship (Data Field 22021),  
420 and the top 40 genetic principal components (PCs; Data Field 22009) on 502,368 individuals in  
421 UKB were accessed and preprocessed through the UKB RAP. Numerical fields such as BMI and

422 age, measured across multiple visits, were averaged. Categorical fields such as sex and ethnic  
423 background were searched for inconsistent readings, and samples with conflicting values across  
424 multiple assessments were dropped. We further filtered samples with 10 or more third-degree  
425 relatives from our analysis based on previously published kinship estimates<sup>14</sup>. We finally divided  
426 the cohort into two populations, based on their ethnic background namely, “British” consisting of  
427 419,228 individuals self-reported to be “White British” based on the UKB data-field 21000, and  
428 the other 68,070 individuals who we categorized as “non-British” population. We note that the  
429 UKB non-British population contains individuals of European ancestry, *for example*, individuals  
430 with self-reported “Irish” ethnic background. ***Obesogenic lifestyle factors*** such as Metabolic  
431 Equivalent Task (MET) scores, sleep duration, alcohol consumption, or smoking habits were  
432 also accessed and preprocessed through the UKB RAP. Numerical fields were averaged and  
433 samples with categorical fields having conflicting values across multiple assessments were  
434 dropped. All fields were then binarized based on their extreme values and their potential effect  
435 on BMI as follows: Numerical fields with the potential to have a directly proportional effect on  
436 BMI, such as “time spent watching television”, were assigned as 1 for an individual, if their  
437 corresponding value was greater than 95% quantile. On the other hand, fields with inversely  
438 proportional effect on BMI, such as “consumption of cooked vegetable”, were assigned as 1 if  
439 the value was less than 5% quantile. Individuals with extreme values from ordinal fields that  
440 potentially lead to increased BMI were assigned as 1, and individuals with all other values were  
441 assigned as 0. *For example*, individuals who consumed alcohol daily or almost daily, the extreme  
442 category for this field, were encoded as 1 for that lifestyle factor. A full description of the  
443 lifestyle factors used in this study and the thresholds or categories used to binarize them is  
444 available in **Supplementary Table 16**. After binarizing, the fields were combined to define one  
445 of the six lifestyle factors, namely, (i) physical activity, (ii) alcohol, (iii) smoke, (iv) diet, (v)  
446 sleep, and (vi) sedentary lifestyle, used in this study. Binarized MET scores, alcohol  
447 consumption frequency, and sleep patterns were directly used as indicators of physical activity,  
448 alcohol, and sleep metrics, respectively. Individuals whose current or past smoking tendencies  
449 were high were considered to have the “smoke” lifestyle. Similarly, those spending more time on  
450 TV or computer were indicative of “sedentary lifestyle”. Diet was denoted as 1 if either the sum  
451 of high processed meat-, beef-, mutton-, and pork-intake binarized columns was greater than 0,  
452 or the sum of low cooked vegetable, salad, fresh fruit, dried fruit, oily fish, or non-oily fish-

453 intake binarized columns was greater than 1. Thus, the diet field used in the study is analogous to  
454 high-meat and/or low vegetable consumption. International Classification of Disease (ICD-10)  
455 10<sup>th</sup> revision summary diagnosis codes corresponding to each individual were also extracted  
456 from the Hospital Episode Statistics (HES) data available in the UKB RAP.

457 *For the AoU cohort*, we obtained BMI, age, genetic sex, previously calculated genetic  
458 ancestry<sup>15</sup>, 16 genetic PCs, and sample IDs of related individuals from the AoU Workbench. We  
459 calculated the age based on the date of birth and the BMI measurement data concepts (“concept”  
460 defined here as a collection of similar data stored together) available in AoU. We used the “sex  
461 at birth” concept, available in AoU, to denote the genetic sex of an individual. Only individuals  
462 with BMI values between 12 and 75 kg/m<sup>2</sup> and “sex at birth” field listed as either “Male” or  
463 “Female” were included to ensure consistency between UKB and AoU. Based on the predicted  
464 genetic ancestry, we divided the AoU data into three populations namely, (1) European  
465 (N=127,644), which included individuals predicted to be of European ancestry, (2) African  
466 (N=54,865), which included individuals predicted to be of African ancestry, and (3) Mixed  
467 (N=52,134), which included individuals predicted to be of Admixed American ancestry and  
468 other ancestries. *We calculated PGS for BMI in individuals from AoU* by using the GWAS  
469 summary statistics from a previous study<sup>4</sup>. Duplicate and ambiguous single nucleotide  
470 polymorphism (SNPs) were filtered from the summary statistics. Genotype data of the AoU  
471 cohort filtered for common variants (MAF>0.01) was obtained as a Hail matrix table. We further  
472 removed variants with Hardy-Weinberg Equilibrium p-value <1x10<sup>-6</sup> and retained variants that  
473 overlapped between summary statistics data and the AoU genotyped data with a call rate greater  
474 than 0.9. We then used the predetermined effect sizes of the SNPs as Bayesian priors, weighed  
475 them based on the alternate allele frequency observed in individuals, and summed the weighted  
476 effect sizes to obtain individual-specific polygenic risk for BMI. All calculations were conducted  
477 using Hail, available in the AoU Researcher Workbench. *Obesogenic lifestyle factors* such as  
478 physical activity, alcohol consumption, and smoking tendency in the AoU cohort were obtained  
479 from survey questionnaires available through the AoU research portal. Individuals falling in the  
480 extreme categories for each of the three factors, (1) Physical activity: “General Physical Health:  
481 Fair” or “General Physical Health: Poor”, (2) Alcohol consumption, “Drink Frequency Past  
482 Year: 4 or More Per Week”, and (3) Smoking tendency: “Smoke Frequency: Every Day”, were  
483 annotated as 1, denoting those carrying the respective obesogenic factor, while others not falling

484 in these extreme categories were annotated as 0. We note that physical activity measurements in  
485 AoU were not equivalent to those in UKB, where a more rigorous assessment of an individual's  
486 physical health was measured using International Physical Activity Questionnaire guidelines.  
487 Additionally, diet and sedentary lifestyle information was not available in AoU, while sleep data  
488 were available on less than 10% of the cohort and these lifestyle factors could not be tested in  
489 AoU. ICD-10 diagnosis codes corresponding to each individual in the AoU cohort was extracted  
490 using the researcher workbench.

491

### 492 **Gene burden association test using REGENIE**

493 We used REGENIE v3.3<sup>16</sup> to conduct gene burden association tests. Similar to most whole  
494 genome regression tools, REGENIE operates in two steps. *First*, it uses SNP data, preferably a  
495 genotyped array of SNPs, from across the genome to fit a null model that estimates a polygenic  
496 score for the trait to be tested (i.e., BMI in our study). This step accounts for population structure  
497 and relatedness between samples. For both UKB and AoU cohorts, we used the genotyped SNP  
498 array files available in DNANexus or Researcher workbench for this step. We lifted over the  
499 SNP array files in UKB from version hg19 to hg38 using Picard's LiftoverVCF tool<sup>52</sup>. In the  
500 *second* step, REGENIE then calculates the association between the genetic variants of interest  
501 and the trait after considering the null model calculated in step 1 and other user defined  
502 covariates. In our study, we used age, genetic sex, and first ten genetic PCs obtained from both  
503 UKB and AoU cohorts as additional covariates. REGENIE is capable of collapsing variants to a  
504 gene-level with user defined masks and annotation files apart from the usual variant data as input  
505 before running association tests. We supplied our already annotated variant file based on their  
506 impact on gene transcripts and defined three variant masks to collapse variants on a gene level,  
507 namely, (a) "lof" variants only, (b) "lof" and "missense strict" variants, and (c) "lof", "missense  
508 strict" and "missense lenient" variants. We performed gene-based association tests for all the five  
509 defined populations across both biobanks and generated statistics for all individual gene-mask  
510 pairs.

511

### 512 **Meta analysis and variance calculation**

513 We first created three meta populations from the previously defined five populations (UKB  
514 British, UKB non-British, AoU European, AoU African, and AoU Mixed), the first composed of

515 individuals from European ancestry (UKB British and AoU European) termed “European meta”,  
516 the second from non-European ancestry (UKB non-British, AoU African and AoU Mixed)  
517 termed “non-European meta”, and the third from all five populations termed “combined meta”.  
518 We note that a minority of individuals (approximately 15%) in the non-European population  
519 were from ethnic groups such as ‘Irish’ or ‘any other white background’, and are therefore likely  
520 to be of European ancestry.

521 Results from the individual populations were pooled into the three meta populations  
522 using an inverse variance weighted random effects model. The meta-analysis calculation was  
523 implemented using the python package statsmodels v0.14.2. Each gene-variant mask was  
524 associated with meta statistics for three meta populations. Any gene-variant mask that passed the  
525 Bonferroni multiple testing correction threshold of  $8.34 \times 10^{-7}$ , accounting for 20,000 genes and  
526 three variant collapsing models, *either* in the European meta *or* the non-European meta *and* in  
527 the combined meta population were considered to be significantly associated with BMI across  
528 ancestries.

529 We calculated the interpopulation variance in effect sizes across populations for the  
530 discovered genes passing our multiple testing criterion, as well as previously associated BMI  
531 genes, using the “var” method available in the python package, pandas v2.1.1.

532

### 533 **Odds-ratio calculation for obesity clinical categories and obesity-related disorders**

534 To calculate risk across the obesity clinical categories among carriers of PTVs in the discovered  
535 genes compared to non-carriers, we first categorized each individual into their respective clinical  
536 category based on their BMI values: underweight or normal (<25), overweight (25 to 30), obese  
537 (30 to 40) and severely obese (>40). We then created 2x2 contingency tables using their carrier  
538 status as the first variable and whether they belong to the respective obesity category as opposed  
539 to any other category as the second variable. Finally, we calculated the conditional odds ratio and  
540 the significance value using a two-sided Fisher’s exact test.

541 To calculate risk for obesity related disorders, we obtained the specific ICD codes for  
542 comorbidities frequently associated with obesity based on previous BMI related studies and  
543 calculated the odds of PTV carriers diagnosed with the comorbidity compared to non-carriers  
544 using Fisher’s exact test. A list of the comorbidities and the ICD codes used to categorize an



545 individual as carrying the comorbidity is provided in **Supplementary Table 17**. All odds ratio  
546 calculations were conducted using the `scipy v1.11.3` package available in python.

547

### 548 **Functional analysis of the discovered genes**

549 We overlapped the list of discovered genes with a BMI gene list obtained from the GWAS  
550 catalogue<sup>19</sup>, where the SNPs identified in previous BMI GWAS studies are mapped to their most  
551 likely gene targets. We further overlapped the discovered gene list with gene phenotype  
552 associations from Common Metabolic Diseases Knowledge Portal (CMDKP) database<sup>20</sup>.

553 CMDKP aggregates information and scores genotype-phenotype relationships (Human Genetic  
554 Evidence or HuGE scores<sup>21</sup>) based on the evidence acquired from previous rare and common  
555 variant association studies for all common metabolic disorders, including obesity and related  
556 phenotypes. Any gene that showed strong associations (HuGE score  $\geq 10$ , indicating “strong”  
557 evidence of gene-phenotype link) with obesity, BMI, or related phenotypes, such as fasting  
558 insulin adjusted BMI, or waist hip ratio adjusted BMI, were considered an overlapping BMI  
559 related gene. Associations of the discovered genes with previous *in vivo* studies of knockout  
560 (KO) mouse models were first identified using data from the International Mouse Phenotyping  
561 Consortium (IMPC)<sup>22</sup>, which provides comprehensive phenotypic characterization of mouse  
562 KOs, including assessments of body weight and obesity-related traits. IMPC represents an  
563 unbiased resource for investigating the *in vivo* effects of novel genes that have not been studied  
564 in the context of obesity. Further, we conducted a PubMed search of each gene by searching for  
565 the gene name in the abstract or title and “obesity” in the text. This provided additional  
566 functional evidence for the novel genes, which have been phenotypically characterized using  
567 mouse models, or other complementary approaches such as human genetic studies. The BMI-  
568 associated GWAS gene list was directly obtained from NHGRI-EBI catalog, while associations  
569 from CMDKP, IMPC, and PubMed were accessed programmatically through their REST API  
570 using python’s `requests v2.31.0` module.

571

### 572 **Proteomics data analysis of oligogenic combinations**

573 Normalized plasma protein expression data of 1,463 proteins on 50,956 individuals were  
574 accessed through the DNANexus portal of UKB RAP. Multiple replicates of protein readouts  
575 were averaged, and protein levels of corresponding genes were identified. Using linear

576 regression models, the effect of protein expression levels for the discovered genes with available  
577 protein data on BMI was measured, while accounting for covariates such as age, genetic sex, ten  
578 genetic PCs, PGS, and the interaction term between the protein and the PGS. The model was  
579 trained using “ols” function from statsmodels v0.14.2 package available in python.

580

### 581 **Interaction models for obesity risk factors**

582 To assess interactive effects between each gene and PGS, we used a linear regression model to  
583 predict BMI using individual as well as interactive terms between PGS and PTV carrier status  
584 for the discovered genes while accounting for age, genetic sex, and ten genetic PCs. We obtained  
585 separate model coefficients in UKB and AoU cohorts and then used random effect meta-analysis  
586 to combine the statistics from the two cohorts. Interaction terms with meta-analysis  $P$  value less  
587 than 0.05 were considered significant gene-PGS interactions. The models were trained using  
588 “ols” function from statsmodels v0.14.2 package, available in python.

589 To compare interactive effects between lifestyle factors and gene PTV carrier status, we  
590 conducted factorial analysis of covariance using the individual gene and lifestyle terms as well as  
591 the gene-lifestyle interaction term, while accounting for age, genetic sex, first ten genetic PCs  
592 and PGS, separately in UKB and AoU cohorts. Interaction terms that crossed the  $P$  value  
593 threshold (0.05) in both UKB and AoU were considered significant gene-lifestyle interactions.  
594 Model training and calculation of  $P$  value significance and confidence intervals were performed  
595 using “ols” and “anova\_lm” function from statsmodels v0.14.2 package available in python.

596

597 **AUTHOR CONTRIBUTIONS**

598 S.G and D.B conceived the study, acquired, analyzed and interpreted the data, and wrote the  
599 manuscript.

600

601 **ACKNOWLEDGEMENTS**

602 We thank Corrine Smolen and Jiawan Sun for critical reading of the manuscript. We thank the  
603 participants and investigators in the UK Biobank and All of US research studies who made this  
604 work possible. This research has been conducted using the UK Biobank Resource under  
605 Application Number 45023. We also thank the National Institutes of Health's All of Us Research  
606 Program for making available the participant data examined in this study. This work was  
607 supported by NIH R01-GM121907, resources from the Huck Institutes of the Life Sciences and  
608 the Pennsylvania State University to S.G. The authors declare no competing interests.

609

610 **Data and code availability**

611 The UK Biobank and All of Us genetic and phenotypic data analyzed in this study are publicly  
612 available to registered researchers through their respective analysis portals. Additional  
613 information about registration for access to the data is available at <https://www.ukbiobank.ac.uk/>  
614 and <https://www.researchallofus.org/> for UK Biobank and All of Us, respectively. The code for  
615 pre-processing, filtering, annotating genetic data, association tests and statistical analysis is  
616 available on GitHub ([https://github.com/deeprob/BMI\\_monogenic](https://github.com/deeprob/BMI_monogenic)).

617

618

619 **REFERENCES**

- 620 1. Kivimäki, M. *et al.* Body-mass index and risk of obesity-related complex multimorbidity: an  
621 observational multicohort study. *Lancet Diabetes Endocrinol* **10**, 253–263 (2022).
- 622 2. Kilpeläinen, T. O. *et al.* Physical activity attenuates the influence of FTO variants on obesity  
623 risk: a meta-analysis of 218,166 adults and 19,268 children. *PLoS Med* **8**, e1001116 (2011).
- 624 3. Kim, M. S. *et al.* Association of genetic risk, lifestyle, and their interaction with obesity and  
625 obesity-related morbidities. *Cell Metab* **36**, 1494-1503.e3 (2024).
- 626 4. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology.  
627 *Nature* **518**, 197–206 (2015).
- 628 5. Khera, A. V. *et al.* Polygenic Prediction of Weight and Obesity Trajectories from Birth to  
629 Adulthood. *Cell* **177**, 587-596.e9 (2019).
- 630 6. Turcot, V. *et al.* Protein-altering variants associated with body mass index implicate  
631 pathways that control energy intake and expenditure in obesity. *Nat Genet* **50**, 26–41 (2018).
- 632 7. Zhao, Y. *et al.* Protein-truncating variants in BSN are associated with severe adult-onset  
633 obesity, type 2 diabetes and fatty liver disease. *Nat Genet* **56**, 579–584 (2024).
- 634 8. Zhu, N. *et al.* Rare predicted loss of function alleles in Bassoon (BSN) are associated with  
635 obesity. *NPJ Genom Med* **8**, 33 (2023).
- 636 9. Akbari, P. *et al.* Sequencing of 640,000 exomes identifies GPR75 variants associated with  
637 protection from obesity. *Science* **373**, eabf8683 (2021).
- 638 10. Fatumo, S. *et al.* A roadmap to increase diversity in genomic studies. *Nat Med* **28**, 243–250  
639 (2022).
- 640 11. Cohen, J. *et al.* Low LDL cholesterol in individuals of African descent resulting from  
641 frequent nonsense mutations in PCSK9. *Nat Genet* **37**, 161–165 (2005).

- 642 12. Genovese, G. *et al.* A risk allele for focal segmental glomerulosclerosis in African  
643 Americans is located within a region containing APOL1 and MYH9. *Kidney Int* **78**, 698–704  
644 (2010).
- 645 13. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic  
646 Studies. *Cell* **177**, 1080 (2019).
- 647 14. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data.  
648 *Nature* **562**, 203–209 (2018).
- 649 15. All of Us Research Program Genomics Investigators. Genomic data in the All of Us  
650 Research Program. *Nature* **627**, 340–346 (2024).
- 651 16. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and  
652 binary traits. *Nat Genet* **53**, 1097–1103 (2021).
- 653 17. Paule, R. C. & Mandel, J. Consensus Values and Weighting Factors. *J Res Natl Bur Stand*  
654 (1977) **87**, 377–385 (1982).
- 655 18. Kaisinger, L. R. *et al.* Large-scale exome sequence analysis identifies sex- and age-specific  
656 determinants of obesity. *Cell Genom* **3**, 100362 (2023).
- 657 19. Sollis, E. *et al.* The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource.  
658 *Nucleic Acids Res* **51**, D977–D985 (2023).
- 659 20. Common Metabolic Diseases Knowledge Portal (cmdkp.org); (RRID:SCR\_020937).
- 660 21. Dornbos, P. *et al.* Evaluating human genetic support for hypothesized metabolic disease  
661 genes. *Cell Metab* **34**, 661–666 (2022).
- 662 22. International Mouse Phenotyping Consortium.
- 663 23. Li, P. *et al.* Adipocyte NCoR knockout decreases PPAR $\gamma$  phosphorylation and enhances  
664 PPAR $\gamma$  activity and insulin sensitivity. *Cell* **147**, 815–826 (2011).

- 665 24. Sa, M. *et al.* Hypothalamic GABRA5-positive neurons control obesity via astrocytic GABA.  
666 *Nat Metab* **5**, 1506–1525 (2023).
- 667 25. Li, H. *et al.* FGF2 disruption enhances thermogenesis in brown and beige fat to protect  
668 against adiposity and hepatic steatosis. *Mol Metab* **54**, 101358 (2021).
- 669 26. de Haan, W., Bhattacharjee, A., Ruddle, P., Kang, M. H. & Hayden, M. R. ABCA1 in  
670 adipocytes regulates adipose tissue lipid content, glucose tolerance, and insulin sensitivity. *J*  
671 *Lipid Res* **55**, 516–523 (2014).
- 672 27. Cheng, X. *et al.* Ablation of promyelocytic leukemia protein (PML) re-patterns energy  
673 balance and protects mice from obesity induced by a Western diet. *J Biol Chem* **288**, 29746–  
674 29759 (2013).
- 675 28. Diaz-Vegas, A. *et al.* Deletion of miPEP in adipocytes protects against obesity and insulin  
676 resistance by boosting muscle metabolism. *Mol Metab* **86**, 101983 (2024).
- 677 29. An, Y. A. *et al.* Angiopoietin-2 in white adipose tissue improves metabolic homeostasis  
678 through enhanced angiogenesis. *Elife* **6**, e24071 (2017).
- 679 30. Patitucci, C. *et al.* Mtfp1 ablation enhances mitochondrial respiration and protects against  
680 hepatic steatosis. *Nat Commun* **14**, 8474 (2023).
- 681 31. Jurgens, S. J. *et al.* Rare coding variant analysis for human diseases across biobanks and  
682 ancestries. *Nat Genet* **56**, 1811–1820 (2024).
- 683 32. Meyer-Schaller, N. *et al.* The human Dcn1-like protein DCNL3 promotes Cul3 neddylation  
684 at membranes. *Proc Natl Acad Sci U S A* **106**, 12365–12370 (2009).
- 685 33. Monda, J. K. *et al.* Structural conservation of distinctive N-terminal acetylation-dependent  
686 interactions across a family of mammalian NEDD8 ligation enzymes. *Structure* **21**, 42–53  
687 (2013).

- 688 34. Yu, M. *et al.* Emerging role of NEDD8-mediated neddylation in age-related metabolic  
689 diseases. *Ageing Res Rev* **94**, 102191 (2024).
- 690 35. Zhang, S., Yu, Q., Li, Z., Zhao, Y. & Sun, Y. Protein neddylation and its role in health and  
691 diseases. *Signal Transduct Target Ther* **9**, 85 (2024).
- 692 36. Madeira, A., Moura, T. F. & Soveral, G. Aquaglyceroporins: implications in adipose biology  
693 and obesity. *Cell Mol Life Sci* **72**, 759–771 (2015).
- 694 37. Cannon, M. *et al.* DGIdb 5.0: rebuilding the drug-gene interaction database for precision  
695 medicine and drug discovery platforms. *Nucleic Acids Res* **52**, D1227–D1235 (2024).
- 696 38. Ni, B. *et al.* Selective adipocyte loss of Angiopoietin-2 prompts female-specific obesity and  
697 metabolic syndrome. *Mol Metab* **65**, 101588 (2022).
- 698 39. Backman, J. D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants.  
699 *Nature* **599**, 628–634 (2021).
- 700 40. Hail Team. Hail 0.2. (<https://github.com/hail-is/hail>).
- 701 41. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
- 702 42. Liu, X., Li, C., Mou, C., Dong, Y. & Tu, Y. dbNSFP v4: a comprehensive database of  
703 transcript-specific functional predictions and annotations for human nonsynonymous and  
704 splice-site SNVs. *Genome Med* **12**, 103 (2020).
- 705 43. Ng, P. C. & Henikoff, S. Predicting deleterious amino acid substitutions. *Genome Res* **11**,  
706 863–874 (2001).
- 707 44. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes.  
708 *Genome Res* **19**, 1553–1561 (2009).
- 709 45. Shihab, H. A. *et al.* Predicting the functional, molecular, and phenotypic consequences of  
710 amino acid substitutions using hidden Markov models. *Hum Mutat* **34**, 57–65 (2013).

- 711 46. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect  
712 of amino acid substitutions and indels. *PLoS One* **7**, e46688 (2012).
- 713 47. Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for  
714 nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet* **24**, 2125–2137  
715 (2015).
- 716 48. Sundaram, L. *et al.* Predicting the clinical impact of human mutation with deep neural  
717 networks. *Nat Genet* **50**, 1161–1170 (2018).
- 718 49. Raimondi, D. *et al.* DEOGEN2: prediction and interactive visualization of single amino acid  
719 variant deleteriousness in human proteins. *Nucleic Acids Res* **45**, W201–W206 (2017).
- 720 50. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations:  
721 application to cancer genomics. *Nucleic Acids Res* **39**, e118 (2011).
- 722 51. Stromberg, M. *et al.* Nirvana: Clinical Grade Variant Annotator. in *Proceedings of the 8th*  
723 *ACM International Conference on Bioinformatics, Computational Biology, and Health*  
724 *Informatics* 596–596 (ACM, Boston Massachusetts USA, 2017).  
725 doi:10.1145/3107411.3108204.
- 726 52. Picard (<http://broadinstitute.github.io/picard/>).
- 727