Effects of the maternal and fetal proteome on birth weight: a Mendelian randomization analysis

Nancy McBride*^{1,2}, Alba Fernández-Sanlés^{1,3}, Marwa Al Arab², Tom A. Bond^{1,2,11}, Jie Zheng^{5,6}, Maria C. Magnus⁷, Elizabeth C. Corfield^{1,8,9}, Gemma L Clayton^{1,2}, Liang-Dar Hwang¹¹, Robin N. Beaumont¹⁰, David M. Evans^{1,4,11}, Rachel M. Freathy¹⁰, Tom R. Gaunt^{1,2}, Deborah A Lawlor^{1,2}, Maria Carolina Borges^{1,2}

¹MRC Integrative Epidemiology Unit at the University of Bristol

² Population Health Science, Bristol Medical School, University of Bristol

³MRC Unit for Lifelong Health and Ageing, University College London

4 Frazer Institute, University of Queensland, Brisbane, QLD, Australia

⁵Department of Endocrine and Metabolic Diseases, Shanghai Institute of Endocrine and Metabolic Diseases, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, 200025, China

6 Shanghai National Clinical Research Center for Metabolic Diseases, Key Laboratory for Endocrine and Metabolic Diseases of the National Health Commission of the PR China, Shanghai Key Laboratory for Endocrine Tumor, State Key Laboratory of Medical Genomics, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, 200025, China

⁷ Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway ⁸Psychiatric Genetic Epidemiology Group, Research Department, Lovisenberg Diaconal Hospital, Oslo, Norway

⁹PsychGen Centre for Genetic Epidemiology and Mental Health, Norwegian Institute of Public Health, Oslo, Norway

 10 Department of Clinical and Biomedical Sciences, Faculty of Health and Life Sciences, University of Exeter, Exeter, UK

¹¹Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia

*Corresponding author nancy.mcbride@bristol.ac.uk

¹**Abstract**

2 Background

3 Fetal growth is an important indicator of survival, regulated by maternal and fetal genetic and
4 environmental factors. However, little is known about the underlying molecular mechanisms.

- ⁴environmental factors. However, little is known about the underlying molecular mechanisms.
- 5 Proteins play a major role in a wide range of biological processes and could provide key
6 insights into maternal and fetal molecular mechanisms regulating fetal growth.
- insights into maternal and fetal molecular mechanisms regulating fetal growth.
- ⁷Method

8 We used intergenerational two-sample Mendelian randomization to explore the effects of
9 1.139 maternal and fetal genetically-instrumented plasma proteins on birth weight. We used 9 1,139 maternal and fetal genetically-instrumented plasma proteins on birth weight. We used
10 oenome-wide association summary data from the Early Growth Genetics (EGG) consortium 10 genome-wide association summary data from the Early Growth Genetics (EGG) consortium
11 (n=406.063 with maternal and/or fetal genotype), with independent replication in the 11 (n=406,063 with maternal and/or fetal genotype), with independent replication in the
12 Norwegian Mother, Father and Child Cohort Study (MoBa: n=74.932 mothers and n=62.108 12 Norwegian Mother, Father and Child Cohort Study (MoBa; n=74,932 mothers and n=62,108
13 offspring). Maternal and fetal data were adiusted for the correlation between fetal and 13 offspring). Maternal and fetal data were adjusted for the correlation between fetal and
14 maternal genotype to distinguish their independent genetic effects. maternal genotype, to distinguish their independent genetic effects.

15 Results

16 We found that higher genetically-predicted maternal levels of NEC1 increased birth weight
17 (mean-difference: 12a (95% CI [6a, 18al) per 1 standard deviation protein level) as did 17 (mean-difference: 12g (95% CI [6g, 18g]) per 1 standard deviation protein level) as did
18 PRS57 (20a [10a, 31a]) and ULK3 (140a [81a, 199a]). Higher maternal levels of Galectin 4 18 PRS57 (20g [10g, 31g]) and ULK3 (140g [81g, 199g]). Higher maternal levels of Galectin_4
19 decreased birth weight (-206g [-299g, -113g]). In contrast, in the offspring, higher 19 decreased birth weight (-206g [-299g, -113g]). In contrast, in the offspring, higher
20 denetically-predicted offspring levels of NEC1 decreased birth weight (-10g [-16g, -5g]). 20 genetically-predicted offspring levels of NEC1 decreased birth weight (-10g [-16g, -5g]),
21 alongside sLeptin R (-8g [-12g, -4g]), and UBS3B (-78g [-116g, -41g]), Higher fetal levels of 21 alongside sLeptin_R (-8g [-12g, -4g]), and UBS3B (-78g [-116g, -41g]). Higher fetal levels of
22 Galectin 4 increased birth weight (174g [89g, 258g]). We replicated these results in MoBa, 22 Galectin_4 increased birth weight (174g [89g, 258g]). We replicated these results in MoBa,
23 and found supportive evidence for shared causal variants from genetic colocalization 23 and found supportive evidence for shared causal variants from genetic colocalization
24 analyses and protein-protein-network-associations analyses and protein-protein network associations.

25 Conclusions

26 We find strong evidence for causal effects, sometimes in opposing directions, of maternal
27 and fetal genetically-instrumented proteins on birth weight. These provide new insights into 27 and fetal genetically-instrumented proteins on birth weight. These provide new insights into
28 maternal and fetal molecular mechanisms regulating fetal growth, involving glucose 28 maternal and fetal molecular mechanisms regulating fetal growth, involving glucose
29 metabolism energy balance and vascular function that could be used to identify new 29 metabolism, energy balance, and vascular function that could be used to identify new
30 intervention targets to reduce the risk of fetal growth disorders, and their associated adverse 30 intervention targets to reduce the risk of fetal growth disorders, and their associated adverse
31 maternal and fetal outcomes. maternal and fetal outcomes.

- 32 Keywords: fetal growth, birth weight, proteomics, molecular mechanisms, Mendelian
33 randomization, pregnancy, MoBa
- randomization, pregnancy, MoBa

34

35 Introduction

36 Birth weight is a valuable and widely used measure of fetal growth. Healthy fetal growth is
37 essential to minimise adverse perinatal health outcomes. including miscarriage. stillbirth. 37 essential to minimise adverse perinatal health outcomes, including miscarriage, stillbirth,
38 preterm birth and associated neonatal and infant morbidity and mortality (1-7). Large-scale 38 preterm birth and associated neonatal and infant morbidity and mortality (1-7). Large-scale
39 aenome-wide association studies (GWAS) have revealed hundreds of genomic regions 39 genome-wide association studies (GWAS) have revealed hundreds of genomic regions
40 independently associated with birth weight (2, 3, 8-10). In addition, these GWAS have 40 independently associated with birth weight (2, 3, 8-10). In addition, these GWAS have
41 uncovered a complex interplay between maternal and fetal genomes (8, 9), in which some 41 uncovered a complex interplay between maternal and fetal genomes (8, 9), in which some
42 denetic variants have maternal or fetal-specific effects on birth weight. whereas other genetic 42 genetic variants have maternal or fetal-specific effects on birth weight, whereas other genetic
43 variants have both maternal and fetal effects. sometimes in opposite directions. Despite 43 variants have both maternal and fetal effects, sometimes in opposite directions. Despite
44 these advances in mapping fetal and maternal genetic variants, understanding the molecular 44 these advances in mapping fetal and maternal genetic variants, understanding the molecular
45 mechanisms underpinning fetal growth remains a challenge (6). mechanisms underpinning fetal growth remains a challenge (6).

46 Proteins play a major role in a wide range of biological processes (11). They are essential for
47 cellular growth and repair, and could provide key insights into maternal and fetal molecular 47 cellular growth and repair, and could provide key insights into maternal and fetal molecular
48 mechanisms regulating fetal growth (10). Mendelian Randomization (MR) is a method which 48 mechanisms regulating fetal growth (10). Mendelian Randomization (MR) is a method which
49 uses genetic variants associated with exposures as instrumental variables, to test the effect 49 uses genetic variants associated with exposures as instrumental variables, to test the effect
50 of those exposures on human traits and diseases. The method aims to mitigate causal 50 of those exposures on human traits and diseases. The method aims to mitigate causal
51 effects being biased due to confounding and reverse causation that may explain 51 effects being biased due to confounding and reverse causation that may explain
52 conventional-observational-associations. Previous-MR-studies-have-established-causality-of-52 conventional observational associations. Previous MR studies have established causality of
53 maternal modifiable factors on offspring birth weight, including higher maternal adiposity and 53 maternal modifiable factors on offspring birth weight, including higher maternal adiposity and
54 bigher maternal circulating glucose levels on higher birth weight, and smoking and higher 54 higher maternal circulating glucose levels on higher birth weight, and smoking and higher
55 blood pressure on lower birth weight (12) (13, 14) (5) (15). Thev have also provided novel 55 blood pressure on lower birth weight (12) (13, 14) (5) (15). They have also provided novel
56 insights, such as a potential effect of higher maternal glutamine levels on higher birth weight 56 insights, such as a potential effect of higher maternal glutamine levels on higher birth weight
57 (5) and metabolically favourable adiposity on lower birth weight (16). As such. MR and 57 (5) and metabolically favourable adiposity on lower birth weight (16). As such, MR and
58 proteomics can be integrated to explore maternal and fetal proteins regulating fetal growth 58 proteomics can be integrated to explore maternal and fetal proteins regulating fetal growth
59 by using genetic variants associated with protein levels (i.e., protein quantitative trait loci -59 by using genetic variants associated with protein levels (i.e., protein quantitative trait loci -
60 pQTL) as instrumental variables. As most drug targets are proteins, identifving the effects of 60 pQTL) as instrumental variables. As most drug targets are proteins, identifying the effects of
61 – maternal circulating proteins on birth weight could identify targets for drug development that 61 Thaternal circulating proteins on birth weight could identify targets for drug development that
62 Thatin brevent fetal arowth restriction and over-growth. and their associated adverse 62 might prevent fetal growth restriction and over-growth, and their associated adverse
63 pregnancyand perinatal outcomes. pregnancy and perinatal outcomes.

64 The aim of this study was to identify causal effects of maternal and fetal proteins on birth
65 weight to highlight underlying molecular mechanisms. To achieve that, we used two-sample 65 weight to highlight underlying molecular mechanisms. To achieve that, we used two-sample
66 MR. in which large-scale GWAS of plasma proteomics and birth weight were combined to 66 MR, in which large-scale GWAS of plasma proteomics and birth weight were combined to
67 examine the causal effect of plasma proteins on offspring birth weight. examine the causal effect of plasma proteins on offspring birth weight.

⁶⁹**Methods**

70 We performed a two-sample MR study to identify maternal and fetal effects of plasma
71 Deroteins on offspring birthweight, examining both maternal and fetal genetic effects 71 proteins on offspring birthweight, examining both maternal and fetal genetic effects
72 (n=406.063 with maternal and/or fetal genotype). For proteins passing multiple testing 72 (n=406,063 with maternal and/or fetal genotype). For proteins passing multiple testing
73 correction (FDR p-value < 5%), we tested whether results replicated in an independent 73 correction (FDR p-value $<$ 5%), we tested whether results replicated in an independent
74 sample (n=74.932 mothers and n=62.108 offspring) and investigated whether MR results 74 sample (n=74,932 mothers and n=62,108 offspring) and investigated whether MR results
75 could be confounded by linkage diseguilibrium using genetic colocalization. For proteins 75 could be confounded by linkage disequilibrium using genetic colocalization. For proteins
76 v which had discovery and replication evidence, and no evidence against colocalization, we 76 which had discovery and replication evidence, and no evidence against colocalization, we
77 used protein network associations to examine plausible pathwavs through which thev could 77 used protein network associations to examine plausible pathways through which they could
78 be influencing birth weight. **Figure 1** summarises the analytical workflow used in this paper. ⁷⁸be influencing birth weight. **Figure 1** summarises the analytical workflow used in this paper.

81 **Figure 1** – Schematic representation of the analyses conducted in this paper. Abbreviations:
82 GWAS, genome wide association studies: pQTL, protein quantitative trait loci

- GWAS, genome wide association studies; pQTL, protein quantitative trait loci
- 83

84 Selection of genetic instruments for plasma proteins

85 We selected genetic instruments from the deCODE GWAS of plasma proteins, which
86 provides data on 4.459 unique proteins, mapped to 13.514 pQTL. For our analyses, we 86 provides data on 4,459 unique proteins, mapped to 13,514 pQTL. For our analyses, we
87 excluded trans instruments (N=2.647 proteins with 6.402 trans pQTL) and secondary 87 excluded trans instruments (N=2,647 proteins with 6,402 trans pQTL) and secondary
88 variants (N=5.238 secondary pQTL) retaining 1.783 cis pQTL defined as all those within 88 variants (N=5,238 secondary pQTL) retaining 1,783 cis pQTL defined as all those within
89 ±1□Mb) for 1.812 proteins (P□<□1□×□10⁻⁷) (Supplementary Figure 1) ±1□Mb) for 1,812 proteins (P□<□1□×□10⁻⁷) (Supplementary Figure 1)

90 Discovery sample
91

92 92 We used publicly available meta-analysed GWAS summary data on birth weight. These data
93 were obtained from the 2019 Early Growth Genetics (EGG) consortium plus UK Biobank 93 were obtained from the 2019 Early Growth Genetics (EGG) consortium plus UK Biobank
94 (UKBB)GWAS This GWAS has publicly available estimates of maternal genetic 94 (UKBB) GWAS. This GWAS has publicly available estimates of maternal genetic
95 associations adiusted for fetal genotype and fetal genetic associations adiusted for maternal 95 associations adjusted for fetal genotype and fetal genetic associations adjusted for maternal
96 aenotype (see below 'Adjustment for maternal/offspring genotype'). genotype (see below 'Adjustment for maternal/offspring genotype').

97 The following exclusions were made prior to running the GWAS: i) twins and other multiple
98 births, ii) offspring born earlier than 37 weeks of gestation, iii) extreme birth weight outliers 98 births, ii) offspring born earlier than 37 weeks of gestation, iii) extreme birth weight outliers
99 (only individuals born between 2.2kg and 4.6kg were included in the maternal GWAS, and 99 (only individuals born between 2.2kg and 4.6kg were included in the maternal GWAS, and
100 individuals born between 2.5 and 4.5kg were included in the offspring GWAS) and iv) babies 100 individuals born between 2.5 and 4.5kg were included in the offspring GWAS) and iv) babies
101 born with congenital anomalies (for the maternal genotype only). Genetic associations with 101 born with congenital anomalies (for the maternal genotype only). Genetic associations with
102 birth weight were reported by the GWAS in standard deviation (SD) units. We multiplied 102 birth weight were reported by the GWAS in standard deviation (SD) units. We multiplied
103 these by 551.36 (the SD of birth weight in the UKB) to obtain results in grams (g). these by 551.36 (the SD of birth weight in the UKB) to obtain results in grams (g).

104 Overall, N=406,063 participants contributed to the models estimating the maternal and fetal
105 oenetic effects on birth weight that were adiusted, respectively, for fetal and maternal 105 genetic effects on birth weight that were adjusted, respectively, for fetal and maternal
106 aenotype. These included N=101.541 participants who reported their own birth weight and 106 genotype. These included N=101,541 participants who reported their own birth weight and
107 birth weight of their first child, N=195.815 participants with their own birth weight (fetal 107 birth weight of their first child, N=195,815 participants with their own birth weight (fetal
108 GWAS), and N=108,707 female participants with data on offspring birth weight (maternal 108 GWAS), and N=108,707 female participants with data on offspring birth weight (maternal 109 GWAS)(8) (17). GWAS)(8) (17).

110 Replication study
111

112 112 For independent replication, we used genetic association data for birth weight from MoBa, a
113 Dopulation-based pregnancy cohort study conducted by the Norwegian Institute of Public 113 population-based pregnancy cohort study conducted by the Norwegian Institute of Public
114 Health. Pregnant women and their partners were recruited from all over Norway from 1999-114 Health. Pregnant women and their partners were recruited from all over Norway from 1999-
115 2008. The women consented to participate in 41% of the pregnancies. The cohort includes 115 2008. The women consented to participate in 41% of the pregnancies. The cohort includes
116 approximately N=114.500 children, N=95.200 mothers, and N=75.200 fathers (18). The 116 approximately N=114,500 children, N=95,200 mothers, and N=75,200 fathers (18). The
117 current study uses version 12 of the quality-assured genetic data files released for research 117 current study uses version 12 of the quality-assured genetic data files released for research
118 in January 2019. Birthweight was obtained from linked clinical records held in Medical Birth 118 in January 2019. Birthweight was obtained from linked clinical records held in Medical Birth
119 Registry of Norway (MBRN, a national health registry established in 1967 that contains 119 Registry of Norway (MBRN, a national health registry established in 1967 that contains
120 information about all births in Norway, which is linked to the MoBa study using unique 120 information about all births in Norway, which is linked to the MoBa study using unique
121 Loersonal identification numbers). Questionnaire responses from mothers and fathers data 121 personal identification numbers). Questionnaire responses from mothers and fathers data
122 were used to identify sex registered at birth, year of birth, birth weight, and multiple births were used to identify sex registered at birth, year of birth, birth weight, and multiple births
123 (only available in the offspring). As in the discovery GWAS, multiple births and extremes of 123 (only available in the offspring). As in the discovery GWAS, multiple births and extremes of
124 birth weight were excluded. Analyses were adjusted for principal components (PCs) and 124 birth weight were excluded. Analyses were adjusted for principal components (PCs) and
125 aenotvping batch.

genotyping batch.

126 In MoBa, plasma blood samples were obtained from both parents during pregnancy, and
127 Ifrom umbilical cord at birth (19). This proiect used MoBa genetic data that was quality 127 from umbilical cord at birth (19). This project used MoBa genetic data that was quality
128 controlled and imputed using the MoBaPsychGen pipeline, which has previously been 128 controlled and imputed using the MoBaPsychGen pipeline, which has previously been
129 described (20). Phasing and imputation were performed using the publicly available 129 described (20). Phasing and imputation were performed using the publicly available
130 Haplotype Reference Consortium release 1.1 panel as a reference (21). The pipeline output ¹³⁰Haplotype Reference Consortium release 1.1 panel as a reference (21). The pipeline output 131 consists of best-guess hard-call genotype data for 6,981,748 variants in 207,409 MoBa
132 participants of European ancestry. GWAS in MoBa were run using regenie software (v 3.1.2) 132 participants of European ancestry. GWAS in MoBa were run using *regenie* software (v 3.1.2)
133 on the TSD server (22). on the TSD server (22).

134 Adjustment for maternal/offspring genotype
135

135 136 Birth weight is influenced by both maternal and fetal genotype. Therefore, to distinguish
137 maternal and fetal genetic effects. in the discovery GWAS a weighted linear model was used 137 maternal and fetal genetic effects, in the discovery GWAS a weighted linear model was used
138 to estimate maternal genetic associations (adjusted for fetal genotype), and fetal genetic 138 to estimate maternal genetic associations (adjusted for fetal genotype), and fetal genetic
139 associations adiusted for maternal genotype (3). From this, genetically-instrumented 139 associations adjusted for maternal genotype (3). From this, genetically-instrumented
140 maternal effects of proteins on birth weight (not mediated or biased due to horizontal 140 maternal effects of proteins on birth weight (not mediated or biased due to horizontal
141 Dleiotropy via fetal effects) and genetically instrumented fetal effects of proteins on birth 141 pleiotropy via fetal effects) and genetically instrumented fetal effects of proteins on birth
142 veight (not confounded by maternal genotype) can be estimated. For our discovery sample weight (not confounded by maternal genotype) can be estimated. For our discovery sample
143 (EGG+UKB GWAS), we downloaded these adiusted estimates from the GWAS consortium 143 (EGG+UKB GWAS), we downloaded these adjusted estimates from the GWAS consortium
144 vebsite. For our replication sample (MoBa), we estimated adiusted associations using the 144 website. For our replication sample (MoBa), we estimated adjusted associations using the
145 weighted linear model implemented via the DONUTS software package (51). As part of this. 145 weighted linear model implemented via the DONUTS software package (51). As part of this,
146 we used genetic covariance intercepts from LD score regression (LDSC) to account for 146 we used genetic covariance intercepts from LD score regression (LDSC) to account for
147 overlap between maternal and fetal GWAS sample (23). The summary maternal adiusted for 147 overlap between maternal and fetal GWAS sample (23). The summary maternal adjusted for
148 fetal, and fetal adjusted for maternal results were then used in our replication two-sample 148 fetal, and fetal adjusted for maternal results were then used in our replication two-sample
149 MR. MR.

150 Two sample MR
151

151 152 We conducted MR analyses to study the causal effects of circulating maternal and fetal
153 proteins on birth weight. All MR analyses were performed using the TwoSampleMR R 153 proteins on birth weight. All MR analyses were performed using the *TwoSampleMR* R
154 package version 0.5.6 (26). We used the Wald ratio estimator for analyses as only one pQTL 154 package version 0.5.6 (26). We used the Wald ratio estimator for analyses as only one pQTL
155 was available per protein. Prior to MR analysis, genetic association data for proteins and 155 was available per protein. Prior to MR analysis, genetic association data for proteins and
156 birth weight were harmonized so that the effect of each genetic variant on the exposure and 156 birth weight were harmonized so that the effect of each genetic variant on the exposure and
157 outcome were relative to the same effect allele. In some instances, data could not be 157 outcome were relative to the same effect allele. In some instances, data could not be
158 harmonised (due to being palindromic SNPs. A/T or G/C, with allele frequencies close to 158 harmonised (due to being palindromic SNPs, A/T or G/C, with allele frequencies close to
159 50%), and these pQTL and their corresponding proteins were removed from further analyses 159 50%), and these pQTL and their corresponding proteins were removed from further analyses
160 (Supplementary Table 2). After excluding pQTL that failed harmonisation, we retained data ¹⁶⁰(**Supplementary Table 2)**. After excluding pQTL that failed harmonisation, we retained data 161 for 1,139 proteins instrumented by the same number of *cis* pQTL (**Figure 1** and
162 **Supplementary Figure 1**). We then analysed maternal and fetal data separately. MR results **Supplementary Figure 1**). We then analysed maternal and fetal data separately. MR results

163 with a FDR-adjusted p-value ≤0.05 were carried forward to the replication MR and
164 subsequent analyses. subsequent analyses.

165 We sought replication of the discovery MR results using genetic association data for birth
166 weight from MoBa participants. There is no overlap between the discovery and replication 166 weight from MoBa participants. There is no overlap between the discovery and replication
167 participants. The same procedures for data harmonisation and two-sample MR analyses 167 participants. The same procedures for data harmonisation and two-sample MR analyses
168 described in the discovery stage were used in the replication analyses. To report MR results 168 described in the discovery stage were used in the replication analyses. To report MR results
169 as the mean difference in birth weight (g) per 1 SD increase in protein level, we multiplied 169 as the mean difference in birth weight (g) per 1 SD increase in protein level, we multiplied
170 MR effect estimates and standard error by 565.63 (g), which corresponds to 1 SD of birth 170 MR effect estimates and standard error by 565.63 (g), which corresponds to 1 SD of birth
171 weight in MoBa. We then conducted a fixed effects meta-analysis of estimates from 171 weight in MoBa. We then conducted a fixed effects meta-analysis of estimates from
172 discovery and replication samples using the meta R package (version 6.5-0). Our criteria for 172 discovery and replication samples using the *meta* R package (version 6.5-0). Our criteria for
173 – replication were: i) directionally consistent results between discovery and replication. ii) 173 replication were: i) directionally consistent results between discovery and replication, ii)
174 associations with a p-value threshold <0.05 in the meta-analysis, as commonly used in 174 associations with a p-value threshold <0.05 in the meta-analysis, as commonly used in 175 GWAS, and reflecting the relatively modest size of the replication sample in comparison to 175 GWAS, and reflecting the relatively modest size of the replication sample in comparison to
176 the main analysis sample and (iii) there was no strong evidence supporting heterogeneity 176 the main analysis sample and (iii) there was no strong evidence supporting heterogeneity
177 between discovery and replication MR estimates (Cochrane's Q p-value > 0.05). between discovery and replication MR estimates (Cochrane's Q p-value > 0.05).

178 Genetic colocalization
179

180 ¹⁸⁰We undertook genetic colocalization analyses using the *coloc* R package (24) to investigate 181 whether our MR findings were compatible with a shared causal variant or confounded by
182 linkage diseguilibrium (LD) (24). This could happen if a selected pQTL is in LD with another ¹⁸²linkage disequilibrium (LD) (24). This could happen if a selected pQTL is in LD with another 183 genetic variant influencing birth weight via another biological mechanism. *Coloc* computes
184 all possible configurations of causal variants for each of two traits and uses a Bavesian 184 all possible configurations of causal variants for each of two traits and uses a Bayesian
185 approach to calculate support for each causal model (H_o: no association: H₁: association 185 approach to calculate support for each causal model (H₀: no association; H₁: association
186 with protein only: H₂: association with birth weight only: H₃: both traits owing to distinct 186 with protein only; H₂: association with birth weight only; H₃: both traits owing to distinct 187 causal variants: H₄: both traits owing to distinct 187 causal variants; H₄: both traits owing to a shared causal variant). We restricted the analyses
188 to a ±500kb window around the pQTL, and assumed a prior probability that any pQTL in the 188 to a ±500kb window around the pQTL, and assumed a prior probability that any pQTL in the
189 region is associated with the protein only (p₁) of 1 x 10⁻⁴, birth weight only (p₂) of 1 x 10⁻⁴, or region is associated with the protein only (p_1) of 1 \times 10⁻⁴, birth weight only (p_2) of 1 \times 10⁻⁴, or
190 both traits (p_{12}) 1 \times 10⁻⁵. We considered a posterior probability of association (PPA) for H both traits (p_{12}) 1 × 10⁻⁵. We considered a posterior probability of association (PPA) for H₄ > 191. 0.70 to provide strong evidence for colocalization, while a PPA for H₃ > 0.5 was considered 191 0.70 to provide strong evidence for colocalization, while a PPA for $H_3 > 0.5$ was considered
192 as evidence against colocalization. To facilitate visualisation of genomic signals in each 192 as evidence against colocalization. To facilitate visualisation of genomic signals in each
193 region, we generated stacked recombination plots using the *locuscompareR* R package 193 region, we generated stacked recombination plots using the *locuscompareR* R package
194 (http://locuscompare.com/) (25). (http://locuscompare.com/) (25).

195 Where available, we extracted genetic association data for the genomic regions selected for
196 colocalization analyses for plasma proteins and birth weight using the IEU Open GWAS 196 colocalization analyses for plasma proteins and birth weight using the IEU Open GWAS
197 (https://awas.mrcieu.ac.uk/). The summarv statistics for the genetic associations were 197 (https://gwas.mrcieu.ac.uk/). The summary statistics for the genetic associations were
198 downloaded from the deCODE website https://www.decode.com/summarydata/. ¹⁹⁸downloaded from the deCODE website https://www.decode.com/summarydata/.

199 Fetal expression lookup
200

201 One of the assumptions of MR is that the genetic instruments are statistically strongly
202 associated with the phenotype in the population to which inference would be made (i.e. the ²⁰²associated with the phenotype in the population to which inference would be made (i.e. the 203 relevance assumption). We were unable to determine whether our genetic instruments
204 associated with fetal (in utero) circulating proteins, due to the absence of large-scale 204 associated with fetal (in utero) circulating proteins, due to the absence of large-scale
205 datasets with both offspring genetic data and cord-blood measures of circulating proteins. 205 datasets with both offspring genetic data and cord-blood measures of circulating proteins.
206 Instead, we explored whether the proteins with evidence from discovery and replication were 206 Instead, we explored whether the proteins with evidence from discovery and replication were
207 expressed in fetal tissue. We used the single-cell transcriptomics data from the human cell 207 expressed in fetal tissue. We used the single-cell transcriptomics data from the human cell
208 atlas of fetal gene expression (26). The atlas was developed from 121 human fetal samples 208 atlas of fetal gene expression (26). The atlas was developed from 121 human fetal samples
209 (post-conceptional age = 72-129 davs) from 15 organs (eve. heart. intestine. kidnev. liver. ²⁰⁹(post-conceptional age = 72-129 days) from 15 organs (eye, heart, intestine, kidney, liver, 210 Iung, muscle, pancreas, placenta, spleen, stomach, thymus, adrenal, cerebellum and
211 cerebrum), profiling 4 million single cells using a sci-RNA-seg3 protocol. The authors used a 211 cerebrum), profiling 4 million single cells using a sci-RNA-seq3 protocol. The authors used a
212 clustering approach building on a form of principal components analysis (uniform manifold 212 clustering approach building on a form of principal components analysis (uniform manifold
213 approximation and proiection - UMAP) to generate a resource where fetal gene expression

213 approximation and projection - UMAP) to generate a resource where fetal gene expression
214 can be mapped across different organs, tissues and cell types 214 can be mapped across different organs, tissues and cell types 215 (https://descartes.brotmanbaty.org/). We used this resource to check that our proteins of
216 interest were expressed in fetal tissue, as a further line of evidence. 216 interest were expressed in fetal tissue, as a further line of evidence.
217

218 Protein networks
219 We explored inf

219 We explored information on protein-protein associations (referred to as interactions in
220 denetic research) from the STRING database (https://string-db.org/cgi) to gain insights into 220 genetic research) from the STRING database (https://string-db.org/cgi) to gain insights into
221 presumed biological processes underlying our findings. STRING collates information from all 221 presumed biological processes underlying our findings. STRING collates information from all
222 known protein-protein interactions databases and provides a visual network of predicted 222 known protein-protein interactions databases and provides a visual network of predicted
223 associations for a particular group of proteins, whereby the proteins are represented as 223 associations for a particular group of proteins, whereby the proteins are represented as
224 thodes' and the lines which ioin them are the predicted functional associations. The colour of ²²⁴'nodes' and the lines which join them are the predicted functional associations. The colour of 225 the line denotes the source of the interaction information (see figure legend) (27). We set the
226 confidence score of a given association to be 'high confidence' and restricted to associations 226 confidence score of a given association to be 'high confidence' and restricted to associations
227 from curated databases (either KEGG, Reactome, BioCvc and Gene Ontology, as well as 227 from curated databases (either KEGG, Reactome, BioCyc and Gene Ontology, as well as
228 legacy datasets from PID and BioCarta), experimentally determined (where evidence comes 228 legacy datasets from PID and BioCarta), experimentally determined (where evidence comes
229 from lab experiments), reported from gene fusions (from reported gene fusion events), gene 229 from lab experiments), reported from gene fusions (from reported gene fusion events), gene
230 co-occurrence (where gene families have similar occurrence patterns across genomes), co-230 co-occurrence (where gene families have similar occurrence patterns across genomes), co-231 expression (where gene are shown to be correlated in expression across a large number of 232 datasets) and protein homology (where proteins may have similar structures). datasets) and protein homology (where proteins may have similar structures).

²⁰⁰

²³⁴**Results**

235 Identifying causal effects of maternal circulating proteins on offspring birth weight
236

237 237 We found evidence of causal effects of 5 maternal proteins (instrumented by 5 cis-pQTLs)
238 on offspring birth weight that passed our 5% false discovery rate (FDR) adiusted p-value 238 on offspring birth weight that passed our 5% false discovery rate (FDR) adjusted p-value
239 (Fiqure 2). We found evidence that higher NEC1 increased birth weight (mean-difference: ²³⁹**(Figure 2).** We found evidence that higher NEC1 increased birth weight (mean-difference: 240 12g (95% CI: 6g, 18g) per 1 standard deviation protein level) as did PRS57 (20g (10g, 31g))
241 and ULK3 (140g (81g, 199g)). Higher maternal levels of Galectin 4 decreased birth weight (-241 and ULK3 (140g (81g, 199g)). Higher maternal levels of Galectin_4 decreased birth weight (-
242 206g (-299g, -113g)), as did PTN9 (-159g (-237g, -82g)), (Figure 2),

²⁴²206g (-299g, -113g)), as did PTN9 (-159g (-237g, -82g)). **(Figure 2)**.

²⁴⁴**Figure 2** – Two sample MR results using maternal cis pQTLs against offspring birth weight. 245 FDR corrected p-values transformed to their -log10. Results plotted in green passed 5%
246 FDR corrected p-value threshold. Abbreviations: SD. standard deviation FDR corrected p-value threshold. Abbreviations: SD, standard deviation

248 MR results for maternal proteins NEC_1, Galectin_4, ULK3 and PTN9 identified in discovery
249 analyses met our replication criteria based on meta-analysed results from pooled discovery 249 analyses met our replication criteria based on meta-analysed results from pooled discovery
250 and replication. However, the replication result for PRS57 was not directionally consistent 250 and replication. However, the replication result for PRS57 was not directionally consistent
251 and there was evidence of high heterogeneity in the meta-analysis for PRS57, ULK3, PTN9 251 and there was evidence of high heterogeneity in the meta-analysis for PRS57, ULK3, PTN9
252 (Figure 3 and Supplementary Table 6 and 8). ²⁵²(**Figure 3** and **Supplementary Table 6 and 8**).

253

254
255

²⁵⁵**Figure 3** Mendelian randomization results estimating the effect of genetically instrumented 256 maternal proteins (in standard deviation units) on offspring birth weight (in grams) in the 257 discovery (orange dots N=210,248 mothers), replication (red dots; N=87,651 mothers) and 257 discovery (orange dots N=210,248 mothers), replication (red dots; N=87,651 mothers) and
258 meta-analysed discovery and replication data. Replication was tested for proteins passing 258 meta-analysed discovery and replication data. Replication was tested for proteins passing
259 FDR-corrected p-value <0.05 in the discovery data. Our criteria for replication were: i) 259 FDR-corrected p-value <0.05 in the discovery data. Our criteria for replication were: i)
260 directionally consistent results between discovery and replication. ii) associations with a p-260 directionally consistent results between discovery and replication, ii) associations with a p-
261 value threshold <0.05 in the meta-analvsis, as commonly used in GWAS, and reflecting the 261 value threshold <0.05 in the meta-analysis, as commonly used in GWAS, and reflecting the 262 relatively modest size of the replication sample in comparison to the main analysis sample 262 Trelatively modest size of the replication sample in comparison to the main analysis sample
263 Tand (iii) there was no strong evidence supporting heterogeneity between discovery and 263 and (iii) there was no strong evidence supporting heterogeneity between discovery and 264 replication MR estimates (Cochrane's Q p-value > 0.05). replication MR estimates (Cochrane's Q p-value > 0.05).

266 For maternal effects, we undertook colocalization analysis for all proteins which passed our
267 FDR-adiusted p-value in the discovery analysis. For colocalization analyses, results are 267 FDR-adjusted p-value in the discovery analysis. For colocalization analyses, results are
268 Shown in Table 1 as posterior probabilities (PP) of no association (H₀), association with 268 shown in **Table 1** as posterior probabilities (PP) of no association (H₀), association with 269 protein only (H₁), birth weight only (H₂), both traits owing to distinct causal variants (H₃), or protein only (H_1) , birth weight only (H_2) , both traits owing to distinct causal variants (H_3) , or

- 270 both traits owing to a shared causal variant (H_4) (24). The colocalization analyses supported
271 causal effects for maternal circulating NEC1. Galectin 4 and PRS57 with offspring birth
-
- 271 causal effects for maternal circulating NEC1, Galectin_4 and PRS57 with offspring birth 272 weight. with high H₄ of 0.82, 0.84, and 0.76 respectively (Table 1). For PTN9 and ULK3, weight, with high H₄ of 0.82, 0.84, and 0.76 respectively **(Table 1)**. For PTN9 and ULK3,
273 there was strong evidence that the association stemmed from distinct causal variants.
- there was strong evidence that the association stemmed from distinct causal variants.
- ²⁷⁴**Table 1** Genetic colocalization results for maternal proteins with birth weight

275 Abbreviations – PP, posterior probabilities; H, hypothesis. Results are expressed as 276 posterior probabilities (PP) of no association (H0), association with protein only (H1), birth 276 posterior probabilities (PP) of no association (H0), association with protein only (H1), birth 277 weight only (H2), both traits owing to a 277 weight only (H2), both traits owing to distinct causal variants (H3), or both traits owing to a
278 shared causal variant (H4). shared causal variant (H4).

279 Maternal discovery and replication MR, and colocalization results of are summarised in
280 Table 2. Together these provide strong evidence for maternal circulating proteins Galectin 4 ²⁸⁰**Table 2.** Together these provide strong evidence for maternal circulating proteins Galectin_4 281 and NEC1 being identified in discovery, replication, and not being biased due to confounded
282 by LD (H₄ 0.84 and 0.82, respectively). There was statistical support for PRS57 in the 282 by LD (H₄ 0.84 and 0.82, respectively). There was statistical support for PRS57 in the
283 discovery and colocalization analysis (H₄ 0.76), but not the replication. Whilst there was 283 discovery and colocalization analysis (H_4 0.76), but not the replication. Whilst there was
284 good replication for ULK3, colocalization results for ULK3 and PTN9 suggested results were 284 good replication for ULK3, colocalization results for ULK3 and PTN9 suggested results were
285 being driven by distinct causal variants (H₃ > 0.9). being driven by distinct causal variants (H₃ > 0.9).

-
- 289
- 290
-

294

297
298

298 Identifying causal effects of fetal circulating proteins on offspring birth weight
299

300 300 We found evidence of causal effects of 14 proteins on offspring birth weight. Higher
301 genetically-predicted offspring levels of NEC1 decreased birth weight (-10g (-16g, -5g), 301 genetically-predicted offspring levels of NEC1 decreased birth weight (-10g (-16g, -5g), 302 alongside sLeptin R (-8g (-12g, -4g)), UBS3B (-78g (-116g, -41g)), FUT5 (-19g (-29g, -9g)), 302 alongside sLeptin_R (-8g (-12g, -4g)), UBS3B (-78g (-116g, -41g)), FUT5 (-19g (-29g, -9g)),
303 IFN a b R1 (-53g (-84g, -23g)), TWEAK (-24g, (-37g, -11g)) and MK13 (-69g (-108g, -303 IFN_a_b_R1 (-53g (-84g, -23g)), TWEAK (-24g, (-37g, -11g)) and MK13 (-69g (-108g, -
304 30g)). Higher fetal levels of Galectin 4 increased birth weight (174g (89g, 258g)), alongside 304 30g)). Higher fetal levels of Galectin_4 increased birth weight (174g (89g, 258g)), alongside
305 ACADV (227g (142g, 311g)), PLCG1 (133g (72g, 193g)), BCMA (61g (27g, 96g)) Cystatin C 305 ACADV (227g (142g, 311g)), PLCG1 (133g (72g, 193g)), BCMA (61g (27g, 96g)) Cystatin C
306 (37g (16g, 58g), and FKB1B (46g (26g, 67g)) (**Figure 4**). ³⁰⁶(37g (16g, 58g), and FKB1B (46g (26g, 67g)) (**Figure 4**).

307

³⁰⁸**Figure 4** – Fetal two sample MR results for cis pQTL. FDR corrected p-values transformed 309 to their -log10. Results plotted in green passed FDR corrected p-value threshold.
310 Abbreviations: SD standard-deviation Abbreviations: SD, standard deviation

312 As in the maternal effects analysis, we conducted replication in MoBa and then ran a meta-
313 analysis. We found meta-analysed MR results consistent with discovery for NEC1, 313 analysis. We found meta-analysed MR results consistent with discovery for NEC1,
314 Galectin 4. sLeptin R. TWEAK. ACADV. UBS3B. BCMA (same direction of effect and with a 314 Galectin_4, sLeptin_R, TWEAK, ACADV, UBS3B, BCMA (same direction of effect and with a
315 meta-analysed FDR adiusted pyalue <0.05 and no evidence of between discovery and 315 meta-analysed FDR adjusted pvalue <0.05 and no evidence of between discovery and
316 Freplication results heterogeneity), PLCG1, MK13, FUT5, FKB1B, and Cystatin c did not 316 replication results heterogeneity). PLCG1, MK13, FUT5, FKB1B, and Cystatin_c did not
317 replicate, and the replication sample estimate had the opposite direction of effect to the 317 replicate, and the replication sample estimate had the opposite direction of effect to the
318 discovery sample (Fiqure 5 and Supplementary Table 7). The pQTL for protein 318 discovery sample (**Figure 5** and **Supplementary Table 7**). The pQTL for protein
319 IFN a b R1 was not in our replication data so could not be tested or meta-analysed 319 IFN_a_b_R1 was not in our replication data so could not be tested or meta-analysed
320 (Supplementary Table 9). ³²⁰(**Supplementary Table 9**).

322 **Figure 5** Mendelian randomization analyses estimating the effect of genetically-instrumented
323 fetal proteins (in standard deviation units) on offspring birth weight (in grams) in the
324 discovery (N=297.356 offspr 323 fetal proteins (in standard deviation units) on offspring birth weight (in grams) in the
324 discovery (N=297,356 offspring, orange dots), replication (N=62,108 offspring, red dots) and
325 meta-analyses (blue dots) of 324 discovery (N=297,356 offspring, orange dots), replication (N=62,108 offspring, red dots) and
325 meta-analyses (blue dots) of discovery and replication samples. Replication was tested for
326 proteins passing FDR-corre 326 proteins passing FDR-corrected p-value <0.05 in the discovery data. Our criteria for
327 replication were: i) directionally consistent results between discovery and replication. ii) 327 replication were: i) directionally consistent results between discovery and replication, ii)
328 associations with a p-value threshold <0.05 in the meta-analysis, as commonly used in 328 associations with a p-value threshold <0.05 in the meta-analysis, as commonly used in
329 GWAS, and reflecting the relatively modest size of the replication sample in comparison to 328 associations with a p-value threshold <0.05 in the meta-analysis, as commonly used in
329 GWAS, and reflecting the relatively modest size of the replication sample in comparison to
330 the main analysis sample and (iii 329 GWAS, and reflecting the relatively modest size of the replication sample in comparison to
330 the main analysis sample and (iii) there was no strong evidence supporting heterogeneity
221 between discovery and replicat 330 the main analysis sample and (iii) there was no strong evidence supporting heterogeneity
331 between discovery and replication MR estimates (Cochrane's Q p-value > 0.05). 331 between discovery and replication MR estimates (Cochrane's Q p-value > 0.05).

-
- 332
- 333 For fetal effects, we undertook colocalization analysis for all proteins which passed our FDR
334 adjusted p-value in the discovery analysis. There was strong evidence of colocalization for
- 334 adjusted p-value in the discovery analysis. There was strong evidence of colocalization for
335 NEC1, ACADV, PLCG1, FUT5, IFN_a_b_R1, AIF1L and UBS3B (all H₄ > 0.75). There was
- 335 NEC1, ACADV, PLCG1, FUT5, IFN_a_b_R1, AIF1L and UBS3B (all $H_4 > 0.75$). There was moderate evidence of colocalization for sLeptin_R (H_4 0.65). There was evidence that the
- 336 moderate evidence of colocalization for sLeptin_R (H₄ 0.65). There was evidence that the
337 MR results for fetal Galectin_4, Cystatin-C, BCMA and MK13 might be explained by LD 337 MR results for fetal Galectin_4, Cystatin-C, BCMA and MK13 might be explained by LD
338 confounding (H₃>0.5 or high H₁) (**Table 3**).
- 338 confounding (H₃ > 0.5 or high H₁) (**Table 3**).
339 **Table 3** Genetic colocalization results for
- **Table 3** Genetic colocalization results for offspring proteins with birth weight.

340 Abbreviations – PP, posterior probabilities; H, hypothesis. Results are expressed as
341 posterior probabilities (PP) of no association (H0), association with protein only (H1), birth
342 weight only (H2), both traits

341 posterior probabilities (PP) of no association (H0), association with protein only (H1), birth
342 weight only (H2), both traits owing to distinct causal variants (H3), or both traits owing to a
343 shared causal 342 weight only (H2), both traits owing to distinct causal variants (H3), or both traits owing to a
343 shared causal variant (H4). 343 shared causal variant (H4).

- 345 Fetal discovery and replication MR, and colocalization results of are summarised in **Table 4.**
- 346 Together these provide strong evidence for fetal circulating proteins.
347
-

348
349
350

Fetal expression lookup
As a sanity check of our fetal results, we were able to check whether these genes are 351
352 expressed in fetal tissue using an atlas of human gene expression during development 353 (https://descartes.brotmanbaty.org/). We wanted to see if tissue expression data supported 353 (<u>https://descartes.brotmanbaty.org/</u>). We wanted to see if tissue expression data supported
354 expression of the proteins with evidence from discovery and replication or colocalization in 354 expression of the proteins with evidence from discovery and replication or colocalization in
355 fetal tissue. We found this was the case for genes encoding NEC1 (in pancreas), FUT5 (in 355 fetal tissue. We found this was the case for genes encoding NEC1 (in pancreas), FUT5 (in
356 lung and thymus), Galectin_4, UBS3B, sLeptin_R, AIF1L, PLCG1, IFN_a_b_R1 and ACADV 356 lung and thymus), Galectin_4, UBS3B, sLeptin_R, AIF1L, PLCG1, IFN_a_b_R1 and ACADV
357 (all expression all across multiple tissues) (**Supplementary Figures 20-29**).

357 (all expression all across multiple tissues) (**Supplementary Figures 20-29**).

358 Protein networks of identified maternal and fetal proteins affecting birth weight
359 We undertook protein-protein association analyses for the proteins that replicated in the MR 359 We undertook protein-protein association analyses for the proteins that replicated in the MR
360 results, and had supporting colocalization results in either maternal (NEC1, Galectin_4) or 360 results, and had supporting colocalization results in either maternal (NEC1, Galectin_4) or
361 fetal analyses (NEC1, ACADV, UBSH3B).

361 fetal analyses (NEC1, ACADV, UBSH3B).
362 We found evidence that NEC1 (proprotein convertase 1, the same enzyme as PCSK1, as in 362 We found evidence that NEC1 (proprotein convertase 1, the same enzyme as PCSK1, as in
363 the **Figure 6B** below) associates with proteins regulating glucose homeostasis, such as 363 the **Figure 6B** below) associates with proteins regulating glucose homeostasis, such as

364 insulin (INS), and glucagon (GCG) - hormones produced by the pancreas to regulate blood 365 sugar levels. NEC1 is a stimulator of INS (28, 29). It also associates with proteins regulating ³⁶⁶food intake, such as pro-opiomelanocortin (POMC) and ghrelin (GHRL) (28-31) (**Figure 6B**). ³⁶⁷UBS3B associated with EGFR (part of the epidermal growth factor family), CBL (critical for ³⁶⁸cell signalling pathways), and UBC (which helps regulate cell stress) (**Figure 6C**). ACADV 369 had the largest network of associations, including many proteins associated with fatty acid 370 metabolism (ACOX1, CPT2, and other proteins within the thiolase family (enzymes which 371 break down fatty acids) ACAA1 and ACAA2) (Figure 6A). However, Galectin_4 and PRS57 372 had no protein networks which survived our criteria (high confidence, and from sources 373 detailed in **Figure 6** footnote).

³⁸²**Figure 6** - Protein-protein associations - proteins with evidence from both discovery, 383 replication, and colocazation analyses. Different line colours represent different sources of
384 evidence for these predictions – known associations from curated databases (light blue) and 384 evidence for these predictions – known associations from curated databases (light blue) and
385 experimentally determined (pink): predicted associations from gene neighbourhood (green). 385 experimentally determined (pink); predicted associations from gene neighbourhood (green), 386 aene fusions (red), gene co-occurrence (dark blue), co-expression, and protein homology 386 gene fusions (red), gene co-occurrence (dark blue), co-expression, and protein homology
387 (lilac). Associations are for proteins iointly contributing to a shared function, rather than 387 (lilac). Associations are for proteins jointly contributing to a shared function, rather than
388 specifically binding to each other. Splice isoforms or post-translational modifications are 388 specifically binding to each other. Splice isoforms or post-translational modifications are
389 collapsed, i.e. each node represents all the proteins produced by a single, protein-coding 389 collapsed, i.e. each node represents all the proteins produced by a single, protein-coding
390 oene locus. The physical distances between two nodes along an edge in a graph have no 390 gene locus. The physical distances between two nodes along an edge in a graph have no 391 meaning. meaning.

³⁹²**Discussion**

³⁹³We have used two-sample MR to explore the causal effect of maternal pregnancy and fetal 394 plasma proteins on offspring birth weight. We found evidence of potential causal effects, 395 unlikely to be explained by chance or confounding by LD, for 2 maternal circulating proteins ³⁹⁶(NEC1, Galectin_4) and 3 fetal circulating proteins (NEC1, ACADV, and UBS3B). For the 397 remainder of the discussion, we focus mainly on these proteins which had the strongest 398 evidence across all analyses.

399 Proteins with opposing maternal and fetal effects on birthweight
400 Two proteins, NEC1 and Galectin_4, instrumented by cis-pQTLs, appear to be acting in 400 Two proteins, NEC1 and Galectin_4, instrumented by cis-pQTLs, appear to be acting in
401 opposite directions when comparing maternal and fetal MR results. They suggest that more 401 opposite directions when comparing maternal and fetal MR results. They suggest that more
402 NEC1 from the mother leads to higher birth weight, yet more NEC1 from the fetus leads to 402 NEC1 from the mother leads to higher birth weight, yet more NEC1 from the fetus leads to
403 lower birth weight – as in the case with insulin and other regulators of fetal growth. This is 404 consistent with previous work showing the role of NEC1 in regulating proinsulin processing 404 consistent with previous work showing the role of NEC1 in regulating proinsulin processing
405 (32). Indeed, a new drug (Setmelanotide) being tested to treat obesity is known to perturb 405 (32). Indeed, a new drug (Setmelanotide) being tested to treat obesity is known to perturb
406 PCSK1, because PCSK1 mutations lead to deficient signalling in melanocortin pathways 406 PCSK1, because PCSK1 mutations lead to deficient signalling in melanocortin pathways
407 (e.g. because of a lack of α-MSH). Setmelanotide activates MC4R to restore appetite 407 (e.g. because of a lack of α-MSH). Setmelanotide activates MC4R to restore appetite
408 control, addressing genetic obesity caused by disrupted PCSK1 function (33). This is further 408 control, addressing genetic obesity caused by disrupted PCSK1 function (33). This is further
409 evidence that our findings could represent molecular targets for birth weight. The STRING 409 evidence that our findings could represent molecular targets for birth weight. The STRING
410 protein-protein associations in this study show this link between NEC1 and insulin (INS) and 411 glucagon (GCG) (Figure 6B), demonstrating this known function in glucose homeostasis. It 412 is important to note that whilst these protein-protein associations show co-expression (which 412 is important to note that whilst these protein-protein associations show co-expression (which
413 may be downstream of a certain mechanism), they do not always imply interaction. 413 may be downstream of a certain mechanism), they do not always imply interaction.
414 Furthermore, the different sources of evidence are important when interpreting these 414 Furthermore, the different sources of evidence are important when interpreting these
415 interaction networks. We did not include text-mining and gene neighborhood interactions in 416 the STRING networks so we could follow the most robust lines of evidence, and set a high 417 confidence threshold. This may explain why we did not identify any associations for some 417 confidence threshold. This may explain why we did not identify any associations for some
418 proteins that replicated and were supported by colocalization, or because proteins like 418 proteins that replicated and were supported by colocalization, or because proteins like
419 Galectin_4 are poorly characterized (27). 119 Galectin_4 are poorly characterized (27).
420 In our analysis, Galectin_4 was shown to the opposite direction of effect to NEC1 on birth

421 weight. More Galectin_4 from the mother leads to lower birthweight and more Galectin_4 421 weight. More Galectin_4 from the mother leads to lower birthweight and more Galectin_4
422 from the fetus leads to higher birthweight. Galectin_4 is part of the Galectin family which 422 from the fetus leads to higher birthweight. Galectin_4 is part of the Galectin family which
423 binds carbohydrates, lactose and sugars (34). Previous work has shown that the Galectin 423 binds carbohydrates, lactose and sugars (34). Previous work has shown that the Galectin
424 family of β -galactoside binding proteins are important in modulating diverse developmental 425 processes. They contribute to placentation and regulate maternal immune tolerance (34) and 425 processes. They contribute to placentation and regulate maternal immune tolerance (34) and
426 have been previously suggested as biomarkers for adverse pregnancy outcomes like 426 have been previously suggested as biomarkers for adverse pregnancy outcomes like
427 gestational diabetes (35) and infertility (34). Other expression studies have shown the 427 gestational diabetes (35) and infertility (34). Other expression studies have shown the
428 importance of Galectins in pregnancy, and the importance of a maternal-fetal balancing act 429 for ideal pregnancy outcomes. For example, lower levels of fetal expression of Galectins 3, 8 430 and 9 have previously been associated with intrauterine growth restriction (34). However, 430 and 9 have previously been associated with intrauterine growth restriction (34). However,
431 most of the previous studies of plasma Galectin_4 concentrations in pregnancy have come 431 most of the previous studies of plasma Galectin_4 concentrations in pregnancy have come
432 from mouse or in vitro models (35). It has not been well characterised either inside or outside 432 from mouse or in vitro models (35). It has not been well characterised either inside or outside
433 of pregnancy, demonstrated by the lack of reported associations from the protein-protein 433 of pregnancy, demonstrated by the lack of reported associations from the protein-protein

434 association models. Our study provides an additional line of evidence of a potential role of
435 Galectin_4 in fetal growth however there was a lack of colocalization in the fetal effects 435 Galectin_4 in fetal growth however there was a lack of colocalization in the fetal effects
436 analysis (34). analysis (34).
137 – PRS57 had strong evidence in the discovery and colocalization analyses for increasing birth

438 weight, however it was the only protein in the maternal analyses which did not replicate. 438 weight, however it was the only protein in the maternal analyses which did not replicate.
439 Whilst other members of the serine protease family (a large family of protein-cleaving 439 Whilst other members of the serine protease family (a large family of protein-cleaving
440 enzymes) appear critical in pregnancy (e.g. high levels of other serine protease proteins are 440 enzymes) appear critical in pregnancy (e.g. high levels of other serine protease proteins are
441 associated with preeclampsia (36)), to our knowledge, PRS57 has not been previously 441 associated with preeclampsia (36)), to our knowledge, PRS57 has not been previously
442 linked with birth weight, or other disorders of pregnancy. In contrast, PTN9 and ULK3 had 442 linked with birth weight, or other disorders of pregnancy. In contrast, PTN9 and ULK3 had
443 evidence from discovery and replication, but not from colocalization. Unfortunately, we had 443 evidence from discovery and replication, but not from colocalization. Unfortunately, we had
444 limited power to explore further bias in MR results due to confounding by LD, but high H₃ 444 limited power to explore further bias in MR results due to confounding by LD, but high H₃
445 results suggest this was down to the traits not sharing the same causal variant, although it 445 results suggest this was down to the traits not sharing the same causal variant, although it
446 could also be related to violations of coloc's assumptions due to multiple causal variants. could also be related to violations of coloc's assumptions due to multiple causal variants.
447 Proteins with fetal-specific effects on birthweight

447 Proteins with fetal-specific effects on birthweight
448 UBS3B and ACADV had good evidence from o

448 UBS3B and ACADV had good evidence from discovery, replication, and colocalization. In
449 our analysis, we found higher UBS3B lowered birth weight. UBS3B is a protein that is not yet 450 well characterised for its role in pregnancy, but known to be associated with cellular 451 processes for mammalian development (37). ACADV (very long-chain acyl-coenzyme A 451 processes for mammalian development (37). ACADV (very long-chain acyl-coenzyme A
452 dehydrogenase, also known as VCLAD) catalyses the first step in beta-oxidation of long-452 dehydrogenase, also known as VCLAD) catalyses the first step in beta-oxidation of long-
453 chain fatty acids to release energy (38). 453 chain fatty acids to release energy (38).
454 In our fetal analysis, we found evidence that higher fetal cis-pQTL instrumented sLeptin_R

455 reduced birth weight. sLeptin_R has a well-established role in energy metabolism, body 455 reduced birth weight. sLeptin_R has a well-established role in energy metabolism, body
456 weight, and is secreted in fetal tissue and the placenta (39). The strong birth weight-456 weight, and is secreted in fetal tissue and the placenta (39). The strong birth weight-
457 associated locus near CCNL1 from previous GWAS is also associated with leptin levels in 457 associated locus near *CCNL1* from previous GWAS is also associated with leptin levels in 458 adults, providing orthogonal evidence for a key role of this signalling pathway in fetal growth
459 (40). There was good evidence for replication in our study, however the colocalization results 459 (40). There was good evidence for replication in our study, however the colocalization results
460 fell slightly short of our threshold (H₄ 0.65, as opposed to 0.70). 460 fell slightly short of our threshold (H_4 0.65, as opposed to 0.70).
461 Strengths and limitations

461 Strengths and limitations
462 VVe have integrated the largest-scale proteomics and birthweight genetic association data to 463 shed light on maternal and fetal molecular mechanisms regulating fetal growth. Furthermore, 463 shed light on maternal and fetal molecular mechanisms regulating fetal growth. Furthermore,
464 we have used large scale birth weight data from three large European cohorts (EGG, UK 464 we have used large scale birth weight data from three large European cohorts (EGG, UK
465 Biobank and MoBa) for our discovery, replication analyses and meta analyses. We have 465 Biobank and MoBa) for our discovery, replication analyses and meta analyses. We have

466 undertaken sensitivity analyses, including checking for confounding by LD and attempted to
467 further explain the underlving biology of our molecular findings. 467 further explain the underlying biology of our molecular findings.
468 In terms of limitations. our analysis has a limited coverage of th

468 In terms of limitations, our analysis has a limited coverage of the proteome due to the use of
469 GWAS data on plasma proteins. For example, because they are not covered in existing 470 proteomic GWAS, we were unable to assess the effects of two key placental expressed 471 proteins - sFlt-1 and PIGF -, which have biological support for affecting fetal growth, and 471 proteins - sFlt-1 and PIGF - , which have biological support for affecting fetal growth, and
472 are now used clinically to predict risk of fetal growth restriction and pre-eclampsia in many 472 are now used clinically to predict risk of fetal growth restriction and pre-eclampsia in many
473 countries (41-43). To explore effects of these and potentially other placental proteins we 473 countries (41-43). To explore effects of these and potentially other placental proteins we
474 would need large proteomic GWAS of maternal pregnancy and cord-blood proteins. It is also 474 would need large proteomic GWAS of maternal pregnancy and cord-blood proteins. It is also
475 likely that there are additional causal proteins that are not placental-/pregnancy-specific, 475 likely that there are additional causal proteins that are not placental-/pregnancy-specific,
476 which might be identified in future, where studies have access to even larger proteomics 476 which might be identified in future, where studies have access to even larger proteomics
477 panels, or using cell- or tissue-specific proteomics.

477 panels, or using cell- or tissue-specific proteomics.
478 Another potential limitation is related to epitope effects intrinsic to the proteomic technology. 479 Whilst we aimed to reduce the likelihood of misclassification of protein levels through epitope 479 Whilst we aimed to reduce the likelihood of misclassification of protein levels through epitope
480 effects (i.e. differences in antigen recognition, rather than protein levels) we cannot rule out 480 effects (i.e. differences in antigen recognition, rather than protein levels) we cannot rule out
481 some bias from these effects (44). We removed trans-pQTL and only retained cis-pQTL due 481 some bias from these effects (44). We removed trans-pQTL and only retained cis-pQTL due
482 to the higher possibility of bias due to pleiotropic mechanisms (45). For each protein we only 482 to the higher possibility of bias due to pleiotropic mechanisms (45). For each protein we only
483 had one SNP, so methods that have been developed to explore bias due to horizontal 483 had one SNP, so methods that have been developed to explore bias due to horizontal
484 pleiotropy are not possible. 484 pleiotropy are not possible.
485 – In using population level proteomics data, a further limitation is that in the exposure data,

486 In whilst we can assume in the maternal analysis that the offspring proteins won't affect the 487 levels of the maternal proteins, in the offspring analysis, the assumption that maternal 487 levels of the maternal proteins, in the offspring analysis, the assumption that maternal
488 proteins won't affect offspring proteins may not hold. However, there is no way to directly 488 proteins won't affect offspring proteins may not hold. However, there is no way to directly
489 test this. 489 test this.
490 Furthermore, we have assumed that genetic variants identified in protein GWAS conducted

491 in males and non-pregnant women are similarly associated with circulating proteins in 491 in males and non-pregnant women are similarly associated with circulating proteins in
492 fetuses and pregnant women. Ideally, we would test this in studies that have GWAS data 492 fetuses and pregnant women. Ideally, we would test this in studies that have GWAS data
493 and circulating protein levels in women during pregnancy and cord-blood. However, such 493 and circulating protein levels in women during pregnancy and cord-blood. However, such
494 data is not currently available or only available in small numbers (for example, there is 495 protein data profiled during pregnancy in the Born in Bradford (BiB) cohort, however the 495 protein data profiled during pregnancy in the Born in Bradford (BiB) cohort, however the
496 Olink coverage is fewer than 453 proteins in 4000 women, with none of these proteins 496 Olink coverage is fewer than 453 proteins in 4000 women, with none of these proteins
497 overlapping with our top hits). However we found evidence that some of the proteins (i.e. 497 overlapping with our top hits). However we found evidence that some of the proteins (i.e.
498 NEC1, Galectin_4) that our MR results suggest affect birth weight, are expressed in fetal 498 NEC1, Galectin_4) that our MR results suggest affect birth weight, are expressed in fetal
499 tissue, which provides some support for their relevance to pregnancy and fetal intrauterine 499 tissue, which provides some support for their relevance to pregnancy and fetal intrauterine

500 mechanisms. Additionally, supporting evidence from other studies, where genetic variants
501 are instrumental variables for social, behavioural, and molecular exposures during 501 are instrumental variables for social, behavioural, and molecular exposures during
502 pregnancy, reveals consistent associations with those identified in GWAS of both women 502 pregnancy, reveals consistent associations with those identified in GWAS of both women
503 and men. This suggests that the assumption is unlikely to be violated. Significantly, with 503 and men. This suggests that the assumption is unlikely to be violated. Significantly, with
504 respect to this paper, we have previously demonstrated that among 89 associations 505 involving genetic instruments for one or more of 9 amino acids, there was substantial 505 involving genetic instruments for one or more of 9 amino acids, there was substantial 506 consistency between those GWAS results derived in men and non-pregnant women, to
507 outcomes in women only (46). This consistency was also observed in women who were 507 outcomes in women only (46). This consistency was also observed in women who were
508 pregnant at the time of amino acid measurement for 67 out of the 89 associations (75%). In 508 pregnant at the time of amino acid measurement for 67 out of the 89 associations (75%). In
509 the remaining 22 associations where heterogeneity was evident, it was likely attributed to the 509 the remaining 22 associations where heterogeneity was evident, it was likely attributed to the
510 poor quality of genetic imputation for certain variants in the pregnancy study (46). 510 poor quality of genetic imputation for certain variants in the pregnancy study (46).
511 Conclusions

511 Conclusions
512 We found strong evidence for causal effects of several proteins on offspring birth weight 512 We found strong evidence for causal effects of several proteins on offspring birth weight
513 across a range of analyses, highlighting proteins with opposing maternal and offspring 513 across a range of analyses, highlighting proteins with opposing maternal and offspring
514 effects (NEC1, Galectin_4), and fetal (sLeptin_R, UBS3B, ACADV) effects. These proteins 514 effects (NEC1, Galectin_4), and fetal (sLeptin_R, UBS3B, ACADV) effects. These proteins
515 are involved in multiple biological processes governing glucose homeostasis, energy 515 are involved in multiple biological processes governing glucose homeostasis, energy
516 metabolism, endothelial function and adipocyte differentiation. These findings provide new 516 metabolism, endothelial function and adipocyte differentiation. These findings provide new
517 insights into maternal and fetal molecular mechanisms regulating fetal growth. 517 insights into maternal and fetal molecular mechanisms regulating fetal growth.

518 **Abbreviations**

- 519 EGG: Early Growth Genetics consortium
520 eQTL: expression quantitative trait loci
- 520 eQTL: expression quantitative trait loci
521 GWAS: genome-wide association stud
- 521 GWAS: genome-wide association study
522 LD: linkage disequilibrium
- 522 LD: linkage disequilibrium
523 MoBa: Norwegian Mother.
- 523 MoBa: Norwegian Mother, Father and Child Cohort Study
524 MR: Mendelian Randomisation
- 524 MR: Mendelian Randomisation
525 PP: predicted probability
- 525 PP: predicted probability
- 526 pQTL: protein quantitative trait loci
527 SNP: single nucleotide polymorphis
- 527 SNP: single nucleotide polymorphism
528 UKBB: UK Biobank
- UKBB: UK Biobank

529 **Declarations**

530 **Ethics approval and consent to participate**
531 All human research was approved by the relevant institutional review boards and conducted

532 according to the Declaration of Helsinki. Participants of all studies in the Early Growth 533 Genetics (EGG) consortium provided written informed consent. The UK Biobank has 533 Genetics (EGG) consortium provided written informed consent. The UK Biobank has
534 approval from the North West Multi-Centre Research Ethics Committee, which covers the 534 approval from the North West Multi-Centre Research Ethics Committee, which covers the
535 United Kingdom. The establishment of the Norwegian Mother Father and Child (MoBa) birth 535 United Kingdom. The establishment of the Norwegian Mother Father and Child (MoBa) birth
536 Cohort and initial data collection was based on a license from the Norwegian Data Protection 537 Agency and approval from The Regional Committees for Medical and Health Research 537 Agency and approval from The Regional Committees for Medical and Health Research
538 Ethics. The MoBa cohort is currently regulated by the Norwegian Health Registry Act. The 538 Ethics. The MoBa cohort is currently regulated by the Norwegian Health Registry Act. The
539 current study was approved by The Regional Committees for Medical and Health Research 539 current study was approved by The Regional Committees for Medical and Health Research
540 Ethics (223273). 540 Ethics (223273).
541 Consent for publication

 542 NA 542 NA

543 **Availability of data and materials**

544 **Data availability**
545 Data on birth weight in our discovery analysis was contributed by the EGG Consortium using 546 the UK Biobank Resource and was downloaded from www.egg-consortium.org. The 546 the UK Biobank Resource and was downloaded from <u>www.egg-consortium.org</u>. The
547 genotype and phenotype data are available on application from the UK Biobank 547 genotype and phenotype data are available on application from the UK Biobank
548 (http://www.ukbiobank.ac.uk/). Individual cohorts participating in the EGG Consortium should 548 (http://www.ukbiobank.ac.uk/). Individual cohorts participating in the EGG Consortium should 550 summary data from which genetic instruments were selected for pQTL are available from 551 EpiGraphDB pQTL browser (https://epigraphdb.org/pqtl). Data from MoBa are available from 551 EpiGraphDB pQTL browser (<u>https://epigraphdb.org/pqtl</u>). Data from MoBa are available from
552 the Norwegian Institute of Public Health after application to the MoBa Scientific Management 552 the Norwegian Institute of Public Health after application to the MoBa Scientific Management
553 Group (see its website https://www.fhi.no/en/op/data-access-from-health-registries-health-553 Group (see its website https://www.fhi.no/en/op/data-access-from-health-registries-health-
554 studies-and-biobanks/data-access/applving-for-access-to-data/ for details). studies-and-biobanks/data-access/applying-for-access-to-data/ for details).

Acknowledgments
556 We gratefully acknowledge the participants in the studies, who were directly or indirectly 556 We gratefully acknowledge the participants in the studies, who were directly or indirectly
557 participating, via consortia summary data, contributed to this research. We also are grateful 557 participating, via consortia summary data, contributed to this research. We also are grateful
558 to researchers who made their data open access. The Norwegian Mother, Father and Child 558 to researchers who made their data open access. The Norwegian Mother, Father and Child

559 Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the
560 Ministry of Education and Research. We are grateful to all the participating families in 561 Norway who take part in this on-going cohort study. We thank the Norwegian Institute of 562 Public Health (NIPH) for generating high-quality genomic data. This research is part of the 563 HARVEST collaboration, supported by the Research Council of Norway (#229624). We also 564 thank the NORMENT Centre for providing genotype data, funded by the Research Council of 564 thank the NORMENT Centre for providing genotype data, funded by the Research Council of
565 Norway (#223273), South East Norway Health Authorities and Stiftelsen Kristian Gerhard 565 Norway (#223273), South East Norway Health Authorities and Stiftelsen Kristian Gerhard
566 Jebsen. We further thank the Center for Diabetes Research, the University of Bergen for 566 Jebsen. We further thank the Center for Diabetes Research, the University of Bergen for
567 providing genotype data and performing quality control and imputation of the data funded by 568 the ERC AdG project SELECTionPREDISPOSED, Stiftelsen Kristian Gerhard Jebsen, Trond 568 the ERC AdG project SELECTionPREDISPOSED, Stiftelsen Kristian Gerhard Jebsen, Trond
569 Mohn Foundation, the Research Council of Norway, the Novo Nordisk Foundation, the 569 Mohn Foundation, the Research Council of Norway, the Novo Nordisk Foundation, the
570 University of Bergen, and the Western Norway Health Authorities. This work was carried out 570 University of Bergen, and the Western Norway Health Authorities. This work was carried out
571 using the computational facilities of the Advanced Computing Research Centre, University of 572 Bristol - http://www.bristol.ac.uk/acrc/ and the TSD (Tjeneste for Sensitive Data) facilities, 573 owned by the University of Oslo, operated and developed by the TSD service group at the 574 University of Oslo, IT Department (USIT) (tsd-drift@usit.uio.no). The computations on this 574 University of Oslo, IT Department (USIT) (<u>tsd-drift@usit.uio.no</u>). The computations on this
575 platform were performed on resources provided by Sigma2 - the National Infrastructure for 575 platform were performed on resources provided by Sigma2 - the National Infrastructure for
576 High-Performance Computing and Data Storage in Norway. 576 High-Performance Computing and Data Storage in Norway.

577 **Funding**
578 This work was supported by the University of Bristol and UK Medical Research Council (579 MC_UU_00032/05), the US National Institute for Health (R01 DK10324), the European 580 Research Council via Advanced Grant 101021566, the British Heart Foundation 581 (AA/18/7/34219 and CS/16/4/32482) and the National Institute of Health Research Bristol 581 (AA/18/7/34219 and CS/16/4/32482) and the National Institute of Health Research Bristol
582 Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the 582 Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the
583 University of Bristol.

583 University of Bristol.
584 RMF and RNB were funded by a Wellcome Trust and Royal Society Sir Henry Dale 585 Fellowship (WT104150). RMF is supported by a Wellcome Senior Research Fellowship 586 (WT220390). This research was supported by the National Institute for Health and Care 587 Research Exeter Biomedical Research Centre. This research was funded in part by the 587 Research Exeter Biomedical Research Centre. This research was funded in part by the
588 Wellcome Trust [WT220390]. J.Z. is supported by grants from the National Key Research 588 Wellcome Trust [WT220390]. J.Z. is supported by grants from the National Key Research
589 and Development Program of China (2022YFC2505203). MCB is funded by a University of 589 and Development Program of China (2022YFC2505203). MCB is funded by a University of
590 Bristol Vice Chancellor's Fellowship. EC is supported by funding from the Research Council 590 Bristol Vice Chancellor's Fellowship. EC is supported by funding from the Research Council
591 of Norway (#274611) and the South-Eastern Norway Regional Health Authority (#2021045). 591 of Norway (#274611) and the South-Eastern Norway Regional Health Authority (#2021045).

- 592 The views expressed in this paper are those of the authors and not necessarily those of the 593 any of the funders listed above.
- 593 any of the funders listed above.
594 For open access. the authors h
- 594 For open access, the authors have applied a CC BY public copyright licence to any Author
595 Accepted Manuscript version arising from this submission.
- Accepted Manuscript version arising from this submission.

596 **Author contributions**
1997 Study conception: DAL, MCB, RMF; Analysis: NM, MCB; Manuscript drafting: NM; 597 Study conception: DAL, MCB, RMF; Analysis: NM, MCB; Manuscript drafting: NM;
598 Manuscript revisions: NM, MCB, DAL, AFS, MA, RB, RMF, TB, TRG, JZ, DE, DH, GC, EC, 598 Manuscript revisions: NM, MCB, DAL, AFS, MA, RB, RMF, TB, TRG, JZ, DE, DH, GC, EC,
599 MM 599 MM

600 **Conflict of interest**

The authors report no conflict of interest.

602 **Code availability**

603 All code is available on the GitHub repo https://github.com/MRCIEU/MR-PREG

604 **References**
605 1. Küpers LK, Monnereau C, Sharp GC, Yousefi P, Salas LA, Ghantous A, et al. Meta-analysis of 606 epigenome-wide association studies in neonates reveals widespread differential DNA methylation
607 associated with birthweight. Nature Communications. 2019:10(1):1893. 607 essociated with birthweight. Nature Communications. 2019;10(1):1893.
608 2. Leite DFB. Cecatti JG. Fetal Growth Restriction Prediction: How t

608 a. Leite DFB, Cecatti JG. Fetal Growth Restriction Prediction: How to Move beyond. The 609 associations. 2019;2019:1519048. 608 2. Leite DFB, Cecatti JG. Fetal Growth Restriction Prediction: How to Move beyond. The

610 Scientific Communication Construction Constitution 3. Warrington NM, Freathy RM, Neale M
611 Digital Actionate maternal and fetal effects on 610 3. Warrington NM, Freathy RM, Neale MC, Evans DM. Using structural equation modelling to
611 jointly estimate maternal and fetal effects on birthweight in the UK Biobank. International Journal of
612 Epidemiology. 2018 Fig. 3. Jointly estimate material and fetal effects on birth material effects on birthmaterial effects.
612 Lepidemiology. 2018;47(4):1229-41.
613 Lepidemiology. 2018;47(4):1229-41.

613 $\frac{67}{4}$ Wilcox AJ. On the importance
614 Lournal of Enidemiology, 2001:30(6) 614 Journal of Epidemiology. 2001;30(6):1233-41.
615 5. Barry CA-O, Lawlor DA, Shanland CA-O, Sanderson E. Borges MA-O, Using Mendelian.

 615 $5.$ Barry CA-O, Lawlor DA, Shapland CA-C
616 Bandomisation to Prioritise Candidate Matern 615 5. Barry CA-O, Lawlor DA, Shapland CA-O, Sanderson E, Borges MA-O. Using Mendelian
616 Randomisation to Prioritise Candidate Maternal Metabolic Traits Influencing Offspring Birthweight.
617 LI

617 LID - 10.3390/metabo12060537 [doi] LID - 537. (2218-1989 (Print)).
618 6. Bond TA, Karhunen V, Wielscher M, Auvinen J, Männikkö M 618 6. Bond TA, Karhunen V, Wielscher M, Auvinen J, Männikkö M, Keinänen-Kiukaanniemi S, et al.
619 Exploring the role of genetic confounding in the association between maternal and offspring body 619 618 6. Exploring the role of genetic confounding in the association between maternal and offspring body
620 600 mass index: evidence from three birth cohorts. International Journal of Epidemiology. 620 mass index: evidence from three birth cohorts. International Journal of Epidemiology.
621 02019;49(1):233-43.

622 7. Iliodromiti S, Mackay DF, Smith GCS, Pell JP, Sattar N, Lawlor DA, et al. Customised and 622 7. Iliodromiti S
623 Noncustomised Birt Fig. 1. Interaction of the little state of the SSA.
624 The Cohort Study of 979.912 Term Singleton Pregnancies in Scotland, PLOS Medicine 624 A Cohort Study of 979,912 Term Singleton Pregnancies in Scotland. PLOS Medicine.
625 2017;14(1):e1002228.

124 A Communication of 979,912 Term Singleton Pregnancies in Scotland. Plus Medicine.
625 - 2017;14(1):e1002228.
626 - 8 - Warrington NM. Beaumont RN. Horikoshi M. Day ER. Helgeland Ø. Laurin C. end and an architecture.
626 28. Warrington NN
627 2nd fetal senetic effect 626 8. Warrington NM, Beaumont RN, Horikoshi M, Day FR, Helgeland Ø, Laurin C, et al. Maternal
627 and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. Nature
628 genetics. 2019; genetics 2019;51(5):804-14. $\frac{6}{5}$

629 629 9. Beaumont RN, Warrington NM, Cavadino A, Tyrrell J, Nodzenski M, Horikoshi M, et al.
630 Genome-wide association study of offspring birth weight in 862577 women identifies five novel loci
631 and highlights maternal 9. 631 and highlights maternal genetic effects that are independent of fetal genetics. (1460-2083)
632 (Electronic)). 632 (Electronic)). 633 and highlights material general ge $\begin{array}{ll}634 & \text{Large-scale integration of the plasma proteome with genetics and disease. Nature Genetics.\end{array}$ 635 2021;53(12):1712-21.
636 Large-scale integration of the plansma protes and protes and disease. Nature Genetics and disease. Nature 635 2021;53(12):1712-21. 637 Mendelian randomization mapping the influence of the plasma proteome on complex disease
638 Mature Genetics, 2020;52/10):1122-31 638 Mature Genetics. 2020;52(10):1122-31.
639 12 Thang G. Bacelis L. Lengvel C. Teramo K. Hallman M. Helgeland M. et al. Assessing the Ca 639 22. Zhang G, Bacelis J, Lengyel C, Te
640 Belationship of Maternal Height on Birtl 639 12. Zhang G, Bacelis J, Lengyel C, Teramo K, Hallman M, Helgeland Ø, et al. Assessing
640 Relationship of Maternal Height on Birth Size and Gestational Age at Birth: A Mendelian
641 Randomization Analysis. (1549-1676 641 - Relationship of Maternal Heght on Birth Size and Gestational Hgc and Gestation Analysis. (1549-1676 (Electronic)).
642 - 13 - Tyrrell L Richmond RC Palmer TM. Feenstra B. Rangaraian L Metrustry S. et al. (ender die Kanadisticke Analysis. (1544)
643 – Evidence for Causal Relationshins Between Materr 643 Evidence for Causal Relationships Between Maternal Obesity-Related Traits and Birth Weight. (1538-
644 3598 (Electronic)). 645 $14.$ Ardissino¹¹
646 Hypertension Incre 645 14. Ardissino MA-O, Slob EA-O, Millar O, Reddy RA-OX, Lazzari L, Patel KHK, et al. Maternal 646 Hypertension Increases Risk of Preeclampsia and Low Fetal Birthweight: Genetic Evidence From
647 Mendelian Bandomization Study (1524-4563 (Electronic)) 647 Mendelian Randomization Study. (1524-4563 (Electronic)).
648 15. Brand JA-O, Gaillard RA-O, West JA-O, McEachan R 648 15. Brand JA-O, Gaillard RA-O, West JA-O, McEachan RA-O, Wright JA-O, Voerman E, et al.
649 Associations of maternal quitting, reducing, and continuing smoking during pregnancy with 649 Associations of maternal quitting, reducing, and continuing smoking during pregnancy with
650 Iongitudinal fetal growth: Findings from Mendelian randomization and parental negative co 650 Iongitudinal fetal growth: Findings from Mendelian randomization and parental negative control
651 studies. (1549-1676 (Electronic)). es dongitudinal fetal growth: Findings from Mendelian randomization from Mendelian randomization and para
652 longitudinal randomization and parameters from Mendelian randomization and parameters from particle control
652 652 16. Thompson WA-O, Beauments.
653 Higher maternal adinosity reduces 652 16. Thompson WA-O, Beaumont RA-O, Kuang AA-O, Warrington NA-OX, Ji Y, Tyrrell JA
653 Higher maternal adiposity reduces offspring birthweight if associated with a metabolically
654 favourable profile. (1432-0428 (Electr $\begin{bmatrix} 654 & 6 \end{bmatrix}$ favourable profile. (1432-0428 (Electronic)).
655 $\begin{bmatrix} 17 & 0 \end{bmatrix}$ allen NE, Sudlow C, Beakman T, Collins B, Biobank UK, UK biobank data: come and end
655 favour - Allen NE, Sudlow C, Peakman T, Colli
656 filteraal Med, 6. United States2014 in 224ed4 1986 18. Allen Ne, Sudan T, Sudania, Sudania, Sudania, Sudania, Sudania, Sudania, Sudania, Sudania, 1986
656 17. Martin Raugan A. Birke C. Alsaker E. Højseth G. Knudsen GP. Tambs K. et al. Cobort Profile Undate: .
657 18. Haugan A, Birke C, Alsaker E, Høiseth
658 The Nonvesian Mother and Child Cohort Stud 657 18. Haugan A, Birke C, Alsaker E, Høiseth G, Knudsen GP, Tambs K, et al. Cohort Profile Upo
658 The Norwegian Mother and Child Cohort Study (MoBa). International Journal of Epidemiology.
659 2016;45(2):382-8. 660 $19.$ Paltiel L, A 660 19. Paltiel L, Anita H, Skjerden T, Harbak K, Bækken S, Nina Kristin S, et al. The biobank of the 661 Norwegian Mother and Child Cohort Study – present status. Norsk Epidemiologi. 2014;24(1-2).
662 – 20. – Corfield EC, Erei O. Shadrin AA, Bahman Z, Lin A, Athanasiu L, et al. The Nonvegian Mothe 661 Norwegian Mother and Child Cohort Study – present status. Norsk Epidemiologi. 2014;24(1-2).
662 20. Corfield EC, Frei O, Shadrin AA, Rahman Z, Lin A, Athanasiu L, et al. The Norwegian Mother,
663 Father. and Child coho 663 Father, and Child cohort study (MoBa) genotyping data resource: MoBaPsychGen pipeline v.1.
664 bioRxiv. 2022:2022.06.23.496289. Father, and completence comp_{le}ment, gener, ping data resources in each pipeline v.2.
664 – bioRxiv. 2022:2022.06.23.496289.
665 – 21 – McCarthy S. Das S. Kretzschmar W. Delaneau O. Wood AR. Teumer A. et al. A refer er and the contract of the con 665 21. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A referen
666 panel of 64,976 haplotypes for genotype imputation. Nature Genetics. 2016;48(10):1279-83.
667 22. Mbatchou J. Barnard L. Backman 667 22. Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, et al.
668 Computationally efficient whole-genome regression for quantitative and binary traits. Na 668 Computationally efficient whole-genome regression for quantitative and binary traits. Nature 669 Genetics. 2021;53(7):1097-103. $\frac{1}{669}$ Genetics. 2021;53(7):1097-103.
670 – 23 – Bulik-Sullivan BK Joh P-R. Finucane HK. Rinke S. Vang L. Patterson N. et al. I.D. Score erral discrete contract the state of the state of the 1600 state of the 1600 state of the 1600 state of the 16
671 – Integrassion distinguishes confour 671 regression distinguishes confounding from polygenicity in genome-wide association studies. Nature 672 Genetics. 2015:47(3):291-5. ة المستوى التي تصدر المستوى .
672 - Genetics، 2015;47(3):291-5.
673 - مستوى العام العام العام العام العام والمستوى العام العام العام . erectives. 2016. 2016. 2016. 2016. 2016. 2016. 2016. 2016. 2016. 2016. 2016. 2016. 2016. 2016. 2016. 2016. 20
1914 - Enidemiology, 2013:37(8):80 ers and the communications of Shared General Testing of Shared Center Control for Potential Control for Potent
675 - 25. Liu B. Gloudemans M.I. Rao AS. Ingelsson E. Montgomery SR. Abundant associations with 675 25. Liu B, Gloudemans MJ, Raccords 2014
676 – Sene expression complicate GWAS 675 25. Liu B, Gloudemans MJ, Rao AS, Ingelsson E, Montgomery SB. Abundant as
676 gene expression complicate GWAS follow-up. Nature Genetics. 2019;51(5):768-9.
677 26. Cao J, O'Day DR, Pliner HA, Kingsley PD, Deng M, Daza 677 26. Cao J, O'Day DR, Pliner HA, Kingsley PD, Deng M, Daza RM, et al. A human cell atlas of fetal
678 gene expression. Science. 2020;370(6518):eaba7721. gene expression. Science. 2020;370(6518):eaba7721. $\frac{1}{2}$

679 679 27. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING
680 v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Research.
681 2014;43(D1):D4 $27₁$ ese and the state of the state
682 28. Benzinou M, Cre
683 C. Dina C. Fau - Durand E 682 28. Benzinou M, Creemers Jw Fau - Choquet H, Choquet H Fau - Lobbens S, Lobbens S Fau - Dina er and the creamers in the matter of the program, the presention of the choquet H, Common S, Lobbens S, Lobben
683 C, Dina C Fau - Durand E, Durand E Fau - Guerardel A, et al. Common nonsynonymous variants in
684 PCSK1 $\overline{684}$ PCSK1 confer risk of obesity. (1546-1718 (Electronic)).
685 - 29 - Stijnen P, Ramos-Molina B, O'Rabilly S, Creemers IWM, PCSK1 Mutations and Human 685 - Partie conference risk of operating the oriental process.
685 - 29. - Stijnen P, Ramos-Molina B, O'Rahilly S, Creeme
686 - Endocrinopathies: From Obesity to Gastrointestinal Dis 685 29. Stijnen P, Ramos-Molina B, O'Rahilly S, Creemers JWM. PCSK1 Mutations and Human
686 Endocrinopathies: From Obesity to Gastrointestinal Disorders. Endocrine Reviews. 2016;37(4):347-
687 71. 688
689 30. Wei X, Ma X, Lu R, Bai G, Zhang J, Deng R, et al. Genetic Variants in PCSK1 Gene Are 689 Associated with the Risk of Coronary Artery Disease in Type 2 Diabetes in a Chinese Han Population:
690 A Case Control Study. PLOS ONE. 2014;9(1):e87168. 689 Associated with the Risk of Coronary Artery Disease in Type 2 Diabetes in a Chinese Han Population: For the control of the Control Study. Planet Control Study.
691 a Case Control Study. Planet Control on Marine Study. Next Control of Success Met 691 31. Gjesing AP, Vestmar MA, Jørgensen T, Heni M, Holst JJ, Witte DR, et al. The Effect of
692 Variants on Waist, Waist-Hip Ratio and Glucose Metabolism Is Modified by Sex and Glucose
693 Tolerance Status. PLOS ONE. 201 $\begin{array}{ll} 693 & \text{Tolerance Status} \end{array}$ PLOS ONE. 2011;6(9):e23907.
694 32 Nead KT Li A. Webner MR. Neupane B. Gustafsson S. Butterworth A. et al. Contribut For the Status Contract Process-Processes.
694 Tolerance Status. Planet Status. PCSK1 to 1695. Common pon-synonymous variants in PCSK1 to 1 694 32. Nead KT, Li A, Wehner MR, Neupane B, Gustafsson S, Butterworth A, et al. Contribution
695 common non-synonymous variants in PCSK1 to body mass index variation and risk of obesity: a
696 systematic review and meta-a 696 systematic review and meta-analysis with evidence from up to 331 175 individuals. (1460-2083 697 (Electronic). 698 33 . Ryan
 699 and Bimagrup 698 33. Ryan DH. Next Generation Antiobesity Medications: Setmelanotide, Semaglutide, Tirzepatide For the computation and the controller measurement of the controller setting and properties.
699 and Bimagrumab: What do They Mean for Clinical Practice? J Obes Metab Syndr. 2021;30(3):196-
700 208 699 and Bimagrumab: What do They Mean for Clinical Practice? J Obes Metab Syndr. 2021;3
100 208.
701 34. Blois SM, Dveksler G, Vasta GR, Freitag N, Blanchard V, Barrientos G. Pregnancy 701
702 702 Galectinology: Insights Into a Complex Network of Glycan Binding Proteins. (1664-3224 (Electronic)).
703 35. Schrader S. Unverdorben L. Hutter S. Knabl J. Schmoeckel E. Meister S. et al. Overexpression 703 35. Schrader S, Unverdorben L, Hutter S, Knabl J, Schmoeckel E, Meister S, et al. Overexpression
704 of galectin-4 in placentas of women with gestational diabetes. (1872-7603 (Electronic)). 704 of galectin-4 in placentas of women with gestational diabetes. (1872-7603 (Electronic)).
705 36. Aiavi F. Kongoasa N. Gaffev T. Asmann YW. Watson WJ. Baldi A. et al. Elevated e 705 36. Ajayi F, Kongoasa N, Gaffey T, Asmann YW, Watson WJ, Baldi A, et al. Elevated expression of
706 serine protease HtrA1 in preeclampsia and its role in trophoblast cell migration and invasion. Am J 706 serine protease HtrA1 in preeclampsia and its role in trophoblast cell migration and invasion. Am J
707 Obstet Gynecol. 2008;199(5):557.e1-10. 707 Obstet Gynecol. 2008;199(5):557.e1-10.
708 37 Mukojević K. Šolijć V. Martinović V. Raguž E. Filipović N. The Uhiquitin Associated and SH3. yol. 2008. – Makojević K, Šoljić V, Martinović
709. – Domain-Containing Proteins (UBASH3) Es 708 37. Vukojević K, Šoljić V, Martinović V, Raguž F, Filipović N. The Ubiquitin-Associated and SH3
709 Domain-Containing Proteins (UBASH3) Family in Mammalian Development and Immune Response.
710 Int J Mol Sci. 2024;25(3) 710 Int J Mol Sci. 2024;25(3).
711 - 38 - Pena LD, van Calcar SC. Hansen L Edick ML Walsh Vockley C. Leslie N. et al. Qutcomes and 711 38. Pena LD, van Cald
712 - genotvne-phenotyne corr 711 38. Pena LD, van Calcar SC, Hansen J, Edick MJ, Walsh Vockley C, Leslie N, et al. Outcomes ar
712 genotype-phenotype correlations in 52 individuals with VLCAD deficiency diagnosed by NBS and
713 enrolled in the IBEM-IS 713 enrolled in the IBEM-IS database. Mol Genet Metab. 2016;118(4):272-81.
714 39. Hoggard N. Hunter L. Duncan JS. Williams LM. Trayhurn P. Mercer 714 39. Hoggard N, Hunter L, Duncan JS, Williams LM, Trayhurn P, Mercer JG. Leptin and leptin
715 enceptor mRNA and protein expression in the murine fetus and placenta. Proceedings 715 receptor mRNA and protein expression in the murine fetus and placenta. Proceedings of 716
716 the National Academy of Sciences. 1997;94(20):11073-8. 716 the National Academy of Sciences. 1997;94(20):11073-8.
717 - 20 - Kilneläinen TO, Carli JEM, Skowronski AA, Sun O, Kriehel J, Feitosa ME, et al. Genome-wide 1997 — Martin Martin Academy of Sciences. 1997;94(20):2001
717 — 40. — Kilpeläinen TO, Carli JFM, Skowronski AA, Sun Q,
718. — meta-analysis uncovers novel loci influencing circulating l 717 40. Kilpeläinen TO, Carli JFM, Skowronski AA, Sun Q, Kriebel J, Feitosa MF, et al. Genome-w
718 meta-analysis uncovers novel loci influencing circulating leptin levels. Nature Communications.
719 2016;7(1):10494. 719 2016;7(1):10494. 720 41. Agrawal S, Cerdeira AS, Redman C, Vatish M. Meta-Analysis and Systematic Review to Assess 721 the Role of Soluble FMS-Like Tyrosine Kinase-1 and Placenta Growth Factor Ratio in Prediction of
722 The Role of Soluble FMS-Like Tyrosine Kinase-1 and Placenta Growth Factor Ratio in Prediction of 722 Preeclampsia: The SaPPPhirE Study. (1524-4563 (Electronic)).
723 december - Andrikos AA-OX, Andrikos D. Schmidt B. Birdir C. Kimi 723 42. Andrikos AA-OX, Andrikos D, Schmidt B, Birdir C, Kimmig R, Gellhaus A, et al. Course of the 724 sFlt-1/PIGF ratio in fetal growth restriction and correlation with biometric measurements, feto-724 sFlt-1/PIGF ratio in fetal growth restriction and correlation with biometric measurements, feto-
725 maternal Doppler parameters and time to delivery. (1432-0711 (Electronic)). 726 43. Chang Y-S, Chen C-N, Jeng S-F, Su Y-N, Chen C-Y, Chou H-C, et al. The sFlt-1/PIGF ratio a
727 medictor for poor pregnancy and neonatal outcomes. Pediatrics & Neonatology, 2017:58(6):52 727 predictor for poor pregnancy and neonatal outcomes. Pediatrics & Neonatology. 2017;58(6):529-33.
728 44. Suhre K, McCarthy MI, Schwenk JM. Genetics meets proteomics: perspectives for large 728 44. Suhre K, McCarthy MI, Schwenk JM. Genetics meets proteomics: perspectives for large
729 population-based studies. Nature Reviews Genetics. 2021:22(1):19-37. population-based studies. Nature Reviews Genetics. 2021;22(1):19-37. 729 population-based studies. Nature Reviews Genetics. 2021;22(1):19-37.

730 45. Fauman EB, Hyde C. An optimal variant to gene distance window derived from an empirical
731 definition of cis and trans protein QTLs. BMC Bioinformatics. 2022;23(1):169.

731 definition of cis and trans protein QTLs. BMC Bioinformatics. 2022;23(1):169.

 732 46. Zhao J, Stewart ID, Baird D, Mason D, Wright J, Zheng J, et al. Causal effects of maternal effects of Λ 733 circulating amino acids on offspring birthweight: a Mendelian randomisation study. (2352-3964

(Electronic)).

⁷³⁵**Supplementary**

736

737

738

- 739 Supplementary figure 1 a workflow of instruments (pQTL) used in our analysis.
740 Abbreviations: pQTL, protein quantitative trait loci
- Abbreviations: pQTL, protein quantitative trait loci

742 Supplementary figure 2 - Genetic associations with NEC1 and offspring birth weight
743 in maternal analysis. Each data point represents one genetic variant. The purple 743 in maternal analysis. Each data point represents one genetic variant. The purple 744 diamond represents the selected pQTL. Colours indicate the $R²$ values for strength diamond represents the selected pQTL. Colours indicate the R^2 values for strength 745 of linkage disequilibrium between the genetic variant and the pQTL. In the left panel. 745 of linkage disequilibrium between the genetic variant and the pQTL. In the left panel, 746 the plot shows the correlation between log 10 p-values for the genetic association 746 the plot shows the correlation between log 10 p-values for the genetic association
747 with protein levels (x axis) and birth weight (y axis). The right panels show 747 with protein levels (x axis) and birth weight (y axis). The right panels show
 748 recombination plots for the protein (top panel) and birth weight (bottom panel) with 748 recombination plots for the protein (top panel) and birth weight (bottom panel) with 749 log10 p-values in the y axis and chromosome positions in the x axis. log10 p-values in the y axis and chromosome positions in the x axis.

750 751 Supplementary figure 3 - Genetic associations with Galectin_4 and offspring birth
752 weight in maternal analysis. Each data point represents one genetic variant. The 752 weight in maternal analysis. Each data point represents one genetic variant. The 753 purple diamond represents the selected pQTL. Colours indicate the $R²$ values for purple diamond represents the selected pQTL. Colours indicate the R^2 values for

754 strength of linkage disequilibrium between the genetic variant and the pQTL. In the 755 left panel, the plot shows the correlation between log 10 p-values for the genetic 755 left panel, the plot shows the correlation between log 10 p-values for the genetic
756 association with protein levels (x axis) and birth weight (y axis). The right panels 756 association with protein levels (x axis) and birth weight (y axis). The right panels
757 show recombination plots for the protein (top panel) and birth weight (bottom panel) 757 show recombination plots for the protein (top panel) and birth weight (bottom panel)
758 with log10 p-values in the y axis and chromosome positions in the x axis.

759 760 Supplementary figure 4 - Genetic associations with PRS57 and offspring birth weight
761 in maternal analysis. Each data point represents one genetic variant. The purple 761 in maternal analysis. Each data point represents one genetic variant. The purple 762 diamond represents the selected pQTL. Colours indicate the $R²$ values for strength diamond represents the selected pQTL. Colours indicate the R^2 values for strength 763 of linkage disequilibrium between the genetic variant and the pQTL. In the left panel, 763 of linkage disequilibrium between the genetic variant and the pQTL. In the left panel,
764 the plot shows the correlation between log 10 p-values for the genetic association 764 the plot shows the correlation between log 10 p-values for the genetic association
765 with protein levels (x axis) and birth weight (y axis). The right panels show 765 with protein levels $(x \text{ axis})$ and birth weight $(y \text{ axis})$. The right panels show 766 recombination plots for the protein (top panel) and birth weight (bottom panel) with 766 recombination plots for the protein (top panel) and birth weight (bottom panel) with 767 log10 p-values in the y axis and chromosome positions in the x axis. log10 p-values in the y axis and chromosome positions in the x axis.

768 769 Supplementary figure 5 - Genetic associations with PTN9 and offspring birth weight
770 in offspring analyses. Each data point represents one genetic variant. The purple 770 in offspring analyses. Each data point represents one genetic variant. The purple 771 diamond represents the selected pQTL. Colours indicate the $R²$ values for strength diamond represents the selected pQTL. Colours indicate the R^2 values for strength 772 of linkage disequilibrium between the genetic variant and the pQTL. In the left panel, 772 of linkage disequilibrium between the genetic variant and the pQTL. In the left panel, 773 the plot shows the correlation between log 10 p-values for the genetic association 773 the plot shows the correlation between log 10 p-values for the genetic association
774 vith protein levels (x axis) and birth weight (y axis). The right panels show 774 with protein levels $(x \text{ axis})$ and birth weight $(y \text{ axis})$. The right panels show 775 recombination plots for the protein (top panel) and birth weight (bottom panel) with 775 recombination plots for the protein (top panel) and birth weight (bottom panel) with 776 log10 p-values in the y axis and chromosome positions in the x axis. log10 p-values in the y axis and chromosome positions in the x axis.

777
778 778 Supplementary figure 6 - Genetic associations with ULK3 and offspring birth weight
779 in maternal analysis. Each data point represents one genetic variant. The purple 779 in maternal analysis. Each data point represents one genetic variant. The purple
780 diamond represents the selected pOTL. Colours indicate the R² values for strepgth 780 diamond represents the selected pQTL. Colours indicate the R^2 values for strength R^2 represents the post of linkage disequilibrium between the genetic variant and the pOTL. In the left panel 781 of linkage disequilibrium between the genetic variant and the pQTL. In the left panel,
782 the plot shows the correlation between log 10 p-values for the genetic association 782 the plot shows the correlation between log 10 p-values for the genetic association
783 with protein levels (x axis) and birth weight (y axis). The right panels show 783 with protein levels (x axis) and birth weight (y axis). The right panels show
784 recombination plots for the protein (top panel) and birth weight (bottom panel) with 784 recombination plots for the protein (top panel) and birth weight (bottom panel) with 785 log10 p-values in the y axis and chromosome positions in the x axis. log10 p-values in the y axis and chromosome positions in the x axis.

786
787 787 Supplementary figure 7 - Genetic associations with NEC1 and offspring birth weight
788 in offspring analysis. Each data point represents one genetic variant. The purple 788 in offspring analysis. Each data point represents one genetic variant. The purple 789 diamond represents the selected pQTL. Colours indicate the $R²$ values for strength diamond represents the selected pQTL. Colours indicate the R^2 values for strength 790 of linkage disequilibrium between the genetic variant and the pQTL. In the left panel, 790 of linkage disequilibrium between the genetic variant and the pQTL. In the left panel, 791 the plot shows the correlation between log 10 p-values for the genetic association 791 the plot shows the correlation between log 10 p-values for the genetic association
792 with protein levels $(x \text{ axis})$ and birth weight $(y \text{ axis})$. The right panels show 792 with protein levels $(x \text{ axis})$ and birth weight $(y \text{ axis})$. The right panels show 793 recombination plots for the protein (top panel) and birth weight (bottom panel) with 793 recombination plots for the protein (top panel) and birth weight (bottom panel) with 794 log10 p-values in the y axis and chromosome positions in the x axis. log10 p-values in the y axis and chromosome positions in the x axis.

796 Supplementary figure 8 - Genetic associations with Galectin_4 and offspring birth
797 weight in offspring analysis. Each data point represents one genetic variant. The 797 weight in offspring analysis. Each data point represents one genetic variant. The 798 purple diamond represents the selected pQTL. Colours indicate the $R²$ values for purple diamond represents the selected pQTL. Colours indicate the R^2 values for

799 strength of linkage diseguilibrium between the genetic variant and the pQTL. In the 799 strength of linkage disequilibrium between the genetic variant and the pQTL. In the 300 left panel, the plot shows the correlation between log 10 p-values for the genetic 800 left panel, the plot shows the correlation between log 10 p-values for the genetic
801 association with protein levels (x axis) and birth weight (y axis). The right panels 801 association with protein levels (x axis) and birth weight (y axis). The right panels 802 show recombination plots for the protein (top panel) and birth weight (bottom panel) 802 show recombination plots for the protein (top panel) and birth weight (bottom panel)
803 with log10 p-values in the y axis and chromosome positions in the x axis. with log10 p-values in the y axis and chromosome positions in the x axis.

805 Supplementary figure 9 - Genetic associations with Cystatin_c and offspring birth
806 weight in offspring analysis. Each data point represents one genetic variant. The 806 weight in offspring analysis. Each data point represents one genetic variant. The 807 purple diamond represents the selected pQTL. Colours indicate the R^2 values for so purple diamond represents the selected pQTL. Colours indicate the R^2 values for so strength of linkage disequilibrium between the genetic variant and the pQTL. In the 808 strength of linkage disequilibrium between the genetic variant and the pQTL. In the 809 left panel, the plot shows the correlation between log 10 p-values for the genetic 809 left panel, the plot shows the correlation between log 10 p-values for the genetic
810 association with protein levels (x axis) and birth weight (y axis). The right panels 810 association with protein levels (x axis) and birth weight (y axis). The right panels 811 show recombination plots for the protein (top panel) and birth weight (bottom panel) 811 show recombination plots for the protein (top panel) and birth weight (bottom panel)
812 with log10 p-values in the v axis and chromosome positions in the x axis. with $log10$ p-values in the y axis and chromosome positions in the x axis.

814 Supplementary figure 10 - Genetic associations with ACADV and offspring birth
815 weight in offspring analysis. Each data point represents one genetic variant. The 815 weight in offspring analysis. Each data point represents one genetic variant. The 816 purple diamond represents the selected pQTL. Colours indicate the R^2 values for s as purple diamond represents the selected pQTL. Colours indicate the R^2 values for strength of linkage disequilibrium between the genetic variant and the pQTL. In the 817 strength of linkage disequilibrium between the genetic variant and the pQTL. In the 818 left panel, the plot shows the correlation between log 10 p-values for the genetic 818 left panel, the plot shows the correlation between log 10 p-values for the genetic 819 association with protein levels (x axis) and birth weight (y axis). The right panels 819 association with protein levels (x axis) and birth weight (y axis). The right panels 820 show recombination plots for the protein (top panel) and birth weight (bottom panel) 820 show recombination plots for the protein (top panel) and birth weight (bottom panel)
821 with log10 p-values in the y axis and chromosome positions in the x axis. with log10 p-values in the y axis and chromosome positions in the x axis.

822
823 823 Supplementary figure 11 - Genetic associations with BCMA and offspring birth 824 weight in offspring analysis. Each data point represents one genetic variant. The 824 weight in offspring analysis. Each data point represents one genetic variant. The 825 purple diamond represents the selected pQTL. Colours indicate the R^2 values for s as purple diamond represents the selected pQTL. Colours indicate the R^2 values for strength of linkage disequilibrium between the genetic variant and the pQTL. In the 826 strength of linkage disequilibrium between the genetic variant and the pQTL. In the 827 left panel, the plot shows the correlation between log 10 p-values for the genetic 827 left panel, the plot shows the correlation between log 10 p-values for the genetic 828 association with protein levels (x axis) and birth weight (y axis). The right panels 828 association with protein levels (x axis) and birth weight (y axis). The right panels
829 show recombination plots for the protein (top panel) and birth weight (bottom panel) 829 show recombination plots for the protein (top panel) and birth weight (bottom panel)
830 with log10 p-values in the y axis and chromosome positions in the x axis. with $log10$ p-values in the y axis and chromosome positions in the x axis.

831
832 832 Supplementary figure 12 - Genetic associations with TWEAK and offspring birth
833 weight in offspring analysis. Each data point represents one genetic variant. The 833 weight in offspring analysis. Each data point represents one genetic variant. The 834 purple diamond represents the selected pQTL. Colours indicate the R^2 values for purple diamond represents the selected pQTL. Colours indicate the R^2 values for 835 strength of linkage disequilibrium between the genetic variant and the pQTL. In the 835 strength of linkage disequilibrium between the genetic variant and the pQTL. In the 836 left panel, the plot shows the correlation between log 10 p-values for the genetic 836 left panel, the plot shows the correlation between log 10 p-values for the genetic
837 association with protein levels (x axis) and birth weight (y axis). The right panels 837 association with protein levels (x axis) and birth weight (y axis). The right panels
838 show recombination plots for the protein (top panel) and birth weight (bottom panel) 838 show recombination plots for the protein (top panel) and birth weight (bottom panel)
839 with log10 p-values in the y axis and chromosome positions in the x axis. with log10 p-values in the y axis and chromosome positions in the x axis.

841 Supplementary figure 13 - Genetic associations with FUT5 and offspring birth weight
842 in offspring analysis. Each data point represents one genetic variant. The purple 842 in offspring analysis. Each data point represents one genetic variant. The purple 843 diamond represents the selected pQTL. Colours indicate the $R²$ values for strength diamond represents the selected pQTL. Colours indicate the R^2 values for strength 844 of linkage disequilibrium between the genetic variant and the pQTL. In the left panel. 844 of linkage disequilibrium between the genetic variant and the pQTL. In the left panel,
845 the plot shows the correlation between log 10 p-values for the genetic association 845 the plot shows the correlation between log 10 p-values for the genetic association
846 with protein levels (x axis) and birth weight (y axis). The right panels show 846 with protein levels $(x \text{ axis})$ and birth weight $(y \text{ axis})$. The right panels show 847 recombination plots for the protein (top panel) and birth weight (bottom panel) with 847 recombination plots for the protein (top panel) and birth weight (bottom panel) with 848 log10 p-values in the y axis and chromosome positions in the x axis. log10 p-values in the y axis and chromosome positions in the x axis.

858
859 859 Supplementary figure 15 - Genetic associations with PLCG1 and offspring birth
860 weight in offspring analysis. Each data point represents one genetic variant. The 860 weight in offspring analysis. Each data point represents one genetic variant. The 861 purple diamond represents the selected pQTL. Colours indicate the $R²$ values for s as purple diamond represents the selected pQTL. Colours indicate the R^2 values for sequently selected pQTL. In the sequently set of the sequently set $\frac{1}{2}$ and the pQTL. In the 862 strength of linkage disequilibrium between the genetic variant and the pQTL. In the 863 left panel, the plot shows the correlation between log 10 p-values for the genetic 863 left panel, the plot shows the correlation between log 10 p-values for the genetic
864 association with protein levels (x axis) and birth weight (y axis). The right panels 864 association with protein levels (x axis) and birth weight (y axis). The right panels
865 show recombination plots for the protein (top panel) and birth weight (bottom panel) 865 show recombination plots for the protein (top panel) and birth weight (bottom panel)
866 with log10 p-values in the v axis and chromosome positions in the x axis. with log10 p-values in the y axis and chromosome positions in the x axis.

868 Supplementary figure 16 - Genetic associations with MK13 and offspring birth weight
869 in offspring analysis. Each data point represents one genetic variant. The purple 869 in offspring analysis. Each data point represents one genetic variant. The purple 870 diamond represents the selected pQTL. Colours indicate the R^2 values for strength 870 diamond represents the selected pQTL. Colours indicate the R^2 values for strength 871 of linkage disequilibrium between the genetic variant and the pQTL. In the left panel, 871 of linkage disequilibrium between the genetic variant and the pQTL. In the left panel, $\frac{1}{2}$ the plot shows the correlation between log 10 p-values for the genetic association 872 the plot shows the correlation between log 10 p-values for the genetic association
873 with protein levels (x axis) and birth weight (y axis). The right panels show 873 with protein levels $(x \text{ axis})$ and birth weight $(y \text{ axis})$. The right panels show 874 recombination plots for the protein (top panel) and birth weight (bottom panel) with 874 recombination plots for the protein (top panel) and birth weight (bottom panel) with 875 log10 p-values in the y axis and chromosome positions in the x axis. log10 p-values in the y axis and chromosome positions in the x axis.

876

877
878 878 Supplementary figure 18 - Genetic associations with AIF1L and offspring birth weight
879 in offspring analysis. Each data point represents one genetic variant. The purple 879 in offspring analysis. Each data point represents one genetic variant. The purple
880 diamond represents the selected pOTL. Colours indicate the R² values for strepgth 880 diamond represents the selected pQTL. Colours indicate the R^2 values for strength $\frac{1}{2}$ of linkage disequilibrium between the genetic variant and the pOTL. In the left panel 881 of linkage disequilibrium between the genetic variant and the pQTL. In the left panel, 882 the plot shows the correlation between log 10 p-values for the genetic association 882 the plot shows the correlation between log 10 p-values for the genetic association 883 with protein levels $(x \text{ axis})$ and birth weight $(y \text{ axis})$. The right panels show 883 with protein levels (x axis) and birth weight (y axis). The right panels show
884 recombination plots for the protein (top panel) and birth weight (bottom panel) with 884 recombination plots for the protein (top panel) and birth weight (bottom panel) with 885 log10 p-values in the y axis and chromosome positions in the x axis. log10 p-values in the y axis and chromosome positions in the x axis.

886 887 Supplementary figure 19 - Genetic associations with sLeptin_R and offspring birth 888 weight in offspring analysis. Each data point represents one genetic variant. The 888 weight in offspring analysis. Each data point represents one genetic variant. The 889 purple diamond represents the selected pQTL. Colours indicate the R^2 values for sse purple diamond represents the selected pQTL. Colours indicate the R^2 values for selected position.
890 strength of linkage disequilibrium between the genetic variant and the pQTL. In the 890 strength of linkage disequilibrium between the genetic variant and the pQTL. In the 891 left panel, the plot shows the correlation between log 10 p-values for the genetic 891 left panel, the plot shows the correlation between log 10 p-values for the genetic
892 association with protein levels (x axis) and birth weight (y axis). The right panels 892 association with protein levels (x axis) and birth weight (y axis). The right panels
893 show recombination plots for the protein (top panel) and birth weight (bottom panel) 893 show recombination plots for the protein (top panel) and birth weight (bottom panel)
894 with log10 p-values in the y axis and chromosome positions in the x axis. with log10 p-values in the y axis and chromosome positions in the x axis.

-
- 896
-
- 898
-
-
- 901
-
-
-
- 904
-
-

907

908 Supplementary figure 20 - Fetal single cell gene expression expressed in the
909 pancreas for NEC1, where the organs are shown in yellow and the single cells 909 pancreas for NEC1, where the organs are shown in yellow and the single cells 910 demonstrating expression in navy blue. demonstrating expression in navy blue.

911

912 Supplementary figure 22 - Fetal single cell gene expression expressed tissue-wide
913 for Galectin 4, where the organs are shown in yellow and the single cells 913 for Galectin_4, where the organs are shown in yellow and the single cells 914 demonstrating expression in navy blue. demonstrating expression in navy blue.

915

916

- 917 Supplementary figure 23 Fetal single cell gene expression expressed tissue-wide
918 for FUT5, where the organs are shown in yellow and the single cells demonstrating
- 918 for FUT5, where the organs are shown in yellow and the single cells demonstrating
919 expression in navy blue.
- expression in navy blue.

920

- 921 Supplementary figure 24 Fetal single cell gene expression expressed tissue-wide
922 for UBS3B, where the organs are shown in vellow and the single cells demonstrating
- 922 for UBS3B, where the organs are shown in yellow and the single cells demonstrating
923 expression in navy blue.
- expression in navy blue.

924

925 Supplementary figure 25 - Fetal single cell gene expression expressed tissue-wide
926 for sLeptin_R, where the organs are shown in yellow and the single cells 926 for sLeptin_R, where the organs are shown in yellow and the single cells 927 demonstrating expression in navy blue. demonstrating expression in navy blue.

928

929 Supplementary figure 26 - Fetal single cell gene expression expressed tissue-wide
930 for AIF1L, where the organs are shown in yellow and the single cells demonstrating

930 for AIF1L, where the organs are shown in yellow and the single cells demonstrating
931 expression in navy blue. expression in navy blue.

932

- 933 Supplementary figure 27 Fetal single cell gene expression expressed tissue-wide
934 for PLCG1, where the organs are shown in yellow and the single cells demonstrating
- 934 for PLCG1, where the organs are shown in yellow and the single cells demonstrating
935 expression in navy blue.
- expression in navy blue.

936

937 Supplementary figure 28 - Fetal single cell gene expression expressed tissue-wide
938 for IFN_a_b_R1, where the organs are shown in yellow and the single cells 938 for IFN_a_b_R1, where the organs are shown in yellow and the single cells
939 demonstrating expression in navy blue. demonstrating expression in navy blue.

940
941

- 941 Supplementary figure 29 Fetal single cell gene expression expressed tissue-wide
942 for ACADV, where the organs are shown in yellow and the single cells demonstrating
- 942 for ACADV, where the organs are shown in yellow and the single cells demonstrating
943 expression in navy blue. expression in navy blue.
- 944
- 945
- 946

948 Supplementary figure 22 - Fetal single cell gene expression expressed tissue-wide
949 for sLeptin R, where the organs are shown in yellow and the single cells 949 for sLeptin_R, where the organs are shown in yellow and the single cells 950 demonstrating expression in navy blue. demonstrating expression in navy blue.