

1 Rare variant associations with birth weight identify  
2 genes involved in adipose tissue regulation, placental  
3 function and insulin-like growth factor signalling

4  
5 Katherine A. Kentistou\*<sup>1</sup>, Brandon E. M. Lim\*<sup>2</sup>, Lena R. Kaisinger\*<sup>1</sup>, Valgerdur  
6 Steinthorsdottir<sup>3</sup>, Luke N Sharp<sup>2</sup>, Kashyap A. Patel<sup>2</sup>, Vinicius Tragante<sup>3</sup>, Gareth  
7 Hawkes<sup>2</sup>, Eugene J. Gardner<sup>1</sup>, Thorhildur Olafsdottir<sup>3</sup>, Andrew R. Wood<sup>2</sup>, Yajie Zhao<sup>1</sup>,  
8 Gudmar Thorleifsson<sup>3</sup>, Felix R. Day<sup>1</sup>, Susan E. Ozanne<sup>6</sup>, Andrew T. Hattersley<sup>2</sup>,  
9 Stephen O’Rahilly<sup>6</sup>, Kari Stefansson<sup>3,4</sup>, Ken K. Ong\*<sup>1,5</sup>, Robin N. Beaumont\*<sup>2</sup>, John R.  
10 B. Perry\*<sup>1,6</sup>, Rachel M. Freathy\*<sup>2</sup>

11  
12 1. MRC Epidemiology Unit, Box 285 Institute of Metabolic Science, University of Cambridge  
13 School of Clinical Medicine, Cambridge CB2 0QQ, UK

14 2. Department of Clinical and Biomedical Sciences, Faculty of Health and Life Sciences,  
15 University of Exeter, Exeter, UK

16 3. deCODE genetics/Amgen, Inc., 102 Reykjavik, Iceland

17 4. Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland

18 5. Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK

19 6. MRC Metabolic Diseases Unit, Institute of Metabolic Science, University of Cambridge  
20 School of Clinical Medicine, University of Cambridge, Cambridge CB2 0QQ, UK

21 \* These authors contributed equally to this work.

## 22 Abstract

23 Investigating the genetic factors influencing human birth weight may lead to biological  
24 insights into fetal growth and long-term health. Genome-wide association studies of birth  
25 weight have highlighted associated variants in more than 200 regions of the genome, but the  
26 causal genes are mostly unknown. Rare genetic variants with robust evidence of association  
27 are more likely to point to causal genes, but to date, only a few rare variants are known to  
28 influence birth weight. We aimed to identify genes that harbour rare variants that impact birth  
29 weight when carried by either the fetus or the mother, by analysing whole exome sequence  
30 data in UK Biobank participants.

31  
32 We annotated rare (minor allele frequency <0.1%) protein-truncating or high impact missense  
33 variants on whole exome sequence data in up to 234,675 participants with data on their own  
34 birth weight (fetal variants), and up to 181,883 mothers who reported the birth weight of their  
35 first child (maternal variants). Variants within each gene were collapsed to perform gene  
36 burden tests and for each associated gene, we compared the observed fetal and maternal  
37 effects. We identified 8 genes with evidence of rare fetal variant effects on birth weight, of  
38 which 2 also showed maternal effects. One additional gene showed evidence of maternal  
39 effects only. We observed 10/11 directionally concordant associations in an independent  
40 sample of up to 45,622 individuals (sign test  $P=0.01$ ).

41  
42 Of the genes identified, *IGF1R* and *PAPPA2* (fetal and maternal-acting) have known roles in  
43 insulin-like growth factor bioavailability and signalling. *PPARG*, *INHBE* and *ACVR1C* (all  
44 fetal-acting) have known roles in adipose tissue regulation and rare variants in the latter two  
45 also showed associations with favourable adiposity patterns in adults. We highlight the dual  
46 role of *PPARG* in both adipocyte differentiation and placental angiogenesis. *NOS3*, *NRK*, and  
47 *ADAMTS8* (fetal and maternal-acting) have been implicated in both placental function and  
48 hypertension.

49  
50 Analysis of rare coding variants has identified regulators of fetal adipose tissue and  
51 fetoplacental angiogenesis as determinants of birth weight, as well as further evidence for the  
52 role of insulin-like growth factors.

## 53 Introduction

54 An improved understanding of the genetic contribution to birth weight may highlight  
55 mechanisms relevant to fetal growth restriction or overgrowth, as well as links between fetal  
56 growth and later disease<sup>1</sup>. Studies of rare, monogenic forms of diabetes have highlighted the  
57 key role of fetal insulin, since rare single gene mutations in the fetus that reduce insulin  
58 secretion are generally associated with a reduced birth weight. These include mutations that  
59 cause neonatal diabetes, where the severity of the insulin secretory defect correlates with the  
60 degree of reduction in birth weight<sup>2</sup>. In babies who produce no insulin at all due to pancreatic  
61 agenesis, or to complete loss-of-function (LoF) mutations in the insulin gene, birth weights  
62 are 50% of those of a healthy term baby<sup>3</sup>.

63  
64 Family studies of mutations in the *GCK* gene have demonstrated the role of maternal genetic  
65 variants, which may influence birth weight independently of variants inherited by the fetus.  
66 Maternal heterozygous mutations in *GCK*, which reduce glucose sensing and cause fasting  
67 glucose to be regulated at a higher set-point, result in greater maternal glucose availability to  
68 the fetus, as glucose crosses the placenta. This causes the fetus, if it does not carry the  
69 mutation, to produce more fetal insulin and grow bigger in response<sup>4</sup>. Conversely, if the fetus  
70 carries a *GCK* mutation, its ability to sense the increase in glucose is impaired, causing  
71 reduced fetal insulin secretion and concomitant reduced fetal growth.

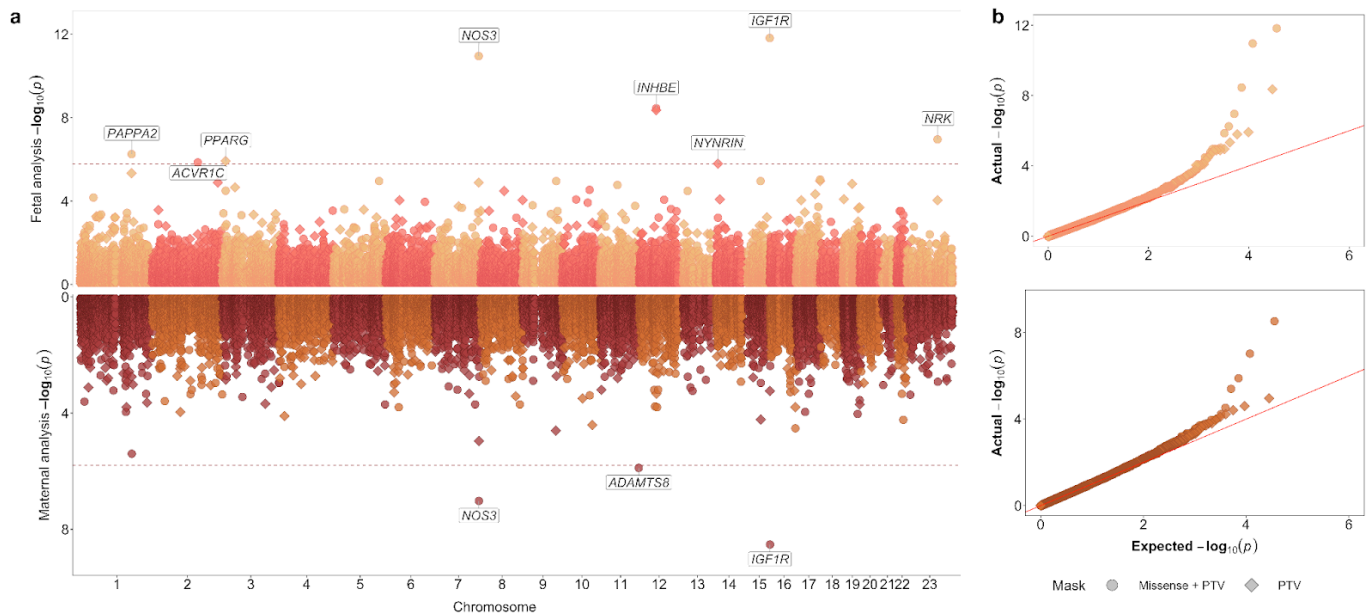
72  
73 Genome-wide association studies of birth weight have identified >200 regions of the genome  
74 where common variants (minor allele frequency (MAF) >1%) are associated with birth  
75 weight<sup>5,6</sup>. Variants at the identified loci influence birth weight by direct effects of the fetal  
76 genotype, indirect effects of the maternal genotype acting on the intrauterine environment, or  
77 a combination of the two. Most of these variants likely influence growth via mechanisms that  
78 are independent of fetal insulin secretion<sup>7</sup>. Parent-of-origin effects in the fetus have been  
79 observed at several birth weight loci<sup>6</sup>. However, the causal genes at the vast majority of  
80 identified loci are unknown.

81  
82 Whole exome sequence (WES) data in biobank-scale studies offer the opportunity to identify  
83 genes that are causally related to birth weight, and the potential to uncover new mechanisms  
84 of importance for the regulation of fetal growth. We performed exome-wide association  
85 studies (ExWAS) of rare variant (MAF <0.01) gene burden with birth weight in up to 234,675  
86 UK Biobank participants who reported their own birth weight (fetal variants), and up to  
87 181,883 female UK Biobank participants who reported the birth weight of their first child  
88 (maternal variants). The identified genes highlight key roles for adipose tissue regulation and  
89 fetoplacental angiogenesis, in addition to the role of insulin-like growth factor bioavailability  
90 and signalling, in regulating human birth weight.

## 91 Results

92 We identified 8 genes in which rare (MAF<0.1%) deleterious LoF fetal variants, defined as  
93 either high-confidence protein truncating variants, or missense variants with a CADD score

94  $\geq 25$ , were associated with birth weight at exome-wide significance (BOLT-LMM  
95  $P < 1.64 \times 10^{-6}$ ): *ACVR1C*, *IGF1R*, *INHBE*, *NOS3*, *NRK*, *NYNRIN*, *PAPPA2*, *PPARG* (**Figure 1**,  
96 **Supplementary Figures 1-3, Supplementary Table 1**). In the analysis of maternal variants  
97 and offspring birth weight, 3 genes showed associations at exome-wide significance  
98 (BOLT-LMM  $P < 1.58 \times 10^{-6}$ ). These included 2 genes also identified in the fetal analysis  
99 (*IGF1R*, *NOS3*) and additionally *ADAMTS8* (**Figure 1, Supplementary Figures 4-5**,  
100 **Supplementary Table 1**). We also confirmed all of these significant associations by  
101 conducting independent analyses using REGENIE (see Methods) and we did not observe  
102 any method-related biases (**Supplementary Figures 1, Supplementary Table 1**).



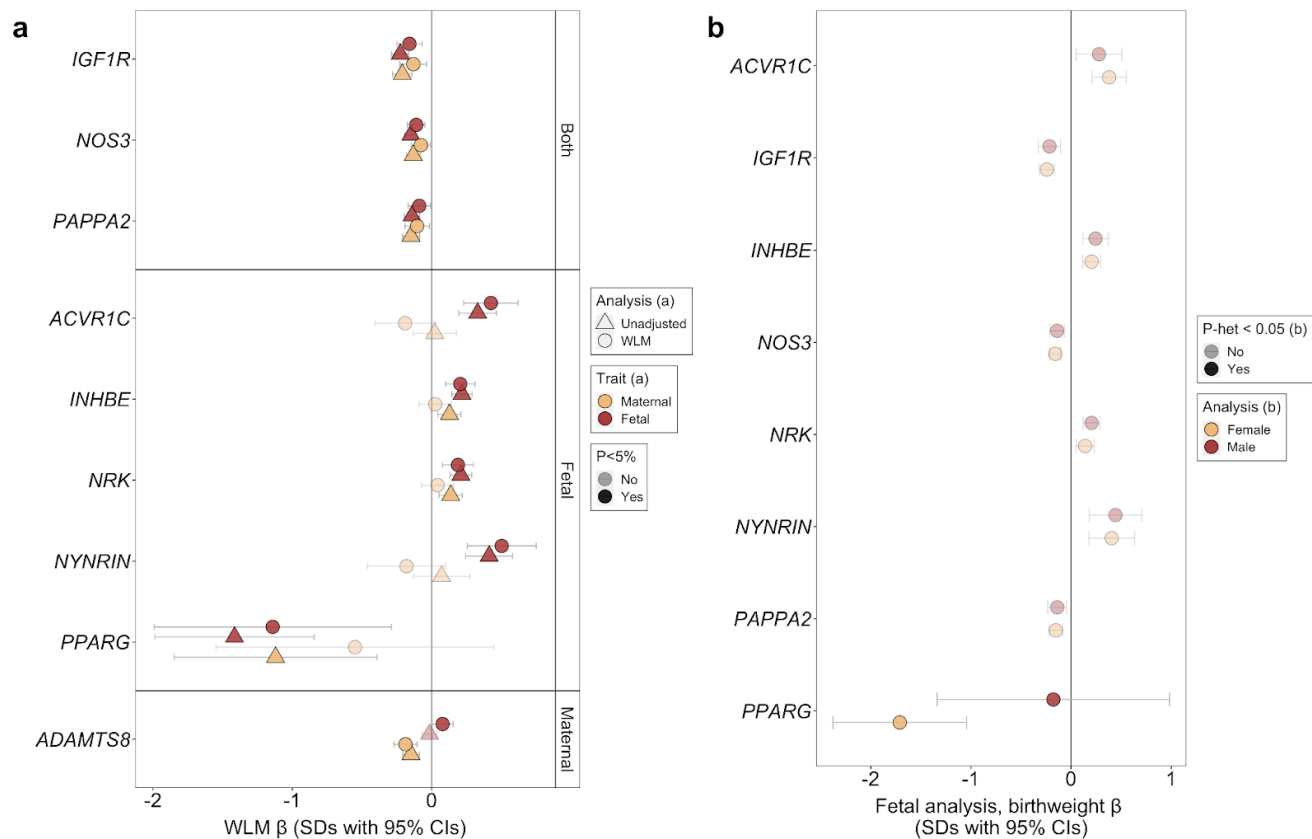
103 **Figure 1 | Exome-wide rare (MAF<0.1%) fetal or maternal variant gene burden**  
104 **associations with birth weight.** (a) Miami plot showing gene burden test results from  
105 BOLT-LMM for birth weight, with the fetal exome-wide analysis (up to  $n=234,675$ ) on the top  
106 panel and the maternal exome-wide analysis (up to  $n=181,883$ ) on the bottom panel. Gene  
107 associations passing exome-wide significance, at  $P < 1.64 \times 10^{-6}$  in the fetal analysis and  
108  $P < 1.58 \times 10^{-6}$  in the maternal analysis, are labelled. The two variant collapsing masks are  
109 indicated by point shapes. (b) QQ plots for exome-wide gene burden associations. Results  
110 for the exome-wide significant genes are included in Supplementary Table 1.

111

112 Since maternal and fetal genotypes are correlated ( $r=0.5$ ), we explored whether the observed  
113 associations represented fetal effects, maternal effects, or both. To do this, we used a  
114 weighted linear model (WLM) to approximately condition the fetal effect on the maternal  
115 genotype, and vice versa. Five out of nine associations (*ACVR1C*, *INHBE*, *NRK*, *NYNRIN*,  
116 *PPARG*) showed evidence of only fetal-genotype effects. *IGF1R*, *PAPPA2*, and *NOS3* were  
117 classified as both fetal- and maternal-acting, with rare variants in all three genes associating  
118 with a lower birth weight in both cases. *ADAMTS8* was the only gene classified as only  
119 maternal-acting (**Figure 2a, Supplementary Table 2**).

120

121 For fetal genes associated with birth weight, we explored whether they might act in a  
 122 sex-specific manner. We found a nominally stronger effect for fetal *PPARG* variants on birth  
 123 weight in females (beta<sub>Female</sub> = -1.710 [0.340] SDs) than in males (beta<sub>Male</sub> = -0.177 [0.592] SDs,  
 124  $p_{\text{het}} = 2.46 \times 10^{-02}$ ), albeit only 3 male carriers were identified (**Figure 2b, Supplementary Table**  
 125 **3**).



126

127 **Figure 2 | Nine genes with rare variant associations with birth weight.** (a)  
 128 Weighted-linear models (WLMs) approximately conditioned the fetal effect on the maternal  
 129 genotype, and vice versa. Marker opacity indicates gene burden associations at  $P < 0.05$ . (b)  
 130 Sexually-dimorphic effects of the associated genes on birthweight. Markers are coloured  
 131 yellow or red to indicate female- or male-only models. Marker opacity indicates associations  
 132 with sexual dimorphism ( $P < 0.05$ ). Units for effect estimates and accompanying 95% CIs are  
 133 SDs in both panels. Relevant data are included in Supplementary Tables 2 and 3,  
 134 accordingly.

### 135 Confirmation of exome associations

136 We aimed to replicate our associations in independent data on up to 45,622 Icelandic  
 137 genomes. Despite the substantially smaller sample size than our discovery sample, we  
 138 observed supportive evidence for the same type of rare LoF exome variants on birth weight

139 for the majority of our associations (10/11 showed directional consistency; Sign test  $P=0.01$ ),  
140 with three genes associated at  $P<0.05$ ; *INHBE*, *NRK* and *PAPPA2* (**Supplementary Figure**  
141 **6, Supplementary Table 4**).

142  
143 To identify whether any of the birth weight genes identified in our exome analyses are also  
144 supported by common variant associations, we examined the latest maternal and fetal GWAS  
145 summary statistics for birth weight<sup>6</sup> and looked for GWAS signals within 500kb of our  
146 identified genes. We observed fetal GWAS signals proximal to *ACVR1C*, *IGF1R*, and  
147 *PAPPA2*, and there was also a maternal GWAS signal proximal to *NOS3* (**Supplementary**  
148 **Figure 7**). In all cases the lead associated variants were located within these genes  
149 (**Supplementary Table 5**) and proximity has been shown to be a good predictor of the causal  
150 gene at a GWAS locus<sup>8,9</sup>. The lead SNPs at the *IGF1R* and *NOS3* loci were also eQTLs for  
151 these genes, with directions of effect concordant with the exome associations, i.e. the alleles  
152 associated with lower expression (LoF-like) were also associated with lower birth weight<sup>10</sup>. Of  
153 these four GWAS signals, *IGF1R*, *ACVR1C*, and *PAPPA2* were classified as having fetal only  
154 effects by Juliusdottir *et al.*<sup>6</sup>. The SNP near *NOS3* was not classified by Juliusdottir *et al.*<sup>6</sup>, but  
155 the *NOS3* lead SNP was classified by Warrington *et al.*<sup>5</sup> as having only maternal effects. We  
156 did not observe attenuation of the rare variant associations with birth weight when  
157 conditioning on the proximal GWAS sentinel SNPs (**Supplementary Table 6**). Hence, these  
158 common variant associations provide independent support for these genes in the regulation  
159 of birth weight.

160  
161 To assess evidence for additional candidate genes influencing birth weight at GWAS loci,  
162 which did not pass our stringent exome-wide significance thresholds, we looked at all genes  
163 within 300 kb of a reported GWAS locus and applied a Benjamini-Hochberg correction for  
164 multiple testing. The results are shown in **Supplementary Table 7**. In the fetal variant  
165 analysis we identified a further eight candidate genes (*CDK6*, *HGS*, *PAPPA*, *PHF19*, *PLAG1*,  
166 *PLCE1*, *PTEN* and *SKP2*), and in the maternal variant analysis we identified one additional  
167 candidate gene (*LMNA*). At six of these genes (*CDK6*, *PAPPA*, *PHF19*, *PLAG1*, *PLCE1* and  
168 *SKP2*) a GWAS lead SNP was located within the same gene, while the lead SNP within  
169 *PLCE1* was also a missense variant, further adding to the evidence that non-synonymous  
170 coding variation in these genes affects birth weight.

171  
172 Finally, we explored whether the biology implicated by the exome-associated genes was  
173 consistent with that from GWAS studies. To do this, we performed a pathway enrichment  
174 analysis, using genes proximal to the GWAS signals from Juliusdottir *et al.*<sup>6</sup>. We saw  
175 enrichment for 185 GO pathways in the fetal GWAS of birth weight and 41 pathways in the  
176 maternal GWAS, with 21 pathways being enriched in both GWAS (fetal and maternal,  
177 **Supplementary Table 8**). In line with biology implicated by our exome associations, we  
178 observed significant enrichment for genes in multiple pathways implicated in insulin and  
179 growth factor response across both fetal and maternal GWAS genes and “circulatory system  
180 development” for fetal GWAS genes, among others.

## 181 Gene burden associations with related traits

182 To understand the wider phenotypic impacts of birth weight-associated rare genomic  
183 variation, we performed phenotypic association lookups in birthweight-related traits using  
184 data from the UK Biobank (see **Methods**).

185  
186 We observed associations for rare LoF variants in *IGF1R* and *PAPPA2*, both of which are  
187 known regulators of insulin-like growth factor (IGF) regulation and bioactivity<sup>11–15</sup>, with adult  
188 circulating IGF-1 levels, adult height and childhood height and body size (**Supplementary**  
189 **Table 9**). The *INHBE*, *NRK*, *NOS3* and *NYNRIN* genes were also associated with adult  
190 height. In all of these genes, apart from *INHBE*, carrying a rare LoF variant was associated  
191 with directionally concordant effects on birth weight and adult height, i.e. higher birth weight  
192 and higher adult height or *vice versa*.

193  
194 Furthermore, we observed associations between variants in *ADAMTS8* and *NOS3* and adult  
195 blood pressure (**Supplementary Table 9**). For both of these genes, rare LoF variants were  
196 associated with lower birth weight and higher blood pressure.

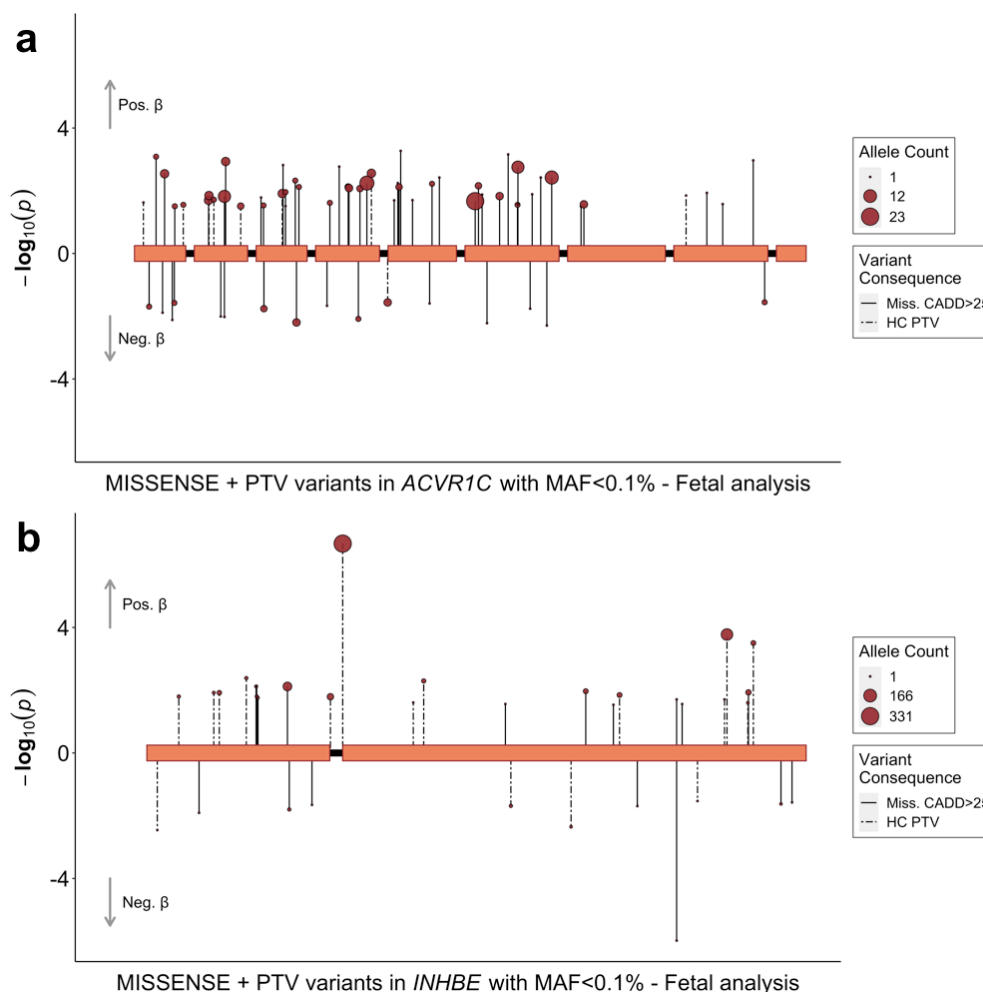
197  
198 Finally, several of the birth weight associated genes were also associated with measures of  
199 adiposity in adulthood, including waist-hip ratio (WHR) adjusted for BMI, and body fat  
200 percentage for rare LoF variants in *ACVR1C*, *INHBE* and *PPARG* (accordingly,  
201 **Supplementary Table 9**). These are further explored below.

## 202 *ACVR1C*, *INHBE* and favourable adiposity

203 *ACVR1C* (birth weight:  $\beta_{\text{fetal}}=0.330$  SDs,  $p=1.4\times 10^{-6}$ , N=200 carriers, **Figure 3a**) encodes  
204 a type I receptor for the TGF-beta family of signalling molecules. It is predominantly  
205 expressed in adipose tissue<sup>10</sup> and its primary function in metabolic regulation is to limit  
206 catabolic activities and preserve energy<sup>16</sup>. Low-frequency heterozygous mutations in humans  
207 have been linked to a favourable metabolic profile including lower WHR adjusted for BMI and  
208 protection against T2D<sup>17,18</sup>. *INHBE* ( $\beta_{\text{fetal}}=0.218$  SDs,  $p=3.5\times 10^{-9}$ , N=687 carriers, **Figure**  
209 **3b**) encodes the inhibin subunit beta E of activin E, a growth factor belonging to the  
210 TGF-beta family. While this gene is predominantly expressed in the liver<sup>10</sup>, it does not appear  
211 to be necessary for normal liver function<sup>19</sup>, but rather acts as a negative regulator of energy  
212 storage in peripheral adipose tissue. Similarly to *ACVR1C*, a candidate receptor,  
213 heterozygous LoF carriers exhibit lower WHR adjusted for BMI and a favourable body fat  
214 distribution and metabolic profile<sup>20,21</sup>. However, recent mouse studies<sup>22,23</sup> showed that  
215 complete ablation of *INHBE* or *ACVR1C* leads to uncontrolled and pathological levels of  
216 lipolysis (see Discussion). We found that rare LoF variants in *ACVR1C* and *INHBE* are  
217 associated with increased birth weight and decreased WHR, indicative of a metabolically  
218 favourable adiposity pattern in adulthood (**Supplementary Table 9**).

219  
220 Two genetics studies<sup>20,21</sup> showed that the association between rare LoF variants in *INHBE*  
221 and WHR is primarily due to a splice acceptor variant (12:57456093:G:C) that substantially  
222 reduces the expression of the gene, attributable to either a change in secretion and/or protein

223 stability. They also reported that the favourable adiposity effect of this gene is almost entirely  
224 attenuated in the absence of this variant (gene-burden  $p=2.01\times 10^{-8}$  with splice variant,  
225  $p=0.34$  without<sup>21</sup>, which we also observe for WHR in our data (**Supplementary Table 10**).  
226 However, while this variant is also strongly associated with higher birth weight ( $p=5.4\times 10^{-6}$ ,  
227 **Figure 3b**), our observed birth weight association with *INHBE* does not fully attenuate in the  
228 absence of this variant (gene-burden  $p=4.6\times 10^{-9}$  with splice variant,  $p=6.3\times 10^{-5}$  without,  
229 **Supplementary Table 10**).  
230



231

232 **Figure 3 | Rare variant associations at *INHBE* and *ACVR1C* with fetal birth weight in**  
233 **the UK Biobank.** Fetal variant-level associations between *ACVR1C* (a), *INHBE* (b) and birth  
234 weight. Included variants had a minor allele frequency (MAF) <0.1% and were annotated to  
235 be damaging variants defined as high-confidence protein truncating variants (PTV) or  
236 missense variants with a CADD score  $\geq 25$ . Each variant is presented as an individual line  
237 extending to its association  $p$ -value ( $-\log_{10}$ ) in the direction indicating the direction of effect on  
238 birth weight in variant carriers. Dashed lines indicate PTVs and solid lines indicate missense  
239 variants. The point size indicates the number of carriers of each variant (i.e. the allele count).  
240



## 241 *PPARG* in birth weight, adipogenesis, and placental formation

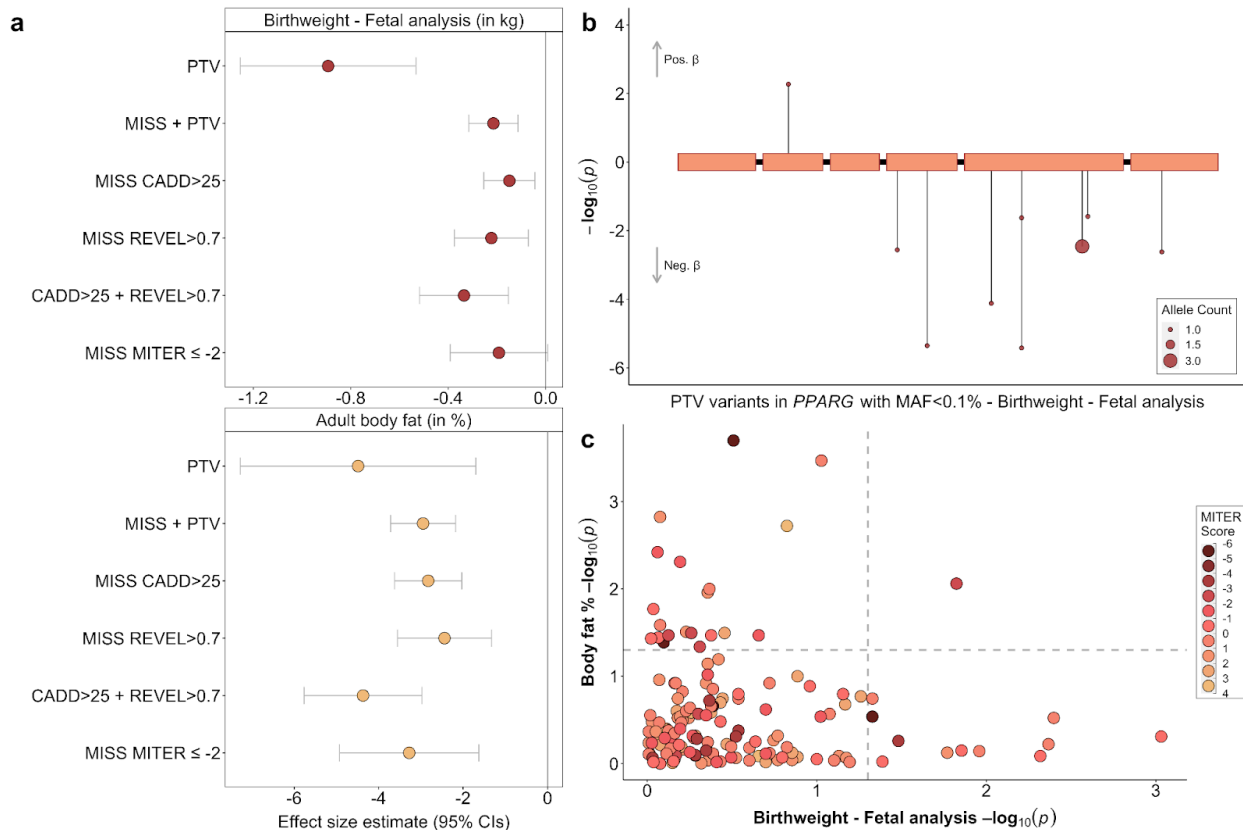
242 *PPARG* (birth weight:  $\beta_{\text{fetal}} = -1.416$  SDs,  $p = 1.2 \times 10^{-6}$ ,  $N = 11$  carriers, **Figure 4b**) encodes the  
243 peroxisome proliferator-activated receptor PPAR-gamma, which acts as a master regulator of  
244 adipogenesis. Heterozygous LoF variants in humans have been linked to familial partial  
245 lipodystrophy (FPLD) characterised by loss of subcutaneous fat from the extremities and  
246 metabolic abnormalities such as insulin resistance and hypertriglyceridemia<sup>24-27</sup>. Individuals  
247 with FPLD due to inherited heterozygous variants have also been shown to have low  
248 birthweight for their gestational age, when in the absence of maternal diabetes, hypertension  
249 or hypertriglyceridemia (e.g. when the *PPARG* variants are paternally inherited)<sup>28</sup>. In addition  
250 to being highly expressed in adipocytes, *PPARG* is also highly expressed in trophoblasts in  
251 both rodent and human placentas<sup>29</sup> where it appears to play an important role in placental  
252 formation and angiogenesis by regulating the expression of proangiogenic factors such as  
253 proliferin (*Prl2c2*) and vascular endothelial growth factor (VEGF)<sup>30</sup>. Its depletion results in  
254 embryonic lethality in mice due to placental dysfunction<sup>31,32</sup>, while PPAR $\gamma$  agonist activity was  
255 reduced in the blood and placentas of patients with severe preeclampsia, compared to those  
256 with normal pregnancy<sup>33</sup>.

257

258 To distinguish if our observed birth weight association could be primarily due to *PPARG*'s  
259 function in fetal insulin resistance or placental function, we characterised rare variants in  
260 *PPARG* using Missense InTerpretation by Experimental Response (MITER) score data<sup>34</sup>, as  
261 the gold-standard indicator of lipodystrophy-causing *PPARG* missense variants. Although  
262 MITER scores are not available for PTV variants, for which we observed the strongest birth  
263 weight association, we note that 8 out of our included 9 PTV variants are classified as  
264 pathogenic based on ACGS<sup>35</sup>. Missense variants with a lipodystrophy-level MITER score  
265 ( $\leq -2$ ) were collectively associated with a decrease in adult body fat percentage of similar  
266 magnitude to other prediction-based missense variant collapsing masks ( $p = 9.70 \times 10^{-5}$ , **Figure**  
267 **4a, Supplementary Table 11**). We found only weak evidence of an association between the  
268 MITER score variants and birth weight ( $p = 0.06$ ), although the effect estimate on birth weight  
269 was not significantly different from the other missense masks (**Figure 4a, Supplementary**  
270 **Table 11**). There was also no discernible correlation pattern between *PPARG* variant effect  
271 estimates on birth weight, body fat percentage and their corresponding MITER scores  
272 (Pearson correlations were,  $R_{\text{BW-MITER}} = 0.026$ ,  $R_{\text{BF\%-MITER}} = 0.187$ ,  $R_{\text{BF\%-BW}} = 0.101$ , **Figure 4c**).  
273 Taken together, our results suggest that the effects of rare *PPARG* PTVs on birth weight  
274 could be due to altered placentation, as well as the established role in adipogenesis.

275

276



277

278 **Figure 4 | Associations with rare LoF variants in *PPARG*.** (a) Different (fetal) variant  
 279 collapsing masks and their associations with birth weight (top) and adult body fat percentage  
 280 (bottom). (b) The genomic location and associations with birth weight for all qualifying fetal  
 281 high confidence PTVs within *PPARG*, in the discovery analysis. (c) Scatterplot of association  
 282 P-values for birth weight (fetal variants) and body fat percentage, for all rare missense  
 283 variants in *PPARG* annotated with a MITER score and found in the UK Biobank population.  
 284 Marker colour indicates the variants MITER score, with scores  $\leq -2$  indicating  
 285 lipodystrophy-level MITER scores. Relevant data is included in **Supplementary Table 11**.

286

287 As with *PPARG*, the *ADAMTS8*, *NRK* and *NOS3* genes have also been implicated in  
 288 placental function. Specifically, *ADAMTS8* shows strong placental expression in the early  
 289 stages of gestation and has been implicated in endometrium decidualization and trophoblast  
 290 invasion<sup>36-39</sup>. *NRK* encodes a serine/threonine kinase of the germinal centre kinase family  
 291 and is expressed in skeletal muscle during development as well as the placenta, specifically  
 292 the spongiotrophoblast layer, a fetus-derived region of the placenta<sup>40-42</sup>. *Nrk*-deficient mouse  
 293 fetuses display enlarged placentas, a likely result of enhanced trophoblast proliferation due to  
 294 upregulation of AKT phosphorylation<sup>43</sup>, as well as a higher birth weight and delayed delivery  
 295 through yet unknown mechanisms<sup>42</sup>. Finally, common variants in *NOS3* have been previously  
 296 associated with the risk of developing preeclampsia<sup>44</sup>, although no association was found at  
 297 this locus in a large-scale preeclampsia GWAS<sup>45</sup>.

## 298 Comparison of burden test results with birth weight effects of pathogenic 299 GCK variants

300 To investigate the sensitivity of our ExWAS approach to detect known rare variant effects on  
301 birth weight, we focused on the *GCK* gene. In the UK, pathogenic variants in *GCK* are the  
302 second most common cause of monogenic diabetes, and have established effects on fetal  
303 growth<sup>4,46</sup>. Mothers with pathogenic *GCK* variants have a stably raised fasting glucose level,  
304 which results in greater glucose availability to the fetus, and increased birth weight due to  
305 fetal insulin secretion, which is raised in response to higher maternal and consequently fetal  
306 glucose. Conversely, fetal growth is reduced in babies who carry pathogenic *GCK* variants  
307 due to their reduced insulin secretion. These birth weight effects are masked if both mother  
308 and fetus carry the same mutation<sup>4</sup>. We annotated and combined pathogenic variants in *GCK*  
309 using clinical guidelines<sup>35</sup> and observed evidence of both fetal and maternal genetic  
310 associations with birth weight in the expected directions ( $\beta_{\text{fetal}} = -0.311$  SDs,  $p = 0.002$ ,  $N = 86$   
311 carriers;  $\beta_{\text{maternal}} = 0.430$  SDs,  $p = 0.001$ ,  $N = 53$  carriers; **Supplementary Figure 8;**  
312 **Supplementary Table 12**). In comparison, our gene burden associations with rare predicted  
313 LoF variants in *GCK* were more modest but in the expected directions (**Supplementary**  
314 **Figure 8; Supplementary Table 12**). We found 63 of the known pathogenic *GCK* variants  
315 present in the UK Biobank WES dataset. Out of 16 variants in the PTV mask, 14 were  
316 annotated as pathogenic and of 55 variants in the Missense + PTV mask, 33 were annotated  
317 as pathogenic. The remaining pathogenic variants were annotated as non-deleterious and  
318 were not included in the masks.

## 319 Discussion

320 In this study, we used whole exome sequencing data from the UK Biobank cohort to  
321 understand the influence of rare genetic variants on human birth weight. We aimed to identify  
322 genes harbouring rare variants that impact birth weight when carried by either the fetus or the  
323 mother. We identified a total of 8 genes with rare fetal LoF variant effects on birth weight, two  
324 of which, *IGF1R* and *NOS3* showed some evidence of maternal effects. We identified one  
325 additional maternal effect gene, *ADAMTS8*. Findings were largely consistent in an  
326 independent cohort of Icelandic genomes, and four of the identified genes harboured  
327 independent GWAS lead SNPs in the largest and most recent birth weight GWAS  
328 meta-analysis, providing additional confirmation that the genes identified in this study are  
329 involved in birth weight regulation.

330  
331 Among the genes associated with birth weight, several have well-established links with fetal  
332 growth. *IGF1R* and *PAPPA2* encode key components of IGF bioavailability and signalling.  
333 *IGF1R* is a transmembrane receptor tyrosine kinase activated by IGF1 (and also IGF2 and  
334 insulin), which mediates anabolic effects. Rare human homozygous, heterozygous or  
335 compound heterozygous mutations in *IGF1R* cause intrauterine growth restriction, reduced  
336 postnatal growth, short stature and microcephaly<sup>47</sup>. *PAPPA2* encodes a metalloproteinase  
337 that regulates the bioavailability of IGFs. Rare homozygous mutations in *PAPPA2* are  
338 reported to cause severe short stature but with unclear effects on birth weight (reduced in 4

339 of 6 affected children)<sup>48</sup>, while the protein originates primarily from trophoblasts in the early  
340 placenta<sup>49,50</sup>. Our findings demonstrate that rare, damaging variants in *IGF1R* and *PAPPA2*  
341 contribute to population variation in birth weight, as well as in childhood and adult height.

342

343 Consistent with the fetal insulin hypothesis<sup>51</sup>, the association between the rare, damaging  
344 fetal variants in *IGF1R* and greater type 2 diabetes risk provides a potential mechanistic  
345 insight into the well-documented link between this later-life disease and lower birth weight.  
346 *IGF1R* is central to the development of several key tissues for glucose metabolism, including  
347 pancreatic islets, adipose tissue and skeletal muscle<sup>52</sup>, and in addition, loss of *IGF1R*  
348 function may lead to compensatory increases in growth hormone levels, which are  
349 associated with insulin resistance<sup>53</sup>. The mechanism underlying the association we observed  
350 between maternal rare, damaging *IGF1R* variants and lower offspring birth weight after  
351 adjustment for fetal genotype is, however, less clear because we would expect such variants  
352 to increase birth weight if they result in higher maternal insulin resistance with concomitant  
353 higher glucose availability. However, there are some reported cases of heterozygous *IGF1R*  
354 mutations leading to hypoglycemia<sup>54</sup>. We recommend that this result is confirmed in  
355 well-powered samples of mother-child pairs before potential mechanisms are investigated.

356

357 *PPARG*, *INHBE*, and *ACVR1C* displayed fetal effects on birthweight and are involved in  
358 adipose tissue differentiation and regulation. Recent genetics studies implicated rare LoF  
359 variants in *INHBE* in adult favourable adiposity traits, i.e. lower WHR adjusted for BMI<sup>20,21</sup>.  
360 They also highlighted a splice variant (12:57456093:G:C) which affects either protein  
361 secretion or stability, as underpinning the favourable adiposity association with this gene. We  
362 found that the birth weight association with *INHBE* largely remains, even in the absence of  
363 this splice variant. This could indicate that some variants within *INHBE* may exert a role in  
364 early (*in utero*) adipogenesis, while others may do so throughout the life course, but further  
365 studies need to confirm this hypothesis.

366

367 Our identification of rare LoF variants in two genes, *ACVR1C* and *INHBE*, that have fetal  
368 associations with higher birth weight and also a more metabolically favourable body fat  
369 distribution in adulthood is in line with recent findings that common genetic variants in the  
370 fetus, which predispose to higher metabolically favourable adiposity in adulthood, are also  
371 associated with higher birth weight. Variants with greater effects on adiposity have greater  
372 effects on birth weight<sup>55</sup> and we therefore hypothesise that the birth weight effects of  
373 *ACVR1C* and *INHBE* are due primarily to fetal fat accretion rather than effects on lean mass.  
374 A fetal genetic predisposition to greater birth weight and later favourable body fat distribution  
375 may underlie epidemiological associations between greater skinfold thickness at birth and  
376 favourable metabolic outcomes in early childhood, especially when controlling for exposure to  
377 maternal glucose levels *in utero*<sup>56</sup>. Further studies should assess the contributions of rare  
378 variation in *ACVR1C* and *INHBE* to the fat and lean mass components of birth weight.

379

380 More generally, the effects of loss of *INHBE* on *Alk7* (*ACVR1C*) signalling on metabolic health  
381 appear to be complex and dose-dependent. Humans haploinsufficient for either of these  
382 genes have a healthy metabolic profile and reduced risk of cardiometabolic disease

383 accompanied by a favourable distribution of body fat. However, it has been shown that the  
384 total absence of either INHBE or ACVR1C from conception results in uncontrolled adipocyte  
385 lipolysis, ectopic fat distribution and insulin resistance, at least in mice<sup>22,23</sup>. This discrepancy  
386 may be explained by the fact that Alk7 signalling in adipose cells suppresses the expression  
387 of both PPAR $\gamma$  and key molecular elements of the lipolytic machinery. We suggest that at  
388 modest levels of INHBE/Alk7 deficiency, such as occurs when only one functional allele is  
389 present, the effect on PPAR $\gamma$  is dominant and carriers of such mutations develop more  
390 subcutaneous adipocytes as a result of the enhanced PPARG tone during development.  
391 Ultimately, this is metabolically beneficial as it will tend to protect the individual from any  
392 effects of any caloric overload in postnatal life. However, in the total absence of INHBE/Alk7  
393 the effect on lipolysis becomes dominant with consequent build-up of ectopic fat and insulin  
394 resistance. While it is not entirely clear why subcutaneous, rather than visceral adipocytes  
395 would be favoured by PPAR $\gamma$  activity, it is worthy of note that the administration of drugs  
396 activating PPAR $\gamma$  preferentially promotes subcutaneous fat deposition<sup>57</sup>.

397  
398 Heterozygous LoF in *PPARG* has previously been linked to familial partial lipodystrophy,  
399 characterised by the loss of subcutaneous fat from the extremities and metabolic  
400 abnormalities such as insulin resistance and hypertriglyceridemia<sup>24–27</sup>. However, *PPARG* has  
401 also been linked to placental development and angiogenesis<sup>30</sup>. Here, we saw a poor  
402 correlation between the variants' lipodystrophy-causing potential and effect on birth weight,  
403 suggesting the birth weight effect could be driven by another mechanism. Given the role of  
404 *PPARG* in placental development, it is possible that *PPARG* LoF variants act via the  
405 placenta to exert an effect on birth weight. Our results are consistent with recent observations  
406 of a high rate of small-for-gestational-age birth weights in patients who had familial partial  
407 lipodystrophy due to inherited *PPARG* mutations and who were not exposed to maternal  
408 diabetes in utero<sup>28</sup>. *PPARG* mutation carriers were also more likely to be born preterm,  
409 regardless of maternal diabetes status, which further supports the role of fetal *PPARG* in  
410 placental development. Data on gestational duration was not available for the UK Biobank  
411 participants, and it is possible that shorter gestation in rare *PPARG* variant carriers  
412 contributed to the associations we observed with birth weight. We note that the common fetal  
413 variant rs1801282 in *PPARG*, which is associated with type 2 diabetes risk in adulthood<sup>58</sup> via  
414 a fat distribution-mediated form of insulin resistance<sup>59</sup>, is weakly associated with lower birth  
415 weight ( $p=0.0259$ )<sup>5</sup> and was not associated with gestational duration in a fetal GWAS of  
416 gestational duration in 84,689 individuals ( $p=0.487$ )<sup>60</sup>. Our observed associations between  
417 birthweight and *ADAMTS8*, *NRK* and *NOS3* are also likely driven by placental mechanisms,  
418 given the known roles of these genes<sup>36–44</sup>. The potential mechanism underlying the association  
419 between birth weight and our other identified gene, *NYNRIN*, requires further study.

420  
421 Our study has several strengths, including a large discovery sample, an independent  
422 replication sample, support from common variant associations proximal to five of the genes,  
423 and approximate adjustment for correlation between maternal and fetal genotype. However,  
424 there are limitations. The sample size available for replication was limited, and although we  
425 observed strong evidence for consistent effect estimates, we recommend confirmation in  
426 larger datasets as they become available, along with proper conditional analyses in

427 sufficiently-powered samples of mother-child pairs. The approximate adjustment for the  
428 correlation between maternal and fetal genotype used here is based on the strong  
429 assumption that all variants with a given annotation have equal effects on birth weight<sup>5</sup>, which  
430 may not be the case. We also acknowledge the limited diversity of our samples. Studies of  
431 rare variants in well-powered samples of diverse ancestries are a priority for future research.  
432 While our study was underpowered to detect very rare pathogenic variants that have known  
433 birth weight effects<sup>2</sup>, our sensitivity analysis of *GCK* showed that the known maternal and  
434 fetal associations were detectable by our method, despite the inevitable lower sensitivity and  
435 specificity of the exome-wide approach, and despite the lack of adjustment for opposing  
436 effects of maternal and fetal genotypes.

437

438 Overall, this study advances our understanding of the fetal and maternal genetic  
439 underpinnings of birth weight, providing strong evidence for causal genes and insights into  
440 biological pathways which are important targets for future research aiming to understand fetal  
441 growth and its links with long-term health.

## 442 Acknowledgements

443 This research has been conducted using the UK Biobank Resource under Application  
444 Numbers 9905 and 7036. UK Biobank data has approval from the North West Multi-centre  
445 Research Ethics Committee (MREC) as a Research Tissue Bank (RTB).

446 Use of the Icelandic data was approved by the National Bioethics Committee (VSN-15-169).  
447 All genotyped participants signed a written informed consent allowing the use of their  
448 samples and data in projects at deCODE genetics approved by the NBC. Data were  
449 anonymized and encrypted by a third-party system, approved and monitored by the Icelandic  
450 Data Protection Authority<sup>61</sup>.

451 B.E.M.L., R.N.B. and R.M.F. were supported by a Wellcome Senior Research Fellowship  
452 (WT220390). This research was funded in part by the Wellcome Trust (WT220390). R.M.F. is  
453 also supported by a grant from the Eunice Kennedy Shriver National Institute Of Child Health  
454 & Human Development of the National Institutes of Health under Award Number  
455 R01HD101669. K.A.K., L.R.K., E.J.G., Y.Z., F.R.D., K.K.O. and J.R.B.P. are supported by the  
456 Medical Research Council (MC\_UU\_00006/2). This study was supported by the National  
457 Institute for Health and Care Research (NIHR) Exeter Biomedical Research Centre and the  
458 NIHR Cambridge Comprehensive Biomedical Research Centre. The views expressed are  
459 those of the authors and not necessarily those of the NIHR or the Department of Health and  
460 Social Care. For the purpose of open access, the authors have applied a CC BY public  
461 copyright licence to any Author Accepted Manuscript version arising from this submission.

462

## 463 Author contributions

464 K.A.K., B.E.M.L., L.R.K., K.K.O., R.N.B., J.R.B.P., and R.M.F. analysed and interpreted the  
465 original data and wrote the manuscript. V.S., V.T., T.O., G.T., and K.S. contributed the  
466 Icelandic replication data. L.S., K.A.P., G.H., E.J.G., A.R.W., Y.Z., F.R.D., S.E.O., A.T.H., and  
467 S.O. contributed to data analysis and/or interpretation. All authors revised, critically reviewed  
468 and approved the manuscript.

469

## 470 Competing interests

471 J.R.B.P. and E.J.G. are employees/shareholders of Insmid. J.R.B.P. also receives research  
472 funding from GSK. Y.Z. is a UK University worker at GSK. S.O. has undertaken remunerated  
473 consultancy work for Pfizer, Third Rock Ventures, AstraZeneca, NorthSea Therapeutics and  
474 Courage Therapeutics. V.S., V.T., T.O., G.T., and K.S. are employees of deCODE genetics, a  
475 subsidiary of Amgen.

476

## 477 Code availability

478 Code for WES data processing and association testing is available on GitHub  
479 (<https://github.com/mrcepid-rap/mrcepid-runassociationtesting>). No custom code was  
480 developed for this study.

## 481 Data availability

482 UK Biobank data are available on application via  
483 <https://www.ukbiobank.ac.uk/enable-your-research/register>.

## 484 Materials and methods

### 485 Exome wide association analyses in the UK Biobank

486 Analyses of the UK Biobank data were performed on the UK Biobank Research Analysis  
487 Platform (RAP; <https://ukbiobank.dnanexus.com/>). Birth weight was derived in the 453,505  
488 UK Biobank participants with genetically defined European ancestry, as follows. Own birth  
489 weight (fetal genotype analysis) was derived using self-reported data (field 20022), while birth  
490 weight of the first child in UK Biobank women (maternal genotype analysis) was taken from  
491 hospital records and self-reports (fields 41284 and 2744). Data from each field were  
492 converted to kg (if in another unit) and if both were available, hospital records were  
493 preferentially retained for maximum accuracy. For both phenotypes, repeated responses  
494 across the assessment centre visits were combined by calculating average values.  
495 Individuals from multiple births (field 1777), who reported substantially different values  
496 between visits (>1kg difference), or extreme birth weight values (<1 or ≥7kg) were excluded  
497 from downstream analyses. After these exclusions, fetal birth weight data was available for  
498 252,329 individuals (234,675 with exome and covariate data) and maternal birth weight data  
499 was available for the first child of 195,653 women (181,883 with exome and covariate data).

500

501 We conducted gene burden tests for both fetal and maternal birth weight, by combining  
502 effects of all rare (MAF <0.1%) variants with predicted deleterious functional consequences  
503 across all protein-coding genes, as described in detail in Gardner *et al.*<sup>53</sup>. Briefly, we queried  
504 population-level VCF files with data for 454,787 individuals provided from the UK Biobank  
505 study. Using bcftools<sup>62</sup> multi-allelic variants were split and left-normalised, and all variants  
506 filtered using a missingness based approach. SNV genotypes with depth <7 and genotype  
507 quality <20 or InDel genotypes with a depth <10 and genotype quality <20 were set to  
508 missing. We also tested for an expected reference and alternate allele balance of 50% for  
509 heterozygous SNVs using a binomial test; SNV genotypes with a binomial test p. value  
510  $\leq 1 \times 10^{-3}$  were set to missing. Following genotype filtering, variants with >50% missing  
511 genotypes were excluded from further analysis. Variants were then annotated with the  
512 ENSEMBL Variant Effect Predictor (VEP<sup>63</sup>, v104) with the 'everything' flag and the LOFTEE  
513 plugin<sup>64</sup>. For each variant we prioritised a single MANE (v0.97) or VEP canonical ENSEMBL  
514 transcript and most damaging consequence as defined by VEP defaults. For the purposes of  
515 defining Protein Truncating Variants (PTVs), we grouped high-confidence (as defined by  
516 LOFTEE) stop gained, splice donor/acceptor, and frameshift consequences. To define  
517 'high-confidence' PTVs, we used the LOFTEE algorithm<sup>64</sup>. All variants were subsequently  
518 annotated using CADD (v1.6, <sup>65</sup>). We then assessed the association between rare variant  
519 burden and birth weight, using BOLT-LMM (v2.3.5, <sup>66</sup>) and a set of dummy genotypes  
520 representing per-gene carrier status, under two overlapping variant collapsing masks; i)



521 high-confidence protein truncating variants (denoted PTV) and ii) High-confidence PTVs plus  
522 missense variants with CADD scores  $\geq 25$  (denoted Missense + PTV). All analyses were  
523 controlled for age, age<sup>2</sup>, the first ten genetic ancestry principal components as calculated in  
524 Bycroft *et al.*<sup>67</sup>, WES batch, and sex when running sex-combined analyses. We excluded  
525 genes with  $< 10$  carriers of variants per mask.

526

527 This led to testing 16,735 genes with qualifying Missense + PTV variants and 13,684 with  
528 PTVs for fetal birth weight, and 17,745 with Missense + PTV variants and 13,968 with PTVs  
529 for maternal birth weight. The exome-wide significance thresholds were thus set to  
530  $p = 1.64 \times 10^{-6}$  ( $0.05 / (16735 + 13684)$ ) for fetal birth weight and  $p = 1.58 \times 10^{-6}$   
531 ( $0.05 / (17745 + 13968)$ ) for maternal birth weight.

### 532 Confirmatory gene burden analyses

533 To replicate the findings from the above analysis and account for potential bias arising from  
534 only using one discovery approach, a second team independently derived the birth weight  
535 phenotypes for the maternal and fetal genetic analyses and conducted additional burden  
536 association analyses, specifically for the significantly associated birth weight genes.

537 Own birth weight (fetal genotype analysis) was derived using self-reported data (field 20022),  
538 excluding individuals who were part of a multiple birth (field 1777). Offspring birth weight  
539 (maternal genotype analysis) was derived from self reported weight of first child (field 2744).  
540 Where individuals reported their birth weight at more than one assessment centre visit,  
541 reported birth weights were averaged unless the weights differed by  $\geq 1$  kg, in which case  
542 they were excluded. Own and offspring birth weight were available in 269,921 (232,876 with  
543 exome and covariate data) and 216,798 (152,585 with exome and covariate data) individuals  
544 respectively.

545

546 Multiallelic variants were similarly split and left-normalised. Variants flagged for exclusion by  
547 UK Biobank were removed, and remaining variants were annotated with their functional  
548 consequence in canonical transcripts using the Ensembl Variant Effect Predictor and the  
549 CADD and LOFTEE plugins (<https://github.com/konradjk/loftee>). Association testing was  
550 performed using REGENIE (v3.1.3,<sup>68</sup>), and variants were grouped according to two  
551 collapsing masks; i) loss of function variants defined by LOFTEE as high confidence  
552 (denoted PTV) and ii) LoF plus missense SNVs or inframe insertions or deletions with CADD  
553 scores  $\geq 25$  (denoted Missense + PTV). All analyses were controlled for maternal age at first  
554 birth (field 3872) or year of birth, sex (for fetal birth weight), assessment centre, WES  
555 sequencing batch, and the first ten principal components.

556

557 Exome-wide significantly associated genes from the above BOLT-LMM analysis were queried  
558 in the REGENIE results.

## 559 Follow-up analyses of identified genes

560 To distinguish between fetal genotyping-acting ('fetal-acting') and maternal genotype-acting  
561 ('maternal-acting') effects at each birth weight-associated gene, we applied a weighted linear  
562 model (WLM) to the burden test results. WLM estimates associations of fetal genotype  
563 conditional on maternal genotype, and vice versa, without the need for data from  
564 mother-child pairs. The WLM approach has been described previously<sup>5</sup>.

565

566 To investigate potential sexual dimorphism, we performed sex-stratified burden tests using  
567 generalised linear models in the discovery sample and for the most strongly associated  
568 mask-gene combination. Sex-stratified effect estimates for each gene were compared using  
569 two-sample T-tests. Associations were deemed dimorphic if the  $p$ -value for the T-test was  
570  $<0.05$ .

571

572 Where described, further missense variant collapsing masks and leave-one out analyses  
573 were tested using BOLT-LMM or linear models in the same individuals and birth weight  
574 outcomes as the discovery analyses. Such masks were defined using a combination of  
575 CADD<sup>65</sup> and REVEL<sup>69</sup>.

576

577 Conditional analyses where the effect of the identified rare exome variant collapsing masks  
578 was conditioned on the genotype at common birth weight GWAS signals were also tested  
579 using linear models in the same individuals and birth weight outcomes as the discovery  
580 analyses. Genotypes at GWAS signals (as identified in <sup>6</sup>) were extracted in this discovery  
581 sample using plink (v1.90b6.18, <sup>70</sup>). Birthweight outcomes were regressed against the  
582 covariates used in discovery analyses and the concurrent effects of exome variant carriage  
583 and GWAS signal genotype were tested against the residual birth weight variance. A change  
584 in effect estimate of more than 10% was considered to show significant attenuation in the  
585 presence of the GWAS variants.

## 586 Related trait PheWAS

587 For all identified exome associations to the two birth weight traits, we performed small-scale  
588 PheWAS on a few predefined relevant phenotypes.

589

590 For BOLT-LMM analyses, phenotypic outcomes were defined as follows. BMI raw values  
591 were used from field 21001. Body fat % was extracted using data from field 23099. Data from  
592 subsequent visits were used, if missing for a given instance. Comparative height at age 10  
593 was recoded to have individuals who in field 1697 reported being 'Shorter' as 0, 'About  
594 average' as 1, and 'Taller' as 2. Individuals who reported 'Do not know' or 'Prefer not to  
595 answer' were set to 'NA'. Following variable recoding, this phenotype was run as a  
596 continuous trait. Comparative size at age 10 was defined as above using data from field  
597 1687. Systolic and diastolic blood pressure was derived using data from fields 4079 and  
598 4080, accordingly. For each, the average blood pressure at each assessment visit was  
599 calculated over the two available measurements and data from subsequent visits were used,  
600 if missing for a given instance. Individuals indicating the use of blood pressure medications in

601 fields 6153 and 6177 were adjusted for in the association models. Height raw values were  
602 used from field 50. IGF-1 levels were derived using data from field 30770, with exclusion of  
603 individuals >5 standard deviations from the mean. T2D was derived as described in<sup>53</sup> and  
604 using data from fields 4041, 10844, 2443, 6177, 6153, 20002, 20003, 41202, 41204,  
605 40001 and 40002. WHR adjusted for BMI; WHR was calculated using data from field 48 and  
606 49. >4 standard deviations from the mean excluded following initial calculation and prior to  
607 adjusting with BMI (field 21001). Paired BMI and WHR data from subsequent visits were  
608 used, if missing for a given instance.

609

610 For REGENIE analyses, phenotypic outcomes are as follows: BMI was taken from the  
611 baseline value for field 21001 and residualised on age at assessment (field 21003), genetic  
612 sex (field 22001), assessment centre (field 54) and PCs 1-5. Blood pressure measurements  
613 were taken from the automated readings for diastolic blood pressure (field 4079) and systolic  
614 blood pressure (field 4080). Where two blood pressure readings were available, the average  
615 between the two was taken, excluding individuals where the measurements differed by >4.56  
616 SD. Individuals with a blood pressure >4.56 SD away from the mean were excluded. Blood  
617 pressure medication was accounted for by adding 10 and 15 mmHg to diastolic and systolic  
618 blood pressure, respectively. Height raw values were taken from baseline values for field 50,  
619 where people whose height was 100 cm and sitting height (field 20015) was 20 cm, were  
620 removed. Raw IGF-1 values were used from field 30770. Size at age 10 was based on a  
621 continuous simulation of childhood BMI. The simulation converted the categorical variables  
622 'thinner', 'the same size' or 'plumper' in field 1687 to a continuous distribution based on the  
623 number of individuals selecting each category. The full methods for the simulation have  
624 previously been reported<sup>71</sup>. WHR was calculated using waist circumference (field 48) and hip  
625 circumference (field 49). If a follow-up visit was recorded, the WHR was taken from the  
626 follow-up data. Otherwise, it was taken from the first recorded WHR measurement.

627

628 The multiple testing threshold was set at  $p < 0.005$ , after correcting for the above 10 queried  
629 phenotypes.

## 630 GWAS associations at identified loci

631 GWAS signals for fetal and maternal birth weight were accessed from Juliusdottir *et al.*<sup>6</sup>.  
632 These signals were lifted back to GRCh37 annotated with all proximal genes, defined as  
633 those within 500kb up- or downstream of the genes start or end sites, using the NCBI RefSeq  
634 gene map for GRCh37 (via <http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/>). For  
635 signals proximal to the exome-associated genes we assessed whether the exome-associated  
636 gene was the closest gene to the GWAS signal. These signals were also queried using eQTL  
637 data from the GTEx study (V8,<sup>10</sup>) and plotted using LocusZoom<sup>72</sup>.

638

639 We performed gene-centric biological pathway enrichment analysis using g:Profiler (via the R  
640 client "gprofiler2"<sup>73</sup>, v0.2.1, accessed on 23/11/2023). Pathway enrichment analyses were  
641 performed using the genes calculated to be the closest per signal to the GWAS signals from  
642 Juliusdottir *et al.*<sup>6</sup>. Pathways with a Bonferroni corrected  $p$ -value < 0.05 were considered  
643 significantly enriched in the birth weight GWAS.

644

645 We also examined all genes within 300 kb of the previously published birth weight loci from  
646 Juliusdottir *et al.*<sup>6</sup>. For this restricted set of loci previously shown to have common genetic  
647 associations with birth weight, we applied a Benjamini-Hochberg procedure to identify genes  
648 with evidence of association.

## 649 Replication of exome associations in deCODE

### 650 Data preparation

651 The genome of the Icelandic population was characterised by whole-genome sequencing  
652 (WGS) of 63,460 Icelanders using Illumina standard TruSeq methodology to a mean depth of  
653 35x (SD 8x) with subsequent long-range phasing<sup>74</sup>, and imputing the information into 173,025  
654 individuals chip-genotyped employing multiple Illumina platforms<sup>75</sup>. Variant calling was  
655 performed using GraphTyper<sup>76</sup> (v2.6) and chip data was phased using SHAPEIT4<sup>77</sup>.

656 We used Variant Effect Predictor (VEP<sup>63</sup>) to attribute predicted consequences to the variants  
657 sequenced in each dataset. We classified as high-impact variants those predicted as  
658 start-lost, stop-gain, stop-lost, splice donor, splice acceptor or frameshift, collectively called  
659 loss-of-function (LOF) variants. Variants of moderate impact are those classified as  
660 missense, splice-region and in-frame-indels.

661 Information on birthweight comes from the Icelandic birth register and Primary Health Care  
662 Clinics of the Capital area. Individual's fetal birth weight was available in up to 36,578  
663 genotyped and 11,184 WGS participants, while maternal birth weight was available in 45,622  
664 genotyped and 17,177 WGS participants.

### 665 Gene burden associations

666 We defined different models to group together various types of variant: (i) only LOF variants,  
667 filtered with LOFTEE<sup>64</sup>; (ii) LOF and MIS (as described on Variant annotation), when  
668 predicted deleterious by CADD<sup>65</sup> score  $\geq 25$ . In all models, we used MAF  $< 0.1\%$  to select  
669 variants for analyses.

670 Quantitative traits were analysed using a linear mixed model implemented in BOLT-LMM<sup>66</sup>. To  
671 estimate the quality of the sequence variants across the entire set we regressed the  
672 alternative allele counts (AD) on the depth (DP) conditioned on the genotypes (GT) reported  
673 by GraphTyper<sup>78</sup>. For a well behaving sequence variant, the mean alternative allele count for  
674 a homozygous reference genotype should be 0, for a heterozygous genotype it should be  
675 DP/2 and for homozygous alternative genotype it should be DP. Under the assumption of no  
676 sequencing or genotyping error, the expected value of AD should be DP conditioned on the  
677 genotype, in other words an identity line (slope 1 and intercept 0). Deviations from the identity  
678 line indicate that the sequence variant is spurious or somatic. We filter variants with slope  
679 less than 0.5. Additionally, GraphTyper employs a logistic regression model that assigns each  
680 variant a score (AAscore) predicting the probability that it is a true positive. We used only  
681 variants that have a AAscore  $> 0.8$ .

## 682 Analysis of birth weight associations with pathogenic *GCK* mutations

683 We used a similar procedure to that used to define pathogenic variants in *GCK* in a recent  
684 paper<sup>79</sup>. We reviewed all heterozygous missense/PTV variants in UK Biobank participants  
685 with whole exome sequencing that were observed at minor allele frequency (MAF) <0.001.  
686 We included variants in the analysis if missense/PTV variants were classified as pathogenic  
687 or likely pathogenic based on ACMG/AMP guidelines by clinical scientists at Exeter  
688 Molecular Genetic laboratory as part of routine clinical diagnostic care (i.e. previously seen in  
689 the MODY probands)<sup>35</sup> and were ultra-rare in the population (maximum allele count of 2 in  
690 gnomADv2.1, MAF <1.4x10<sup>-5</sup>). We manually reviewed sequence read data for all the  
691 pathogenic variants (missense and PTVs) in Integrative Genomics Viewer (IGV) to remove  
692 false-positive variants. We used linear regression to assess the difference in birth weight  
693 associated with being a carrier (separate maternal and fetal analyses) and adjusted for  
694 genotype, sex, mother's diabetes status, batch, year of birth, centre and principal  
695 components in the fetal analysis and genotype, batch, age at first birth, centre and principal  
696 components in the maternal birth weight analysis. We then compared the estimated effects  
697 and 95% confidence intervals for the pathogenic variants with those obtained from our PTV  
698 and PTV+missense masks in the REGENIE analyses.

## 699 References

700

- 701 1. Barker, D. J. *et al.* Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and  
702 hyperlipidaemia (syndrome X): relation to reduced fetal growth. *Diabetologia* **36**, 62–67  
703 (1993).
- 704 2. Hughes, A. E., Hattersley, A. T., Flanagan, S. E. & Freathy, R. M. Two decades since the  
705 fetal insulin hypothesis: what have we learned from genetics? *Diabetologia* **64**, 717–726  
706 (2021).
- 707 3. Hughes, A. E. *et al.* Monogenic disease analysis establishes that fetal insulin accounts  
708 for half of human fetal growth. *J. Clin. Invest.* **133**, e165402 (2023).
- 709 4. Hattersley, A. T. *et al.* Mutations in the glucokinase gene of the fetus result in reduced  
710 birth weight. *Nat. Genet.* **19**, 268–270 (1998).
- 711 5. Warrington, N. M. *et al.* Maternal and fetal genetic effects on birth weight and their  
712 relevance to cardio-metabolic risk factors. *Nat. Genet.* **51**, 804–814 (2019).
- 713 6. Juliusdottir, T. *et al.* Distinction between the effects of parental and fetal genomes on  
714 fetal growth. *Nat. Genet.* **53**, 1135–1142 (2021).
- 715 7. Hughes, A. E. *et al.* Fetal Genotype and Maternal Glucose Have Independent and  
716 Additive Effects on Birth Weight. *Diabetes* **67**, 1024–1029 (2018).
- 717 8. Aragam, K. G. *et al.* Discovery and systematic characterization of risk variants and  
718 genes for coronary artery disease in over a million participants. *Nat. Genet.* **54**, 1803–1815  
719 (2022).
- 720 9. Mountjoy, E. *et al.* An open approach to systematically prioritize causal variants and  
721 genes at all published human GWAS trait-associated loci. *Nat. Genet.* **53**, 1527–1533 (2021).
- 722 10. Lonsdale, J. *et al.* The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45**, 580  
723 (2013).
- 724 11. Sferruzzi-Perri, A. N., Sandovici, I., Constancia, M. & Fowden, A. L. Placental phenotype

- 725 and the insulin-like growth factors: resource allocation to fetal growth. *J. Physiol.* **595**,  
726 5057–5093 (2017).
- 727 12. Conover, C. A. *et al.* Metalloproteinase pregnancy-associated plasma protein A is a  
728 critical growth regulatory factor during fetal development. *Dev. Camb. Engl.* **131**, 1187–1194  
729 (2004).
- 730 13. Bøtkjær, J. A., Noer, P. R., Oxvig, C. & Yding Andersen, C. A common variant of the  
731 pregnancy-associated plasma protein-A (PAPPA) gene encodes a protein with reduced  
732 proteolytic activity towards IGF-binding proteins. *Sci. Rep.* **9**, 13231 (2019).
- 733 14. Yan, X., Baxter, R. C. & Firth, S. M. Involvement of pregnancy-associated plasma  
734 protein-A2 in insulin-like growth factor (IGF) binding protein-5 proteolysis during pregnancy: a  
735 potential mechanism for increasing IGF bioavailability. *J. Clin. Endocrinol. Metab.* **95**,  
736 1412–1420 (2010).
- 737 15. Dauber, A. *et al.* Mutations in pregnancy-associated plasma protein A2 cause short  
738 stature due to low IGF-I availability. *EMBO Mol. Med.* **8**, 363–374 (2016).
- 739 16. Ibáñez, C. F. Regulation of metabolic homeostasis by the TGF- $\beta$  superfamily receptor  
740 ALK7. *FEBS J.* **289**, 5776–5797 (2022).
- 741 17. Justice, A. E. *et al.* Protein-coding variants implicate novel genes related to lipid  
742 homeostasis contributing to body-fat distribution. *Nat. Genet.* **1** (2019).
- 743 18. Emdin, C. A. *et al.* DNA Sequence Variation in ACVR1C Encoding the Activin  
744 Receptor-Like Kinase 7 Influences Body Fat Distribution and Protects Against Type 2  
745 Diabetes. *Diabetes* **68**, 226–234 (2019).
- 746 19. Lau, A. L., Kumar, T. R., Nishimori, K., Bonadio, J. & Matzuk, M. M. Activin betaC and  
747 betaE genes are not essential for mouse liver growth, differentiation, and regeneration. *Mol.*  
748 *Cell. Biol.* **20**, 6127–6137 (2000).
- 749 20. Akbari, P. *et al.* Multiancestry exome sequencing reveals INHBE mutations associated  
750 with favorable fat distribution and protection from diabetes. *Nat. Commun.* **13**, 4844 (2022).

- 751 21. Deaton, A. M. *et al.* Rare loss of function variants in the hepatokine gene INHBE protect  
752 from abdominal obesity. *Nat. Commun.* **13**, 4319 (2022).
- 753 22. Griffin, J. D. *et al.* Hepatic Activin E mediates liver-adipose inter-organ communication,  
754 suppressing adipose lipolysis in response to elevated serum fatty acids. *Mol. Metab.* **78**,  
755 101830 (2023).
- 756 23. Adam, R. C. *et al.* Activin E-ACVR1C cross talk controls energy storage via suppression  
757 of adipose lipolysis in mice. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2309967120 (2023).
- 758 24. Agarwal, A. K. & Garg, A. A novel heterozygous mutation in peroxisome  
759 proliferator-activated receptor-gamma gene in a patient with familial partial lipodystrophy. *J.*  
760 *Clin. Endocrinol. Metab.* **87**, 408–411 (2002).
- 761 25. Barroso, I. *et al.* Dominant negative mutations in human PPARgamma associated with  
762 severe insulin resistance, diabetes mellitus and hypertension. *Nature* **402**, 880–883 (1999).
- 763 26. Hegele, R. A., Cao, H., Frankowski, C., Mathews, S. T. & Leff, T. PPARG F388L, a  
764 transactivation-deficient mutant, in familial partial lipodystrophy. *Diabetes* **51**, 3586–3590  
765 (2002).
- 766 27. Savage, D. B. *et al.* Human metabolic syndrome resulting from dominant-negative  
767 mutations in the nuclear receptor peroxisome proliferator-activated receptor-gamma.  
768 *Diabetes* **52**, 910–917 (2003).
- 769 28. Gosseume, C. *et al.* Perinatal, metabolic, and reproductive features in PPARG-related  
770 lipodystrophy. *Eur. J. Endocrinol.* **188**, lvad023 (2023).
- 771 29. Fournier, T., Tsatsaris, V., Handschuh, K. & Evain-Brion, D. PPARs and the placenta.  
772 *Placenta* **28**, 65–76 (2007).
- 773 30. Nadra, K. *et al.* PPARgamma in placental angiogenesis. *Endocrinology* **151**, 4969–4981  
774 (2010).
- 775 31. Barak, Y. *et al.* PPAR gamma is required for placental, cardiac, and adipose tissue  
776 development. *Mol. Cell* **4**, 585–595 (1999).



- 777 32. Kubota, N. *et al.* PPAR gamma mediates high-fat diet-induced adipocyte hypertrophy  
778 and insulin resistance. *Mol. Cell* **4**, 597–609 (1999).
- 779 33. Waite, L. L., Louie, R. E. & Taylor, R. N. Circulating Activators of Peroxisome  
780 Proliferator-Activated Receptors Are Reduced in Preeclamptic Pregnancy. *J. Clin. Endocrinol.*  
781 *Metab.* **90**, 620–626 (2005).
- 782 34. Majithia, A. R. *et al.* Prospective functional classification of all possible missense  
783 variants in PPARG. *Nat. Genet.* **48**, 1570–1575 (2016).
- 784 35. Ellard, S. *et al.* ACGS Best Practice Guidelines for Variant Classification in Rare Disease  
785 2020. ACGS (2020).
- 786 36. Wen, J., Zhu, H. & Leung, P. C. K. Gonadal steroids regulate the expression of  
787 aggrecanases in human endometrial stromal cells in vitro. *J. Cell. Mol. Med.* **17**, 1325–1334  
788 (2013).
- 789 37. Namli Kalem, M., Kalem, Z., Bakirarar, B. & Demircan, K. Adamts 1, 4, 5, 8, and 9 in  
790 Early Pregnancies. *Fetal Pediatr. Pathol.* **36**, 387–399 (2017).
- 791 38. Qu, H. & Khalil, R. A. Role of ADAM and ADAMTS disintegrin and metalloproteinases in  
792 normal pregnancy and preeclampsia. *Biochem. Pharmacol.* **206**, 115266 (2022).
- 793 39. Tatar, M., Uslu, S. & Öner, J. Expression of placental growth factor and a disintegrin and  
794 metalloprotease with a thrombospondin type motifs 1-4-8 during the three trimesters of rat  
795 pregnancy at the maternal-fetal interface. *Anat. Histol. Embryol.* **52**, 805–814 (2023).
- 796 40. Kanai-Azuma, M. *et al.* NrK: a murine X-linked NIK (Nck-interacting kinase)-related  
797 kinase gene expressed in skeletal muscle. *Mech. Dev.* **89**, 155–159 (1999).
- 798 41. Nakano, K., Yamauchi, J., Nakagawa, K., Itoh, H. & Kitamura, N. NESK, a member of  
799 the germinal center kinase family that activates the c-Jun N-terminal kinase pathway and is  
800 expressed during the late stages of embryogenesis. *J. Biol. Chem.* **275**, 20533–20539  
801 (2000).
- 802 42. Denda, K. *et al.* NrK, an X-linked protein kinase in the germinal center kinase family, is

- 803 required for placental development and fetoplacental induction of labor. *J. Biol. Chem.* **286**,  
804 28802–28810 (2011).
- 805 43. Morioka, Y., Nam, J.-M. & Ohashi, T. Nik-related kinase regulates trophoblast  
806 proliferation and placental development by modulating AKT phosphorylation. *PLoS One* **12**,  
807 e0171503 (2017).
- 808 44. Zeng, F. *et al.* Associations between nitric oxide synthase 3 gene polymorphisms and  
809 preeclampsia risk: a meta-analysis. *Sci. Rep.* **6**, 23407 (2016).
- 810 45. Honigberg, M. C. *et al.* Polygenic prediction of preeclampsia and gestational  
811 hypertension. *Nat. Med.* **29**, 1540–1549 (2023).
- 812 46. Thanabalasingham, G. & Owen, K. R. Diagnosis and management of maturity onset  
813 diabetes of the young (MODY). *BMJ* **343**, d6044 (2011).
- 814 47. Murray, P. G. & Clayton, P. E. Disorders of Growth Hormone in Childhood. in *Endotext*  
815 (eds. Feingold, K. R. *et al.*) (MDText.com, Inc., South Dartmouth (MA), 2000).
- 816 48. Barrios, V. *et al.* Pregnancy-Associated Plasma Protein (PAPP)-A2 in Physiology and  
817 Disease. *Cells* **10**, 3576 (2021).
- 818 49. Overgaard, M. T. *et al.* Pregnancy-associated Plasma Protein-A2 (PAPP-A2), a Novel  
819 Insulin-like Growth Factor-binding Protein-5 Proteinase. *J. Biol. Chem.* **276**, 21849–21853  
820 (2001).
- 821 50. Tóth, E. *et al.* Decreased Expression of Placental Proteins in Recurrent Pregnancy Loss:  
822 Functional Relevance and Diagnostic Value. *Int. J. Mol. Sci.* **25**, 1865 (2024).
- 823 51. Hattersley, A. T. & Tooke, J. E. The fetal insulin hypothesis: an alternative explanation of  
824 the association of low birthweight with diabetes and vascular disease. *Lancet Lond. Engl.*  
825 **353**, 1789–1792 (1999).
- 826 52. Rother, K. I. & Accili, D. Role of insulin receptors and IGF receptors in growth and  
827 development. *Pediatr. Nephrol.* **14**, 558–561 (2000).
- 828 53. Gardner, E. J. *et al.* Damaging missense variants in IGF1R implicate a role for IGF-1

- 829 resistance in the etiology of type 2 diabetes. *Cell Genomics* **2**, None (2022).
- 830 54. Solomon-Zemler, R. *et al.* A novel heterozygous IGF-1 receptor mutation associated with  
831 hypoglycemia. *Endocr. Connect.* **6**, 395–403 (2017).
- 832 55. Thompson, W. D. *et al.* Fetal alleles predisposing to metabolically favorable adiposity are  
833 associated with higher birth weight. *Hum. Mol. Genet.* **31**, 1762–1775 (2022).
- 834 56. Bianco, M. E. *et al.* Hyperglycemia and Adverse Pregnancy Outcome Follow-Up Study:  
835 newborn anthropometrics and childhood glucose metabolism. *Diabetologia* **64**, 561–570  
836 (2021).
- 837 57. Adams, M. *et al.* Activators of peroxisome proliferator-activated receptor gamma have  
838 depot-specific effects on human preadipocyte differentiation. *J. Clin. Invest.* **100**, 3149–3153  
839 (1997).
- 840 58. Altshuler, D. *et al.* The common PPARgamma Pro12Ala polymorphism is associated with  
841 decreased risk of type 2 diabetes. *Nat. Genet.* **26**, 76–80 (2000).
- 842 59. Udler, M. S. *et al.* Type 2 diabetes genetic loci informed by multi-trait associations point  
843 to disease mechanisms and subtypes: A soft clustering analysis. *PLOS Med.* **15**, e1002654  
844 (2018).
- 845 60. Liu, X. *et al.* Variants in the fetal genome near pro-inflammatory cytokine genes on 2q13  
846 associate with gestational duration. *Nat. Commun.* **10**, 3927 (2019).
- 847 61. Gulcher, J. R., Kristjánsson, K., Gudbjartsson, H. & Stefánsson, K. Protection of privacy  
848 by third-party encryption in genetic research in Iceland. *Eur. J. Hum. Genet. EJHG* **8**,  
849 739–742 (2000).
- 850 62. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping  
851 and population genetical parameter estimation from sequencing data. *Bioinforma. Oxf. Engl.*  
852 **27**, 2987–2993 (2011).
- 853 63. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
- 854 64. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in

- 855 141,456 humans. *Nature* **581**, 434–443 (2020).
- 856 65. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the  
857 deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**,  
858 D886–D894 (2019).
- 859 66. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in  
860 large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
- 861 67. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data.  
862 *Nature* **562**, 203–209 (2018).
- 863 68. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative  
864 and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).
- 865 69. Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of  
866 Rare Missense Variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).
- 867 70. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and  
868 Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 869 71. Hawkes, G. *et al.* Genetic evidence that high BMI in childhood has a protective effect on  
870 intermediate diabetes traits, including measures of insulin sensitivity and secretion, after  
871 accounting for BMI in adulthood. *Diabetologia* **66**, 1472–1480 (2023).
- 872 72. Boughton, A. P. *et al.* LocusZoom.js: interactive and embeddable visualization of genetic  
873 association study results. *Bioinforma. Oxf. Engl.* **37**, 3017–3018 (2021).
- 874 73. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2 -- an R package  
875 for gene list functional enrichment analysis and namespace conversion toolset g:Profiler.  
876 *F1000Research* **9**, ELIXIR-709 (2020).
- 877 74. Kong, A. *et al.* Detection of sharing by descent, long-range phasing and haplotype  
878 imputation. *Nat. Genet.* **40**, 1068–1075 (2008).
- 879 75. Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic  
880 population. *Nat. Genet.* **47**, 435–444 (2015).

- 881 76. Eggertsson, H. P. *et al.* GraphTyper2 enables population-scale genotyping of structural  
882 variation using pangenome graphs. *Nat. Commun.* **10**, 5402 (2019).
- 883 77. Delaneau, O., Zagury, J.-F., Robinson, M. R., Marchini, J. L. & Dermitzakis, E. T.  
884 Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 5436 (2019).
- 885 78. Halldorsson, B. V. *et al.* The sequences of 150,119 genomes in the UK Biobank. *Nature*  
886 **607**, 732–740 (2022).
- 887 79. Mirshahi, U. L. *et al.* Reduced penetrance of MODY-associated HNF1A/HNF4A variants  
888 but not GCK variants in clinically unselected cohorts. *Am. J. Hum. Genet.* **109**, 2018–2028  
889 (2022).