medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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 Arab[2], Tom A. Bond[1,2,11], Jie Zheng[5,6], Maria C. Magnus[7], Elizabeth C. Corfield[8,9], Gemma L Clayton[1,2], Liang-Dar Hwang[11], Robin N. Beaumont[10], David M. Evans[1,4,11], Rachel M. Freathy[10], Tom R. Gaunt[1,2], Deborah A Lawlor[1,2], Maria Carolina Borges[1,2]

[1]MRC Integrative Epidemiology Unit at the University of Bristol,
[2]Population Health Science, Bristol Medical School, University of Bristol
[3]MRC Unit for Lifelong Health and Ageing, University College London
[4]Frazer Institute, University of Queensland, Brisbane, QLD, Australia
[5]Department 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
[8]Nic Waals Institute, Lovisenberg Diaconal Hospital, Oslo, Norway
[9]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
[11]Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia

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

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Abstract

Fetal growth is an indicator of fetal survival, regulated by maternal and fetal factors, but little is known about the underlying molecular mechanisms. We used Mendelian randomization to explore the effects of maternal and fetal genetically-instrumented plasma proteins on birth weight using genome-wide association summary data (n=406,063 with maternal and/or fetal genotype), with independent replication (n=74,932 mothers and n=62,108 offspring), and colocalisation. Higher genetically-predicted maternal levels of PCSK1 increased birthweight (mean-difference: 9g (95% CI: 5g, 13g) per 1 standard deviation protein level). Higher maternal levels of LGALS4 decreased birthweight (-54g (-29g, -80g)), as did VCAM1, RAD51D and GP1BA. In the offspring, higher genetically-predicted fetal levels of LGALS4 (46g (23g, 70g)) increased birthweight, alongside FCGR2B. Higher offspring levels of PCSK1 decreased birth weight (-9g (-16g, 4g), alongside LEPR. Results support maternal and fetal protein effects on birth weight, implicating roles for glucose metabolism, energy homeostasis, endothelial function and adipocyte differentiation.

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Introduction

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

Proteins play a major role in a wide range of biological processes (11), are essential for cellular growth and repair, and could provide key insights into maternal and fetal molecular mechanisms regulating fetal growth (10). Mendelian Randomization (MR) is a method which uses genetic variants associated with exposures to test the effect of those exposures on human traits and diseases, aiming to mitigate biased causal effects due to residual confounding and reverse causation that may explain conventional observational associations. Previous MR studies have established causality of well-known maternal modifiable factors on offspring birth weight, including higher maternal general adiposity and circulating glucose levels on higher birth weight, and smoking and higher blood pressure on lower birth weight (12) (13, 14) (5) (15). They have also provided novel insights, such as a potential association of higher maternal glutamine levels with higher birth weight (5) and metabolically favourable adiposity with lower birth weight (16). MR and proteomics can be integrated to explore maternal and fetal proteins regulating fetal growth by using genetic variants associated with protein levels (i.e., protein quantitative trait loci - pQTL) as instrumental variables.

The aim of this study was to identify causal effects of maternal and fetal proteins on birth weight to highlight potential molecular mechanisms. To achieve that, we used two-sample MR, in which large-scale GWAS of plasma proteomics and birth weight were combined to examine the causal effect of 996 plasma proteins on offspring birth weight.

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Results

We investigated the maternal and fetal effects of 996 plasma proteins on offspring birthweight using two-sample MR, examining both maternal and fetal genetic effects (n=406,063 with maternal and/or fetal genotype). For proteins passing multiple testing correction (FDR p-value < 5%), we tested whether results replicated in an independent sample (n=74,932 mothers and n=62,108 offspring) and investigated whether MR results could be confounded by linkage disequilibrium using genetic colocalisation. For proteins which had discovery and replication evidence, and no evidence against colocalisation, we used protein network interactions to examine plausible pathways through which they could be influencing birth weight. **Figure 1** summarises the analytical workflow used in this paper, with full details provided in the methods section.

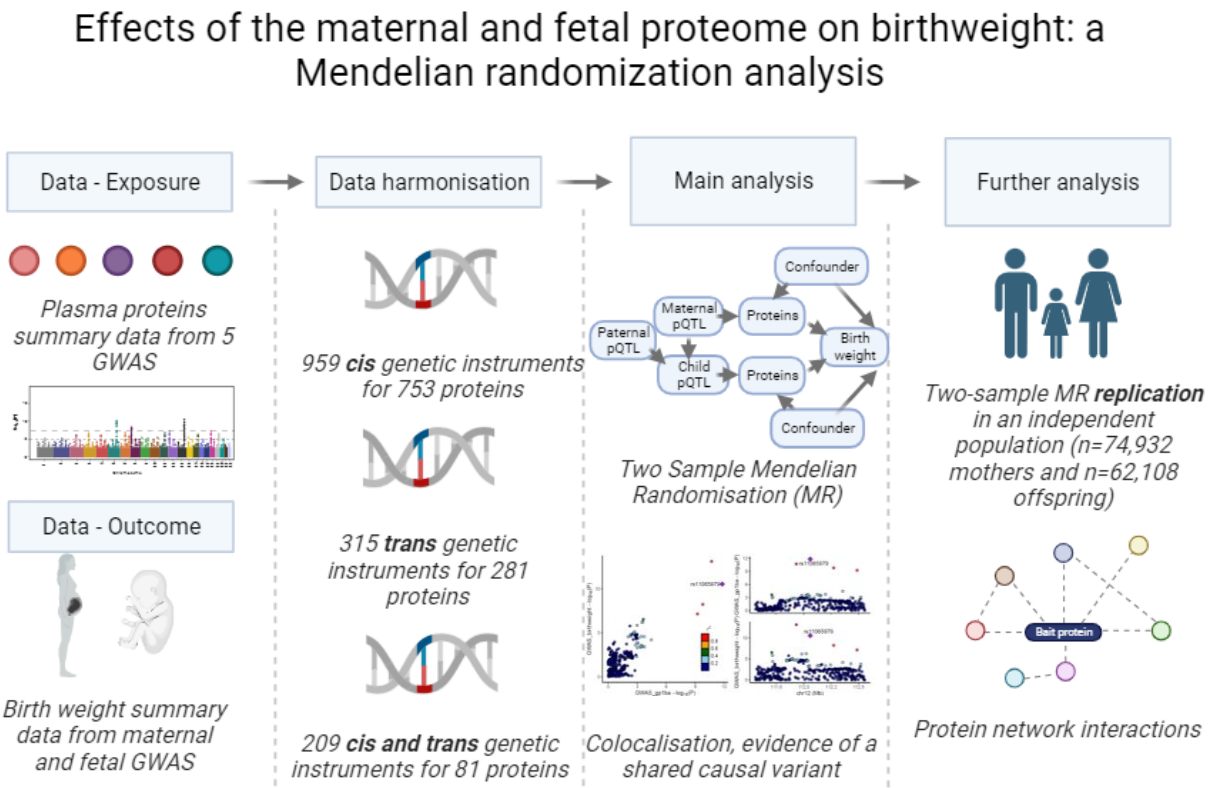


Figure 1 – Schematic representation of the analyses conducted in this paper. Abbreviations: GWAS, genome wide association studies; pQTL, protein quantitative trait loci

Selecting genetic variants to instrument for maternal and fetal plasma proteins

We selected genetic instruments for proteins from a cross-platform meta-analysis, which combined results from five GWAS of plasma protein levels (17-21). These were used as the

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

source of pQTLs data for both maternal and fetal analyses. After excluding pQTL that did not pass quality control, (as detailed in the **Methods** section), we retained genetic instruments (i.e. protein quantitative trait loci (pQTL)) for 996 unique plasma proteins with strong evidence of association with relevant proteins ($p<5x10^{-8}$) (22). These proteins were instrumented by cis-acting pQTL (N=959 pQTL for 753 proteins), trans-acting pQTL (N=315 for 281 proteins), or a combination of both cis-and-trans pQTL (N=209 pQTL for 81 proteins) (further details in **Methods** and **Supplementary Figure 1**). Cis-pQTL, defined in our analyses as pQTL located within ±500kb of the protein encoding gene, are more likely to be specific and credible instruments for MR analyses. In contrast, trans-pQTL (outside this genomic region) can give complementary information to cis-pQTL as they are less likely to emerge due to epitope binding artifacts with the target protein (23), although they present a higher risk of pleiotropy (often affecting levels of multiple proteins). **Supplementary Table 1** provides a list of proteins and pQTL included in our analyses.

To explore maternal and fetal effects, we extracted genetic association data on offspring birth weight for the selected pQTLs for both maternal and fetal genetic effects from the EGG Consortium and UK Biobank meta-analysis. Maternal and fetal genetic association data were adjusted for the correlation between maternal and fetal genotype (3). These data were harmonized with the selected exposure pQTL data, previously derived from 5 GWAS (11), for both maternal and fetal analyses (further details in **Methods**).

Exploring maternal circulating proteins influencing offspring birth weight

We used two-sample MR to estimate effects of genetically instrumented maternal circulating protein levels on birth weight. We found evidence of causal effects of 14 maternal proteins (instrumented by 7 cis-pQTLs, 11 trans-pQTLs and 6 cis-and-trans-pQTLs) on offspring birth weight that passed our 5% false discovery rate (FDR) adjusted p-value **(Figure 2).** Using cis-pQTLs, we found evidence of higher maternal levels of PCSK1, CD59 and GDF15 causing higher birth weight (mean difference in birth weight per 1 standard deviation (SD) higher protein levels: 9g (95% CI: 5g, 13g), 23g (11g, 36g) and 13g (6g, 20g)), respectively, and higher maternal levels of LGALS4, CLMP and PRSS57 causing lower offspring birth weight (-54g (-29, -80), -34g (-16, -54) and -17g (-8g, -26g)), respectively**.** In the trans-pQTL MR analyses, we found evidence that higher maternal levels of CREB3L4, VCAM1, RAD51D, B2M, GP1BA, SEMG2 and RPN1 cause lower birth weight (mean difference -27g (-36g, -17g), -96g (-120g, -72g), -16g (-36g, 5g) , -136g (-174g , -98g), -92g (-119g, -65g), -13g (-21, -5g) and -8g (-12g, -3g), respectively). In cis-and-trans MR analyses, higher levels of maternal RPN1 were related to lower birth weight (-7g (-11g, -2g)) and NPPB to higher birth weight (24g (10g, 37g)) **(Figure 2)**.

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

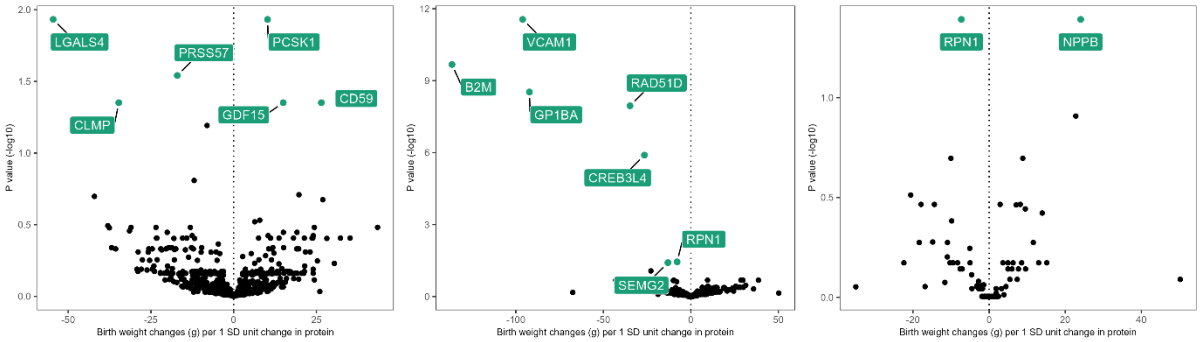


Figure 2 – Two sample MR results using maternal cis (left), trans (middle) and cis-and-trans (right) pQTLs against offspring birth weight. FDR corrected p values transformed to their -log10. Results plotted in green passed 5% FDR corrected p-value threshold. Abbreviations: SD, standard deviation

We explored whether the MR results for the 14 proteins identified in our maternal discovery analyses (5% FDR p-value <0.05) were replicated in an independent sample of 74,932 mothers from the Norwegian Mother, Father and Child Cohort Study (MoBa) (24) (Further details in **Methods**). We considered results to be replicated if they met the following criteria: (i) point estimates from discovery and replication analyses were in the same direction, (ii) the p-value for the meta-analysis pooling discovery and replication results was < 0.05, and (iii) there was no strong evidence supporting heterogeneity between discovery and replication MR estimates (Cochrane's Q p-value > 0.05). We found that MR results replicated for all maternal proteins identified in discovery analyses in the subsequent meta-analysis (**Figure 3** and **Supplementary Table 7**). However, for some proteins passing our replication criteria, i.e. PCSK1 and CD59, effect estimates were close to the null in the replication sample.

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

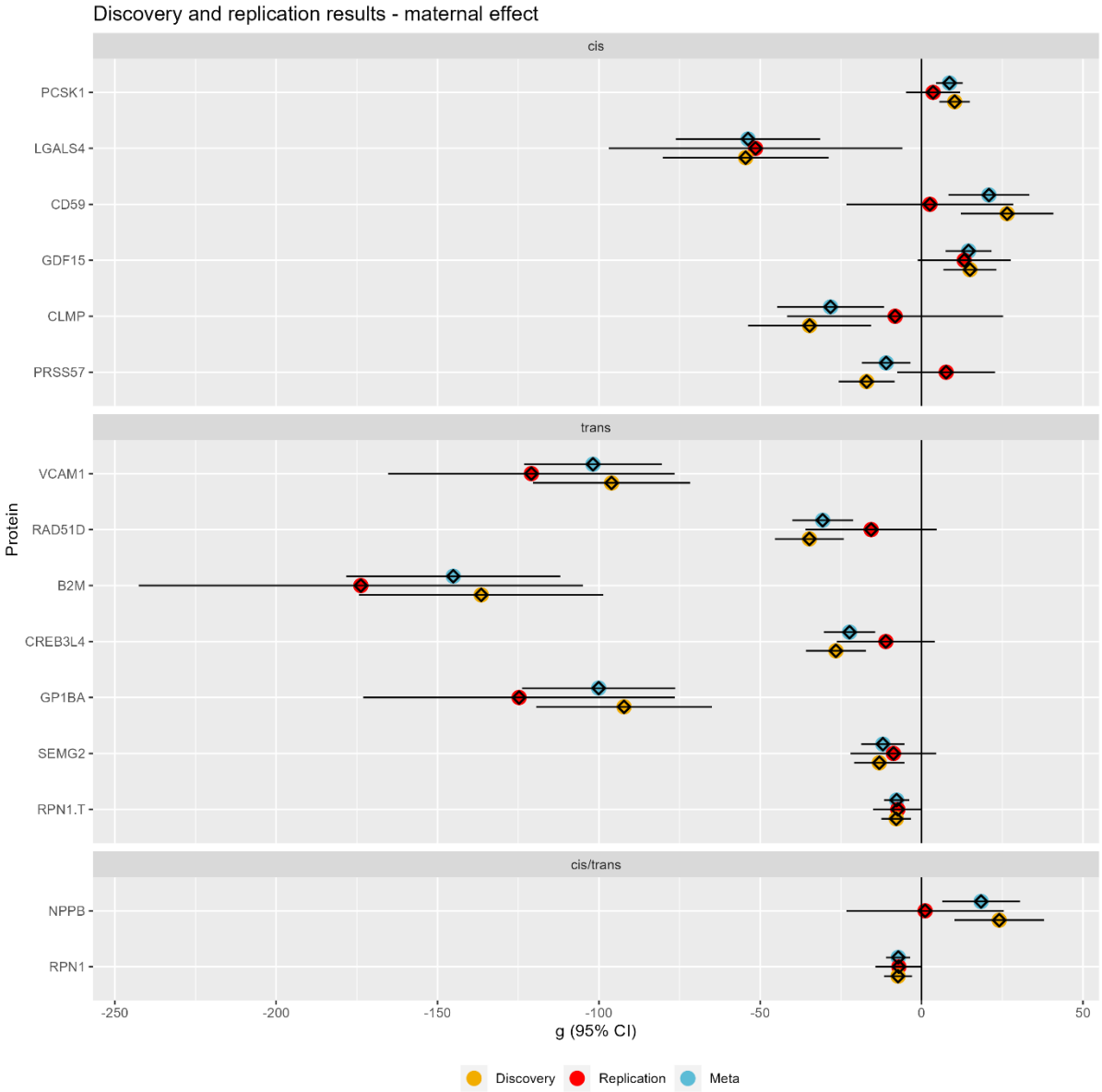


Figure 3 Mendelian randomization analyses estimating the effect of genetically instrumented maternal proteins (in standard deviation units) on offspring birth weight (in grams) in the discovery (N=210,248 mothers), replication (N=87,651 mothers) and meta-analyses of discovery and replication samples. Replication was tested for proteins passing FDR-corrected p-value <0.05 in the discovery sample.

We also explored whether the MR results for the 14 maternal proteins could be biased due confounding by linkage disequilibrium (LD). This could happen if a selected pQTL is in LD with another genetic variant influencing birth weight via another biological mechanism. Colocalisation was performed for 11 of the 14 proteins due to availability of protein GWAS summary statistics (summary statistics were not publicly available from the GWAS we used for PRSS57, B2M or NPPB at the time of the study). For proteins with more than one independent instrument, we used the pQTL with the strongest association p-value with the

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

corresponding protein. Results are shown in **Table 1** as posterior probabilities (PP) of no association (H_0), association with protein only (H_1), birth weight only (H_2), both traits owing to distinct causal variants (H_3), or both traits owing to a shared causal variant (H_4) (25). The colocalisation analyses provided strong evidence supporting a shared causal variant for maternal circulating VCAM1, RAD51D, GP1BA, PCSK1, LGALS4, and CREB3L4 (PP for H_4 1.00, 0.99, 1.00, 0.81, 0.82, and 0.73, respectively) with offspring birth weight **(Table 1)**. For CD59, CLMP, GDF15, SEMG2, and RPN1, evidence for a shared genetic variant was uncertain (PP for H_4 <54% with PP for H_1 >39%) (**Table 1** and **Supplementary Figures 2-11**.

Table 1 – Genetic colocalisation results for maternal proteins with birth weight

	Protein	Cis or trans	PP H0	PP H1	PP H2	PP H3	PP H4
Maternal effect	PCSK1	Cis	0	0.06	0	0.12	0.81
	LGALS4	Cis	0	0.07	0	0.11	0.82
	CD59	Cis	0	0.39	0	0.08	0.54
	CLMP	Cis	0	0.50	0	0.05	0.45
	GDF15	Cis	0	0.44	0	0.07	0.48
	VCAM1	Trans	0	0	0	0	1.00
	RAD51D	Trans	0	0	0	0.01	0.99
	GP1BA	Trans	0	0	0	0	1.00
	CREB3L4	Trans	0	0.02	0	0.27	0.73
	RPN1	Trans	0	0.96	0	0	0.03
	SEMG2	Trans	0	0.88	0	0.04	0.07

Abbreviations – PP, posterior probabilities; H, hypothesis. Results are expressed as posterior probabilities (PP) of no association (H0), association with protein only (H1), birth weight only (H2), both traits owing to distinct causal variants (H3), or both traits owing to a shared causal variant (H4). For proteins with more than one instrument (GDF15, CREB3L4, SEMG2, RPN1 in trans, NPPB, and RPN1 in cis and trans, we took the pQTL with the strongest association p value for colocalisation testing)

Conclusions based on maternal results from discovery, replication MR and colocalisation of maternal plasma proteins on offspring birth weight are summarised in **Table 3**. We have strong evidence that 6 out of 14 maternal associations identified in discovery analyses also replicate and have strong evidence of colocalisation, including two proteins instrumented by cis-pQTL (PCSK1, LGALS4) and four proteins instrumented by trans-pQTL (VCAM1, RAD51D, GP1BA, CREB3L4). Results for 5 maternal proteins identified at discovery stage replicated but had uncertain evidence from colocalisation, possibly due to low power, including three proteins instrumented by cis-pQTL (CD59, GDF15 and CLMP) and two instrumented by trans-pQTL or a combination of trans- and cis-pQTL (SEMG2, RPN1) (**Table 3**).

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Table 3 – Summary of results from analyses exploring the effect of maternal proteins on offspring birth weight

Protein	Uniprot ID	pQTL	Discovery MR direction of effect	Replicated*	Colocalisation
PCSK1	P29120	cis	+	Yes	Yes
LGALS4	P56470	cis	-	Yes	Yes
CD59	P13987	cis	+	Yes	Uncertain
GDF15	Q99928	cis	+	Yes	Uncertain
CLMP	Q9H6B4	cis	-	Yes	Uncertain
PRSS57	Q6UWY2	cis	-	No	NA
VCAM1	P19320	trans	-	Yes	Yes
RAD51D	O75771	trans	-	Yes	Yes
B2M	P61769	trans	-	Yes	NA
CREB3L4	Q8TEY5	trans	-	Yes	Yes
GP1BA	P07359	trans	-	Yes	Yes
SEMG2	Q02383	trans	-	Yes	Uncertain
RPN1	P04843	trans	-	Yes	Uncertain
NPPB	P16860	cis and trans	+	Yes	NA
RPN1	P04843	cis and trans	-	Yes	Uncertain

*Replication in an independent cohort, where the association is directionally consistent, the meta-analysis p value <0.05, and Cochrane's Q p-value >0.05 for between discovery and replication MR estimates. NA denotes where full summary statistics were not available for colocalisation analyses.

Exploring fetal circulating proteins influencing offspring birth weight

We used two-sample MR to explore the effect of genetically instrumented fetal circulating protein levels on offspring own birth weight. We found evidence of causal effects on offspring birth weight for 5 proteins (instrumented by 5 cis-pQTL). We found evidence of higher offspring levels of LGALS4 and FCGR2B causing higher birth weight (mean increase in birth weight per SD of protein levels 46g (95% CI: 23, 70g) and 12g (3, 21g), respectively) and higher levels of PCSK1, UBASH3B and LEPR causing lower birth weight (-9g (95% CI:-16, 4g), -47g (-73, -22g) and -8g (-12, -4g)), respectively **(Figure 4).** These MR results passing our FDR 5% corrected p-values were selected for follow-up analyses, which were all instrumented by cis-pQTL only.

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

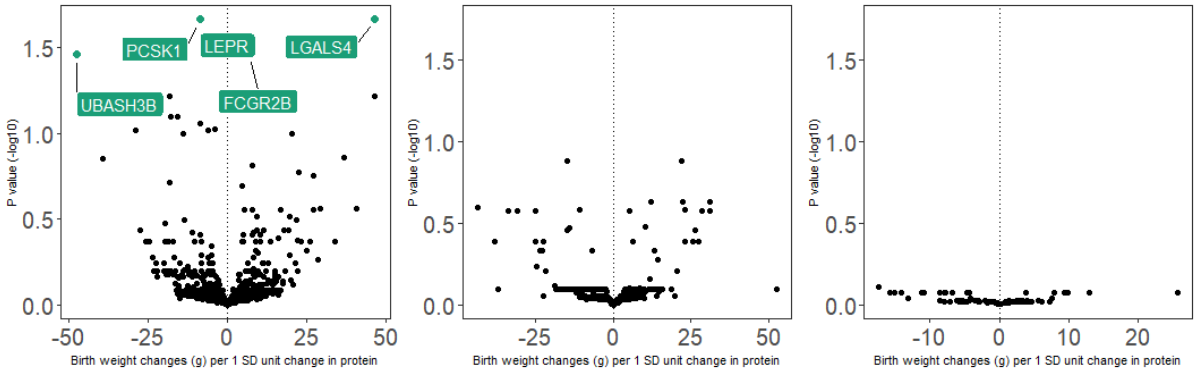
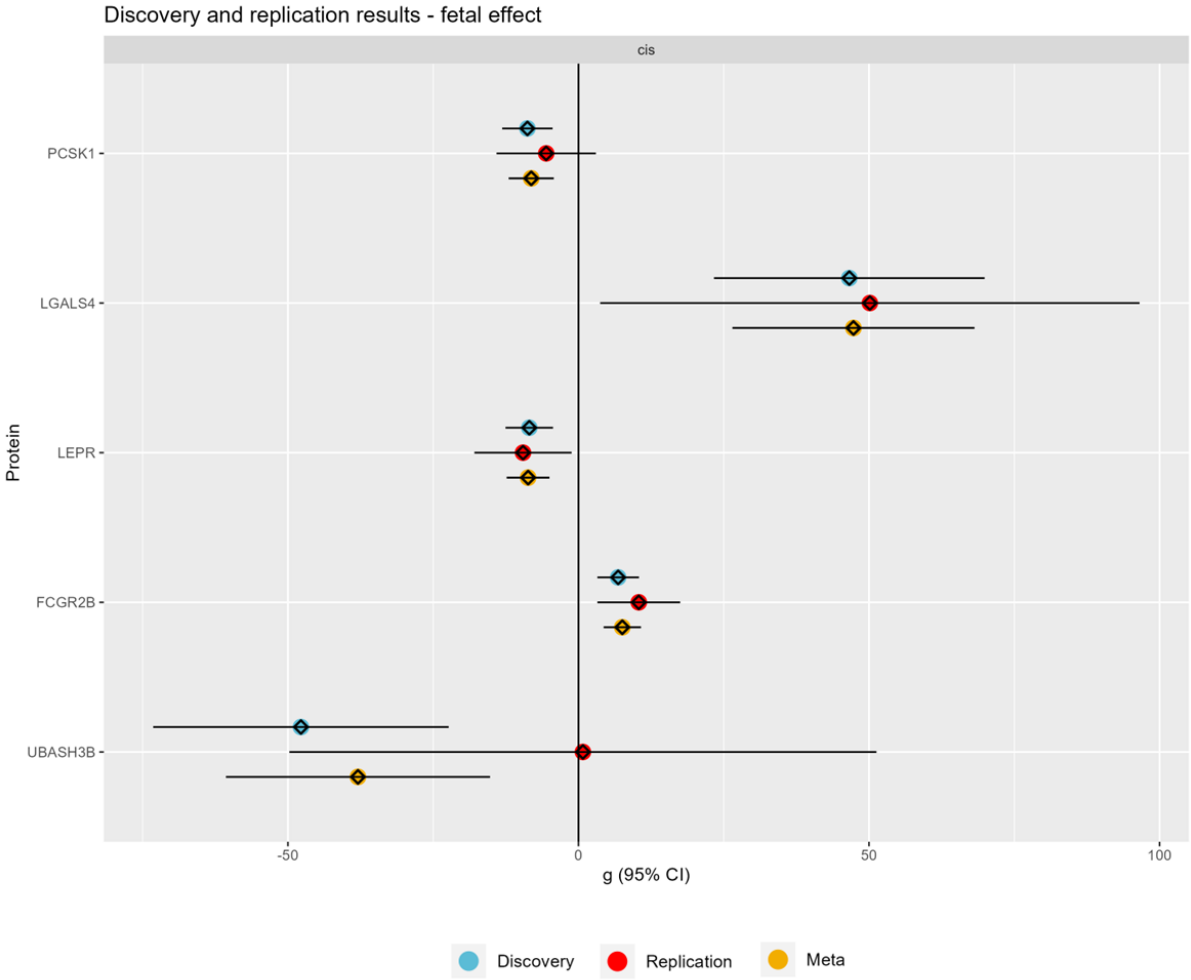


Figure 4 – Fetal two sample MR results for cis (left), trans (middle) and cis-and-trans (left) pQTL. FDR corrected p-values transformed to their -log10. Results plotted in green passed FDR corrected p-value threshold. Abbreviations: SD, standard deviation

We tested for replication of MR results for the 5 proteins identified in our fetal discovery analyses (FDR p-value <0.05) in an independent sample of 62,108 offspring from MoBa. We found that MR results were consistent with discovery in our meta-analysis, for 4 of the 5 proteins (all but UBASH3B) (**Figure 5** and **Supplementary Table 8**).



Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Figure 5 Mendelian randomization analyses estimating the effect of genetically-instrumented fetal proteins (in standard deviation units) on offspring birth weight (in grams) in the discovery (N=297,356 offspring), replication (N=62,108 offspring) and meta-analyses of discovery and replication samples. Replication was tested for proteins passing FDR-corrected p-value<0.05 in the meta-analysis.

For fetal effects, colocalisation was performed for 4 of the 5 proteins which survived replication. There was strong to moderate evidence of colocalisation for FCGR2B (PP H_4: 0.95), PCSK1 (PP H_4: 0.77), and LEPR (PP H_4: 0.68). There was evidence that the MR results for fetal LGALS4 might be explained by LD confounding (PP H_3: 0.58) (**Table 2**) (**Supplementary Figures 12-15**).

Table 2 – Genetic colocalisation results for offspring proteins with birth weight.

	Protein	Cis or trans	PP H0	PP H1	PP H2	PP H3	PP H4
Fetal effect	PCSK1	cis	0	0.15	0	0.07	0.77
	LGALS4	cis	0	0.07	0	0.58	0.34
	LEPR	cis	0	0.04	0	0.31	0.68
	FCGR2B	cis	0	0.05	0	0	0.95

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

Conclusions based on results across all analyses investigating the causal effect of genetically-instrumented fetal proteins on offspring birth weight are summarised in **Table 4**.

Table 4 – Summary of results from analyses exploring the effect of fetal proteins on offspring birth weight

Protein	Uniprot ID	pQTL	Discovery MR direction of effect	Replicated	Colocalisation
PCSK1	P29120	cis	-	Yes	Yes
LGALS4	P56470	cis	+	Yes	No
FCGR2B	P31944	cis	+	Yes	Yes
LEPR	P48357	cis	-	Yes	Yes
UBASH3B	Q8TF42	cis	-	No	NA

Abbreviations – pQTL, protein quantitative trait loci, Mr, Mendelian randomisation, NA, full GWAS summary statistics were not available

In the absence of large-scale datasets with available information on fetal genotype and cord-blood proteome, we could not test whether the selected pQTL are associated with fetal

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

circulating proteins. Instead, we explored whether the selected proteins are expressed in fetal tissue using single cell gene expression data from the human cell atlas of fetal gene expression (26), profiling 4 million single cells including 121 human fetal samples (post-conceptional age =72-129 days) from 15 organs (eye, heart, intestine, kidney, liver, lung, muscle, pancreas, placenta, spleen, stomach, thymus, adrenal, cerebellum and cerebrum) (18). In **Supplementary Figures 15-18**, these maps show where these genes may be expressed in fetal tissue: PCSK1 (pancreas), LGALS4, UBASH3B and LEPR (across multiple tissues), and FCGR2B (heart).

Protein networks of identified maternal and fetal proteins affecting birth weight

To gain insights into biological pathways underlying our findings, we used STRING (27) to explore the presence of protein-protein relationships (typically referred to as protein-protein interactions but statistically reflecting association, rather than statistical evidence of an interaction between two genetic variants on an outcome) for proteins that were identified in discovery analyses, replicated and had supportive or uncertain evidence of colocalisation in either maternal or fetal analyses (PCSK1, CD59, RAD51D, GP1BA, CREB3L4, RPN1, VCAM1, FCGR2B and LEPR – **Figure 6**). The other 4 (LGALS4, GDF15, CLMP, SEMG2) are not shown because they had no protein networks which survived our criteria (high confidence, 0.900, and restricting evidence to the following sources: experimentally determined, databases, gene neighbourhood, fusion or co-expression). As **Figure 6A** shows, PCSK1 associates with proteins regulating glucose homeostasis, such as insulin (INS), of which it is a stimulator, and glucagon (GCG). It also associates with proteins regulating food intake, such as pro-opiomelanocortin (POMC) and ghrelin (GHRL) (28-31). CD59 was associated with SCFD1, previously associated with height (32) (**6B**). GP1BA is in a network with SRC, both also known to play a role in height as well as embryonic development and cell growth (32) (**6C**). CREB3L4 plays a role in protein maturation and metabolism, and interacts with MBTPS2, serine proteases involved in cell proliferation (**6D**). LEPR interacts with JAK2 and LEP, associated with type 2 diabetes (33), BMI (34), known to regulate body weight homeostasis, and produced by the placenta during pregnancy (**6E**) (35). VCAM1 associates with proteins in the ITG family (of which CD59 is a member) (**6F**) and specifically ITGA9, previously implicated in height, waist circumference and early onset pre-eclampsia (32, 36) (37). Unsurprisingly, the RAD51D network includes its paralogs RAD51B, RAD51C, XRCC2M and XRCC3, as well as SFR1, all of which are involved in DNA and double strand sequence repair (38, 39) (**6G**). FCGR2B had a wide network of interactions including LILRB2, part of the LILR family of receptors, which are known to play a role in immune response and tolerance and have been implicated in developmental processes of pregnancy (**6H)** (40) as well as

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

height (32). RPN1 and RPN2 are glycoproteins, and associated with other proteins involved in glycosylation, including DAD1, STT3A, STT3B, MAGT1 and DDOST (**6I**).

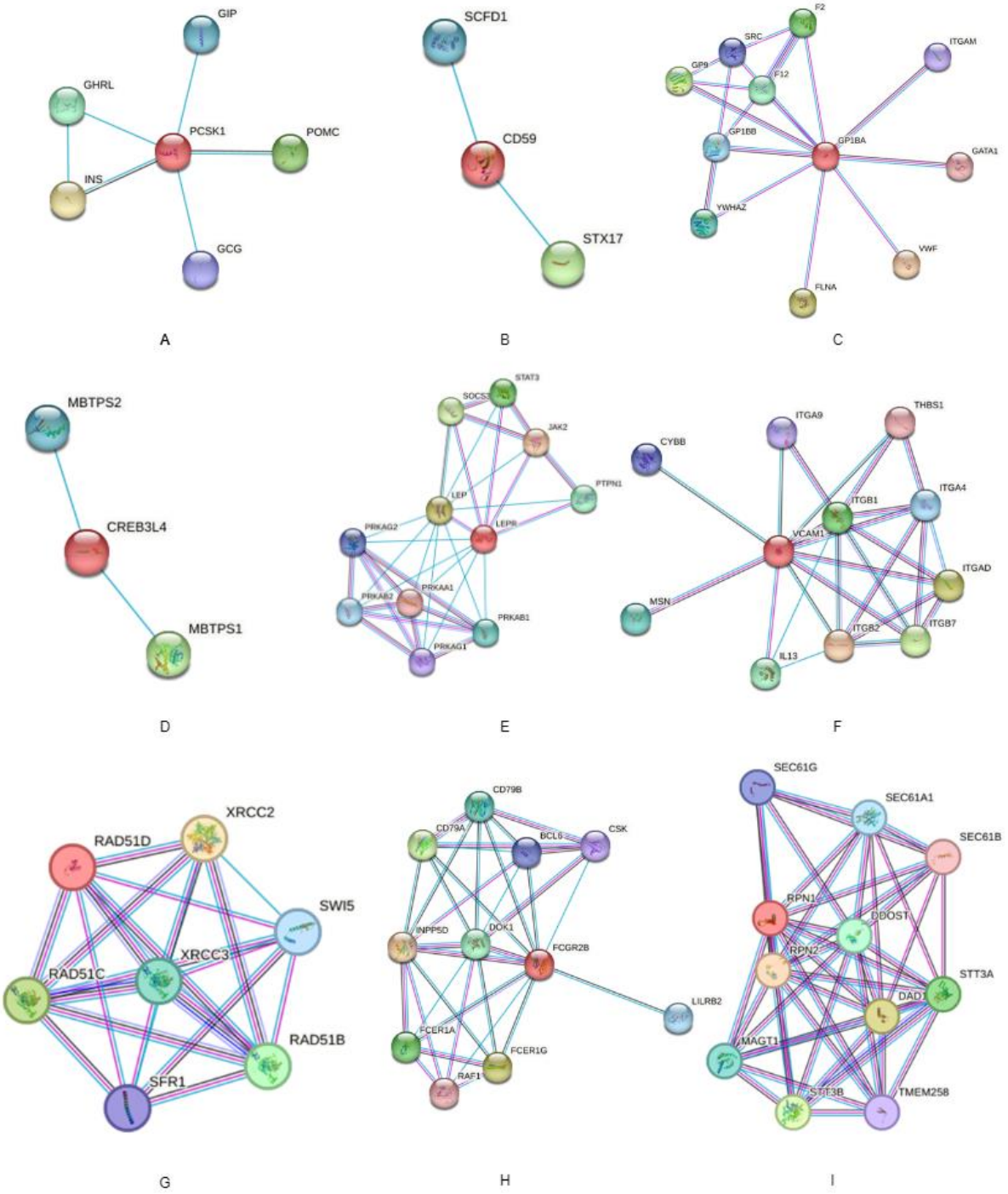


Figure 6 - Protein-protein **associations** around PCSK1 (A), CD59 (B), GP1BA (C), CREB3L4 (D), LEPR (E), VCAM1 (F) RAD51D (G) FCGR2B (H) and RPN1 (I). Different line colours represent different sources of evidence for these predictions – known associations from curated databases (light blue) and experimentally determined (pink); predicted associations from gene neighbourhood (green), gene fusions (red), gene co-occurrence (dark blue), and protein homology (lilac). Associations are for proteins jointly contributing to a shared function, rather than specifically binding to each other. Splice isoforms or post-translational

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

modifications are collapsed, i.e. each node represents all the proteins produced by a single, protein-coding gene locus. The physical distances between two nodes along an edge in a graph have no meaning.

Discussion

We have used MR to explore the causal effect of 996 maternal pregnancy and fetal plasma proteins (753 instrumented by cis-pQTL, 244 instrumented by trans-pQTL and 81 instrumented by cis-and-trans pQTL) on birth weight. We found evidence of potential causal effects, unlikely to be explained by chance or confounding by LD, for 6 maternal circulating proteins (PCSK1, LGALS4, RAD51D, VCAM1, GP1BA, CREB3L4) and 3 fetal circulating proteins (PCSK1, LEPR, FCGR2B). For the remainder of the discussion, we will focus on these proteins which had the strongest evidence across all analyses.

Proteins with opposing maternal and fetal effects on birthweight

Two proteins, PCSK1 and LGALS4, instrumented by cis-pQTLs, appear to be acting in opposite directions when comparing maternal and fetal MR results. They suggest that more PCSK1 from the mother leads to higher birth weight, yet more PCSK1 from the fetus leads to lower birth weight. The STRING protein-protein interactions in this study show a link between PCSK1 and (INS) insulin – a key determinant of fetal growth – and glucagon (GCG) (**Figure 6A**), which suggests a role in glucose homeostasis. However, these protein-protein interactions show co-expression, which may be downstream of a certain mechanism, but they do not always imply interaction. The different sources of evidence are important when interpreting these interaction networks. We did not include text-mining and gene neighborhood interactions in the STRING networks so we could follow the most robust lines of evidence. We tried to be conservative and applied a maximum threshold of 0.900 (highest confidence). This may explain why we did not identify any associations for some proteins that replicated and were supported by colozalization (LGALS4, CLMP and GDF15) or because proteins like LGALS4 are poorly characterized (27).

LGALS4 had the opposite direction of effect to PCSK1. More LGALS4 from the mother leads to lower birthweight and more LGALS4 from the fetus leads to higher birthweight. LGALS4 is part of the galectin family which binds carbohydrates, lactose and sugars. Previous work has shown that the galectin family of β-galactoside binding proteins are important in modulating diverse developmental processes. They contribute to placentation and regulate maternal immune tolerance (41) and have been previously suggested as biomarkers for gestational disorders like gestational diabetes (42) and infertility (41). Other expression studies have shown the importance of galectins in pregnancy, and the importance of a maternal-fetal

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

balancing act for ideal pregnancy outcomes. For example, lower levels of fetal expression of Galectins 3, 8 and 9 have previously been associated with intrauterine growth restriction (41). However, most of the previous studies in LGALS4 in pregnancy have come from mouse or in vitro models and our study provides an additional line of evidence of a potential role of LGALS4 in human fetal growth (41).

Proteins with maternal-specific effects on birthweight

There was also strong evidence for a role of maternal trans-pQTLs instrumented VCAM1, RAD51D, CREB3L4, and GP1BA. Maternal levels of VCAM1 have previously been implicated in endothelial dysfunction, which is linked to disorders of vascular pathogenesis such as preeclampsia and fetal growth restriction (37). RAD51D and its family of paralogs are essential for DNA and double strand repair (38). CREB3L4 is expressed at low levels across all tissues (43). Previous studies have shown CREB3L4 plays a role in adipocyte differentiation, and this is supported by mice studies have shown that CRE3L4-knockout mice had weight gain and higher adiposity than wild type mice (44). GP1BA, as on the same biologic pathway as VCAM1 (11), has previously been associated with preeclampsia (45), and is known to have vascular functions and has been posited to play a role in placental function (46).

There were several other proteins identified in the maternal analyses that provide insights on putative maternal molecular mechanisms regulating fetal growth but for which we had limited power to explore bias in MR results due to confounding by LD. One example is the MR finding that higher maternal GDF15 causes higher birth weight. Recently there has been further research finding previous work showing that GDF15 is implicated in hyperemesis gravidarum (47). HG is associated with low birth weight, potentially because of a mother's inability to retain nutrients during pregnancy (48). Our findings are complementary because genetically higher maternal pre-pregnancy levels of GDF15 are likely to make a mother less sensitive to GDF15 produced by the fetus in the first trimester of pregnancy, which leads to lower levels of nausea and vomiting (47). This may partially explain the relationship we see here with high GDF15 and higher birth weight.

Proteins with fetal-specific effects on birthweight

In our fetal analysis, we found evidence that higher fetal cis-pQTL instrumented LEPR reduced birth weight. LEPR has a well-established role in energy metabolism, body weight, and is secreted in fetal tissue and the placenta (49). It has been reported that individuals with missense mutations that impair receptor signalling have early onset obesity, therefore, we speculate that higher circulating LEPR might be related to lower birthweight through higher

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

leptin cell signalling. We also found evidence for fetal cis-pQTL instrumented FCGR2B increasing birth weight. Whilst there has not been a lot of mechanistic work in this area, a recent GWAS of placental weight identified an eQTL and pQTL for FCGR2B, suggesting a potential role in antibody transport across the placenta (50).

Strengths and limitations

We have integrated large scale proteomics and birthweight genetic association data to shed light on maternal and fetal molecular mechanisms regulating fetal growth. We have used proteomics data from 5 different sources in the INTERVAL studies, which has been previously validated in a useable pipeline (11). Furthermore, we have used large scale birth weight data from EGG, UK Biobank and MoBa, three large European cohorts, for our discovery and replication analyses and meta-analyses. We have undertaken sensitivity analyses on our initial findings including checking for confounding by LD and attempted to further explain the underlying biology of our molecular findings.

In terms of limitations, our analysis has a limited coverage of the proteome due to the use of GWAS data on plasma proteins and restriction to pQTLs with credible evidence of association with the included proteins (further details in **Methods**). This implies that additional putative causal proteins might be identified in future studies using data from more comprehensive plasma proteomics panels or using cell- or tissue-specific proteomics. In addition, this also means we were unable to replicate previous findings on well-established proteins which influence fetal growth such as sFlt-1 and PlGF, the ratio of which is often used to predict preeclampsia and fetal growth restriction (51-53). There were also proteins not present in our data from candidate proteins studies, or from a recent observational proteomics study, such as PRSS8, ACE2, SORT1, LOX-1, TM and PAPP-A (N=857 mother-offspring pairs (11)). In our data, and previously associated with fetal growth in the aforementioned observational study, were MMP-12 and ADM, instrumented by a cis and a trans-pQTL, respectively, but neither were top findings in our MR analysis.

Another limitation is related to potential of epitope effects intrinsic to the proteomic technology. Whilst we selected genetic instruments that had been through several quality control tests, including to reduce the likelihood of misclassification of protein levels through epitope effects (i.e. differences in antigen recognition, rather than protein levels) (23) we cannot rule out some bias from these effects. In addition, for proteins instrumented by trans-pQTL, results from MR analyses need to be interpreted with caution due to the higher possibility of bias due to pleiotropic mechanisms (54). Finally, we have assumed that genetic variants identified in protein GWAS, conducted in males and non-pregnant women, are similarly associated with

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

circulating proteins in fetuses and pregnant women. Ideally, we would test this in samples collected from our target population; however, such data is not available or available in insufficient numbers. For the fetal analyses, we have shown that all proteins identified in the MR analyses are expressed in one or more fetal tissues.

Conclusions

We found strong evidence for causal effects of several proteins on offspring birth weight across a range of analyses, highlighting proteins with opposing maternal and offspring effects (PCSK1, LGALS4), as well as with specific maternal (VCAM1, RAD51D, GP1BA, CREB3L4) or fetal (LEPR, FCGR2B) effects. These proteins are involved in multiple biological processes governing glucose homeostasis, energy metabolism, endothelial function and adipocyte differentiation. In addition, we found some evidence for other maternal proteins having potential effect on birth weight (CD59, GDF15, CLMP, SEMG2, RPN1), which require further exploration with larger sample sizes and/or more pQTL. These findings provide new insights in understanding the maternal and fetal molecular mechanisms regulating fetal growth.

Methods

Ethics

All human research was approved by the relevant institutional review boards and conducted according to the Declaration of Helsinki. Participants of all studies in the EGG consortium provided written informed consent. The UK Biobank has approval from the North West Multi-Centre Research Ethics Committee, which covers the United Kingdom. The establishment of MoBa and initial data collection was based on a license from the Norwegian Data Protection Agency and approval from The Regional Committees for Medical and Health Research Ethics. The MoBa cohort is currently regulated by the Norwegian Health Registry Act. The current study was approved by The Regional Committees for Medical and Health Research Ethics (223273).

Selection of genetic instruments for plasma proteins

We selected pQTL as instruments for proteins using the dataset curated by Zheng *et al*. (**Supplementary Figure 1** (11)), which combined genetic association data for plasma proteins from five GWAS: Sun *et al.* 2018 (N=3,301 individuals), Folkersen *et al.* 2017 (N=3,394), Suhre *et al.* 2017 (N=1,000), Yao *et al.* 2018 (N=6,861) and Emilsson *et al.* 2018 (N=5,457). Zheng *et al.* generated a pipeline to collate the proteomic data from these five GWAS. (Details on the proteomics platforms, sample size, and number of analysed proteins for each GWAS can be found in **Supplementary Table 3**). To explore the reliability of pQTL as genetic

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

instruments for proteins, Zheng *et al*. classified them in three tiers based on specificity and consistency quality control tests, as described in detail in their publication (11). Specificity was defined as pQTL associated with fewer than five proteins, and consistency was defined as pQTL that show homogeneous genetic signals across the five GWAS based on effect size. Tier 1 pQTL passed both specificity and consistency tests, tier 2 passed the specificity test only, and tier 3 pQTL failed both tests. For our analyses, we excluded tier 3 pQTL (N=899 for 182 proteins), as they are likely to be unreliable instruments for the corresponding proteins **(Supplementary Table 1)**. We included: i) *cis*-acting pQTL, located ±500 kb of the transcription start site of the protein encoding gene (N=959 pQTL for 753 proteins), ii) *trans*-acting pQTL, located outside this window (N=315 for 281 proteins), and iii) '*cis*-and-*trans*' pQTL, i.e., with both *cis*- and *trans*-instruments (N=209 pQTL for 81 proteins) **(Supplementary Figure 1)**.

Discovery sample

We used publicly available meta-analysed GWAS data on birth weight. This data was obtained from the 2019 Early Growth Genetics (EGG) consortium plus UK Biobank (UKBB) GWAS. This GWAS was used because it has publicly available estimates of maternal effects adjusted for fetal effects and vice versa. These data included fetal-adjusted maternal and maternal-adjusted fetal genetic associations. Overall, 406,063 participants contributed to the weighted linear model (WLM) analyses, of which 101,541 were UKB participants who reported their own birth weight and birth weight of their first child, 195,815 were UKB and EGG participants with their own birth weight data, and 108,707 were UKB and EGG participants with offspring birth weight data (N=216,611 and N=190,387, respectively) (8).

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

Replication sample

For replication, we used genetic association data for birth weight from MoBa, a population-based pregnancy cohort study conducted by the Norwegian Institute of Public Health. Pregnant women and their partners were recruited from all over Norway from 1999-2008. The women consented to participate in 41% of the pregnancies. The cohort includes approximately

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

114,500 children, 95,200 mothers, and 75,200 fathers (24). The current study is based on version 12 of the quality-assured data files released for research in January 2019. Phenotypic information from the Medical Birth Registry of Norway (MBRN, a national health registry established in 1967 that contains information about all births in Norway and is linked to the MoBa study using unique personal identification numbers) and MoBa questionnaire data were used to identify biological sex, year of birth, reported parent-offspring (PO) relationships, birth weight, and multiple births (only available in the offspring). As in the discovery sample, multiple births and extremes of birth weight were excluded. Analyses were adjusted for genotyping batch.

In MoBa, samples were obtained from both parents during pregnancy (plasma blood samples) and from mothers and children at birth (umbilical cord) (55). This project used MoBa genetic data that was quality controlled and imputed using the MoBaPsychGen pipeline, which has previously been described (56). Phasing and imputation were performed using the publicly available Haplotype Reference Consortium release 1.1 panel as a reference (57). The pipeline output consists of best-guess hard-call genotype data for 6,981,748 variants in 207,409 MoBa participants all of European ancestry. GWAS in MoBa were run in R 4.1.0 using *regenie* software (v 3.1.2) on the TSD server (58).

Adjustment for maternal/offspring genotype

Birth weight is influenced by both maternal and fetal genotype. Therefore, to distinguish maternal and fetal genetic effects, in the discovery sample a weighted linear model was used to adjust maternal data for fetal genetic associations, and fetal data for maternal genetic associations (3). From this, genetically-instrumented maternal effects of proteins on birth weight (not mediated or biased due to horizontal pleiotropy via fetal effects) and genetically instrumented fetal effects of proteins on birth weight (not confounded by maternal genotype) can be estimated. We undertook the same adjustment in the MoBa (replication) summary statistics. We used the maternal and fetal summary statistics from the birth weight GWAS, and used a weighted linear model implemented via the DONUTS software package (51) to estimate mutually adjusted maternal and fetal genetic effects at each SNP. We used genetic covariance intercepts from LD score regression (LDSC) to account for overlap between maternal and fetal GWAS sample (59). The output was then used in our two sample MR replication analyses, with maternal and fetal genetic effects having been adjusted for.

Two sample MR

We conducted MR analyses to study the causal effects of circulating maternal and fetal proteins on birth weight. All MR analyses were performed using the *TwoSampleMR* R package

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

version 0.5.6 (26). Prior to two-sample MR analysis, genetic association data for proteins and birth weight were harmonized so that the effect of each genetic variant on the exposure and outcome were relative to the same effect allele. In some instances, data could not be harmonised and these pQTL, and corresponding proteins, were removed from further analyses (**Supplementary Table 2)**. After excluding pQTL that failed harmonisation (due to being palindromic (A/T, G/C)/ SNPs with intermediate allele frequencies), we retained data for 753 proteins instrumented by 959 *cis*-pQTL, 281 proteins instrumented by 315 *trans*-pQTL, and 81 proteins instrumented by 209 cis-and-trans pQTL (**Figure 1** and **Supplementary Figure 1**). We analysed maternal and fetal data separately. We used the Wald ratio estimator for analyses where only one pQTL was available, and the inverse variance weighted (IVW) model when multiple pQTL were available for the protein. In analyses including multiple *cis*-pQTL in partial LD, we used a generalised IVW model accounting for LD, where the pairwise LD r^2 values were obtained from the 1000 Genomes European super-population. Two-sample MR results which had an FDR-adjusted p-value ≤0.05 were carried forward for replication and subsequent analyses.

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

Genetic colocalisation

We undertook genetic colocalisation analyses using the *coloc* R package (25) to investigate whether our MR findings were compatible with a shared causal variant or confounded by LD (25). *Coloc* computes all possible configurations of causal variants for each of two traits and uses a Bayesian approach to calculate support for each causal model (H_0: association with protein only; H_1: association with birth weight only; H_2: association with both traits owing to distinct causal variants; H_3: association with both traits owing to a single shared causal variant; H_4: no association). We restricted the analyses to a ±500kb window around the pQTL, and

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

assumed a prior probability that any pQTL in the region is associated with the protein only (p_1) of 1×10^{-4}, birth weight only (p_2) of 1×10^{-4}, or both traits (p_{12}) 1×10^{-5}. We considered a posterior probability of association (PPA) for $H_4 > 0.7$ to provide strong evidence for colocalisation, while a PPA for $H_3 > 0.5$ was considered as evidence against colocalisation. To facilitate visualisation of genomic signals in each region, we generated stacked recombination plots using the *locuscompareR* R package (http://locuscompare.com/) (60).

Where available, we extracted genetic association data for the genomic regions selected for colocalisation analyses for plasma proteins and birth weight using the IEU Open GWAS (https://gwas.mrcieu.ac.uk/). The summary statistics for the genetic associations with PCSK1, VCAM1, RAD51D, GP1BA, CREB3L4 and RPN1 were manually downloaded from the Sun *et al* results portal (https://app.box.com/s/u3flbp13zjydegrxjb2uepagp1vb6bj2). We were unable to perform *coloc* for proteins profiled by Emillson *et al*, Yao *et al* or Suhre *et al* (PRSS57, B2M, or NPPB in the maternal analysis, and UBASH3B in the fetal analysis) because the summary statistics were not available.

Fetal expression lookup

We could not test whether pQTL were associated with fetal circulating proteins due to the absence of large-scale datasets with this information. Instead, we explored whether the selected proteins were expressed in fetal tissue using single-cell gene expression data from the human cell atlas of fetal gene expression (26). The atlas was developed from 121 human fetal samples (post-conceptional age = 72-129 days) from 15 organs (eye, heart, intestine, kidney, liver, lung, muscle, pancreas, placenta, spleen, stomach, thymus, adrenal, cerebellum and cerebrum), profiling 4 million single cells using a sci-RNA-seq3 protocol. The authors used a clustering approach building on a form of principal components analysis (uniform manifold approximation and projection - UMAP) to generate a resource where fetal gene expression can be mapped across different organs, tissues and cell types (https://descartes.brotmanbaty.org/). We generated expression maps for the 4/5 proteins which replicated in the fetal meta-analysis: PCSK1, LGALS4, LEPR and FCGR2B (**Supplementary Figures 16-19**).

Protein networks

We explored information on protein-protein associations (referred to as interactions in genetic research) from the STRING database (https://string-db.org/cgi) to gain insights into putative biological processes underlying our findings. STRING collates information from all known protein-protein interactions databases and provides a visual network of predicted associations for a particular group of proteins, whereby the proteins are represented as 'nodes' and the

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

lines which join them are the predicted functional associations. The colour of the line denotes the source of the interaction information (see figure legend) (27). We set the confidence score of a given association to be 'highest confidence' (0.900) and restricted to associations from curated databases (either KEGG, Reactome, BioCyc and Gene Ontology, as well as legacy datasets from PID and BioCarta), experimentally determined (where evidence comes from lab experiments), reported from gene fusions (from reported gene fusion events), gene co-occurrence (where gene families have similar occurrence patterns across genomes), co-expression (where gene are shown to be correlated in expression across a large number of datasets) and protein homology.

Data availability

Data on birth weight in our discovery analysis was contributed by the EGG Consortium using the UK Biobank Resource and was downloaded from www.egg-consortium.org. The genotype and phenotype data are available on application from the UK Biobank (http://www.ukbiobank.ac.uk/). Individual cohorts participating in the EGG Consortium should be contacted directly as each cohort has different data access policies. The exposure summary data from which genetic instruments were selected for pQTL are available from EpiGraphDB pQTL browser (https://epigraphdb.org/pqtl). Data from MoBa are available from the Norwegian Institute of Public Health after application to the MoBa Scientific Management Group (see its website https://www.fhi.no/en/op/data-access-from-health-registries-health-studies-and-biobanks/data-access/applying-for-access-to-data/ for details).

Acknowledgments

We gratefully acknowledge the participants in the studies that were directly used or indirectly used via consortia summary data, and researchers who made their data open access. The Norwegian Mother, Father and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. We are grateful to all the participating families in Norway who take part in this on-going cohort study. We thank the Norwegian Institute of Public Health (NIPH) for generating high-quality genomic data. This research is part of the HARVEST collaboration, supported by the Research Council of Norway (#229624). We also thank the NORMENT Centre for providing genotype data, funded by the Research Council of Norway (#223273), South East Norway Health Authorities and Stiftelsen Kristian Gerhard Jebsen. We further thank the Center for Diabetes Research, the University of Bergen for providing genotype data and performing quality control and imputation of the data funded by the ERC AdG project SELECTionPREDISPOSED, Stiftelsen Kristian Gerhard

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Jebsen, Trond Mohn Foundation, the Research Council of Norway, the Novo Nordisk Foundation, the University of Bergen, and the Western Norway Health Authorities. This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol - http://www.bristol.ac.uk/acrc/ and the TSD (Tjeneste for Sensitive Data) facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT Department (USIT) (tsd-drift@usit.uio.no). The computations on this platform were performed on resources provided by Sigma2 - the National Infrastructure for High-Performance Computing and Data Storage in Norway.

Funding

This work was supported by the University of Bristol and UK Medical Research Council (MC_UU_00032/05, MC_UU_00032/03, MC_UU_00032/01), the US National Institute for Health (R01 DK10324), the European Research Council via Advanced Grant 101021566, the British Heart Foundation (AA/18/7/34219 and CS/16/4/32482) and the National Institute of Health Research Bristol Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol.

RMF and RNB were funded by a Wellcome Trust and Royal Society Sir Henry Dale Fellowship (WT104150). RMF is supported by a Wellcome Senior Research Fellowship (WT220390). This research was supported by the National Institute for Health and Care Research Exeter Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. This research was funded in part by the Wellcome Trust [WT220390]. For open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. J.Z. is supported by grants from the National Key Research and Development Program of China (2022YFC2505203). MCB is funded by a University of Bristol Vice Chancellor's Fellowship. EC is supported by funding from the Research Council of Norway (#274611) and the South-Eastern Norway Regional Health Authority (#2021045).

Author contributions

Study conception: DAL, MCB, RMF; Analysis: NM, MCB; Manuscript drafting: NM; Manuscript revisions: MCB, DAL, AFS, MA, RB, RMF, TB, TRG, JZ, DE, DH, GC, EC, MM

Conflict of interest

The authors report no conflict of interest.

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Code availability

All code is available on the GitHub repo https://github.com/MRCIEU/MR-PREG/[to-be-added]

References

1. Küpers LK, Monnereau C, Sharp GC, Yousefi P, Salas LA, Ghantous A, et al. Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. Nature Communications. 2019;10(1):1893.
2. Leite DFB, Cecatti JG. Fetal Growth Restriction Prediction: How to Move beyond. The Scientific World Journal. 2019;2019:1519048.
3. Warrington NM, Freathy RM, Neale MC, Evans DM. Using structural equation modelling to jointly estimate maternal and fetal effects on birthweight in the UK Biobank. International Journal of Epidemiology. 2018;47(4):1229-41.
4. Wilcox AJ. On the importance—and the unimportance— of birthweight. International Journal of Epidemiology. 2001;30(6):1233-41.
5. Barry CA-O, Lawlor DA, Shapland CA-O, Sanderson E, Borges MA-O. Using Mendelian Randomisation to Prioritise Candidate Maternal Metabolic Traits Influencing Offspring Birthweight. LID - 10.3390/metabo12060537 [doi] LID - 537. (2218-1989 (Print)).
6. Bond TA, Karhunen V, Wielscher M, Auvinen J, Männikkö M, Keinänen-Kiukaanniemi S, et al. Exploring the role of genetic confounding in the association between maternal and offspring body mass index: evidence from three birth cohorts. International Journal of Epidemiology. 2019;49(1):233-43.
7. Iliodromiti S, Mackay DF, Smith GCS, Pell JP, Sattar N, Lawlor DA, et al. Customised and Noncustomised Birth Weight Centiles and Prediction of Stillbirth and Infant Mortality and Morbidity: A Cohort Study of 979,912 Term Singleton Pregnancies in Scotland. PLOS Medicine. 2017;14(1):e1002228.
8. Warrington NM, Beaumont RN, Horikoshi M, Day FR, Helgeland Ø, Laurin C, et al. Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. Nature genetics. 2019;51(5):804-14.
9. Beaumont RN, Warrington NM, Cavadino A, Tyrrell J, Nodzenski M, Horikoshi M, et al. Genome-wide association study of offspring birth weight in 86 577 women identifies five novel loci and highlights maternal genetic effects that are independent of fetal genetics. (1460-2083 (Electronic)).
10. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrmisdottir EL, et al. Large-scale integration of the plasma proteome with genetics and disease. Nature Genetics. 2021;53(12):1712-21.
11. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. Nature Genetics. 2020;52(10):1122-31.
12. Zhang G, Bacelis J, Lengyel C, Teramo K, Hallman M, Helgeland Ø, et al. Assessing the Causal Relationship of Maternal Height on Birth Size and Gestational Age at Birth: A Mendelian Randomization Analysis. (1549-1676 (Electronic)).
13. Tyrrell J, Richmond RC, Palmer TM, Feenstra B, Rangarajan J, Metrustry S, et al. Genetic Evidence for Causal Relationships Between Maternal Obesity-Related Traits and Birth Weight. (1538-3598 (Electronic)).
14. Ardissino MA-O, Slob EA-O, Millar O, Reddy RA-OX, Lazzari L, Patel KHK, et al. Maternal Hypertension Increases Risk of Preeclampsia and Low Fetal Birthweight: Genetic Evidence From a Mendelian Randomization Study. (1524-4563 (Electronic)).

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

15. Brand JA-O, Gaillard RA-O, West JA-O, McEachan RA-O, Wright JA-O, Voerman E, et al. Associations of maternal quitting, reducing, and continuing smoking during pregnancy with longitudinal fetal growth: Findings from Mendelian randomization and parental negative control studies. (1549-1676 (Electronic)).
16. Thompson WA-O, Beaumont RA-O, Kuang AA-O, Warrington NA-OX, Ji Y, Tyrrell JA-O, et al. Higher maternal adiposity reduces offspring birthweight if associated with a metabolically favourable profile. (1432-0428 (Electronic)).
17. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558(7708):73-9.
18. Emilsson V, Ilkov M, Lamb JR, Finkel N, Gudmundsson EF, Pitts R, et al. Co-regulatory networks of human serum proteins link genetics to disease. Science. 2018;361(6404):769-73.
19. Folkersen L, Fauman E, Sabater-Lleal M, Strawbridge RJ, Frånberg M, Sennblad B, et al. Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. PLoS genetics. 2017;13(4):e1006706.
20. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. Nature communications. 2017;8(1):1-14.
21. Yao C, Chen G, Song C, Keefe J, Mendelson M, Huan T, et al. Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. Nature communications. 2018;9(1):1-11.
22. Zhao H, Rasheed H, Nøst TH, Cho Y, Liu Y, Bhatta L, et al. Proteome-wide Mendelian randomization in global biobank meta-analysis reveals multi-ancestry drug targets for common diseases. medRxiv. 2022:2022.01.09.21268473.
23. Suhre K, McCarthy MI, Schwenk JM. Genetics meets proteomics: perspectives for large population-based studies. Nature Reviews Genetics. 2021;22(1):19-37.
24. Haugan A, Birke C, Alsaker E, Høiseth G, Knudsen GP, Tambs K, et al. Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). International Journal of Epidemiology. 2016;45(2):382-8.
25. Wallace C. Statistical Testing of Shared Genetic Control for Potentially Related Traits. Genetic Epidemiology. 2013;37(8):802-13.
26. Cao J, O'Day DR, Pliner HA, Kingsley PD, Deng M, Daza RM, et al. A human cell atlas of fetal gene expression. Science. 2020;370(6518):eaba7721.
27. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Research. 2014;43(D1):D447-D52.
28. Stijnen P, Ramos-Molina B, O'Rahilly S, Creemers JWM. PCSK1 Mutations and Human Endocrinopathies: From Obesity to Gastrointestinal Disorders. Endocrine Reviews. 2016;37(4):347-71.
29. Wei X, Ma X, Lu R, Bai G, Zhang J, Deng R, et al. Genetic Variants in PCSK1 Gene Are Associated with the Risk of Coronary Artery Disease in Type 2 Diabetes in a Chinese Han Population: A Case Control Study. PLOS ONE. 2014;9(1):e87168.
30. Benzinou M, Creemers Jw Fau - Choquet H, Choquet H Fau - Lobbens S, Lobbens S Fau - Dina C, Dina C Fau - Durand E, Durand E Fau - Guerardel A, et al. Common nonsynonymous variants in PCSK1 confer risk of obesity. (1546-1718 (Electronic)).
31. Gjesing AP, Vestmar MA, Jørgensen T, Heni M, Holst JJ, Witte DR, et al. The Effect of PCSK1 Variants on Waist, Waist-Hip Ratio and Glucose Metabolism Is Modified by Sex and Glucose Tolerance Status. PLOS ONE. 2011;6(9):e23907.
32. Yengo LA-O, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, et al. A saturated map of common genetic variants associated with human height. (1476-4687 (Electronic)).
33. Suzuki KA-O, Akiyama M, Ishigaki K, Kanai MA-O, Hosoe J, Shojima N, et al. Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. (1546-1718 (Electronic)).

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

34. Richardson TA-O, Sanderson EA-O, Elsworth BA-O, Tilling KA-O, Davey Smith GA-O. Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study. (1756-1833 (Electronic)).
35. Pérez-Pérez A, Toro A, Vilariño-García T, Maymó J, Guadix P, Dueñas JL, et al. Leptin action in normal and pathological pregnancies. (1582-4934 (Electronic)).
36. Xin Q, Xin G, Li L, Sun W, Jiang W, Wang J, et al. Association study of hypertension susceptibility genes ITGA9, MOV10, and CACNB2 with preeclampsia in Chinese Han population. (1476-4954 (Electronic)).
37. Veas CJ, Aguilera Vc Fau - Muñoz IJ, Muñoz Ij Fau - Gallardo VI, Gallardo Vi Fau - Miguel PL, Miguel Pl Fau - González MA, González Ma Fau - Lamperti LI, et al. Fetal endothelium dysfunction is associated with circulating maternal levels of sE-selectin, sVCAM1, and sFlt-1 during pre-eclampsia. (1476-4954 (Electronic)).
38. Schild D, Lio Y-c, Collins DW, Tsomondo T, Chen DJ. Evidence for Simultaneous Protein Interactions between Human Rad51 Paralogs *. Journal of Biological Chemistry. 2000;275(22):16443-9.
39. Bonilla B, Hengel SR, Grundy MK, Bernstein KA. RAD51 Gene Family Structure and Function. (1545-2948 (Electronic)).
40. Wang J, Zhao S-J, Wang L-L, Lin X-X, Mor G, Liao A-H. Leukocyte immunoglobulin-like receptor subfamily B: A novel immune checkpoint molecule at the maternal-fetal interface. Journal of Reproductive Immunology. 2023;155:103764.
41. Blois SM, Dveksler G, Vasta GR, Freitag N, Blanchard V, Barrientos G. Pregnancy Galectinology: Insights Into a Complex Network of Glycan Binding Proteins. (1664-3224 (Electronic)).
42. Schrader S, Unverdorben L, Hutter S, Knabl J, Schmoeckel E, Meister S, et al. Overexpression of galectin-4 in placentas of women with gestational diabetes. (1872-7603 (Electronic)).
43. Barban N, Jansen R, de Vlaming R, Vaez A, Mandemakers JJ, Tropf FC, et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. Nature Genetics. 2016;48(12):1462-72.
44. Kim TH, Park JM, Jo SH, Kim MY, Nojima H, Ahn YH. Effects of low-fat diet and aging on metabolic profiles of Creb3l4 knockout mice. (2044-4052 (Print)).
45. Li X, Liu L, Whitehead C, Li J, Thierry B, Le TD, et al. Identifying preeclampsia-associated genes using a control theory method. Briefings in Functional Genomics. 2022;21(4):296-309.
46. Luo SA-O, Clarke SA-O, Ramanan AA-O, Thompson SA-O, Langefeld CA-O, Marion MC, et al. Platelet Glycoprotein Ib α-Chain as a Putative Therapeutic Target for Juvenile Idiopathic Arthritis: A Mendelian Randomization Study. (2326-5205 (Electronic)).
47. Fejzo M, Rocha N, Cimino I, Lockhart S, Petry C, Kay R, et al. Fetally-encoded GDF15 and maternal GDF15 sensitivity are major determinants of nausea and vomiting in human pregnancy. bioRxiv. 2023:2023.06.02.542661.
48. Jansen LAW, Nijsten K, Limpens J, van Eekelen R, Koot MH, Grooten IJ, et al. Perinatal outcomes of infants born to mothers with hyperemesis gravidarum: A systematic review and meta-analysis. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2023;284:30-51.
49. Hoggard N, Hunter L, Duncan JS, Williams LM, Trayhurn P, Mercer JG. Leptin and leptin receptor mRNA and protein expression in the murine fetus and placenta. Proceedings of the National Academy of Sciences. 1997;94(20):11073-8.
50. Beaumont RN, Flatley C, Vaudel M, Wu X, Chen J, Moen G-H, et al. Genome-wide association study of placental weight in 179,025 children and parents reveals distinct and shared genetic influences between placental and fetal growth. medRxiv. 2022:2022.11.25.22282723.
51. Agrawal S, Cerdeira AS, Redman C, Vatish M. Meta-Analysis and Systematic Review to Assess the Role of Soluble FMS-Like Tyrosine Kinase-1 and Placenta Growth Factor Ratio in Prediction of Preeclampsia: The SaPPPhirE Study. (1524-4563 (Electronic)).

Public

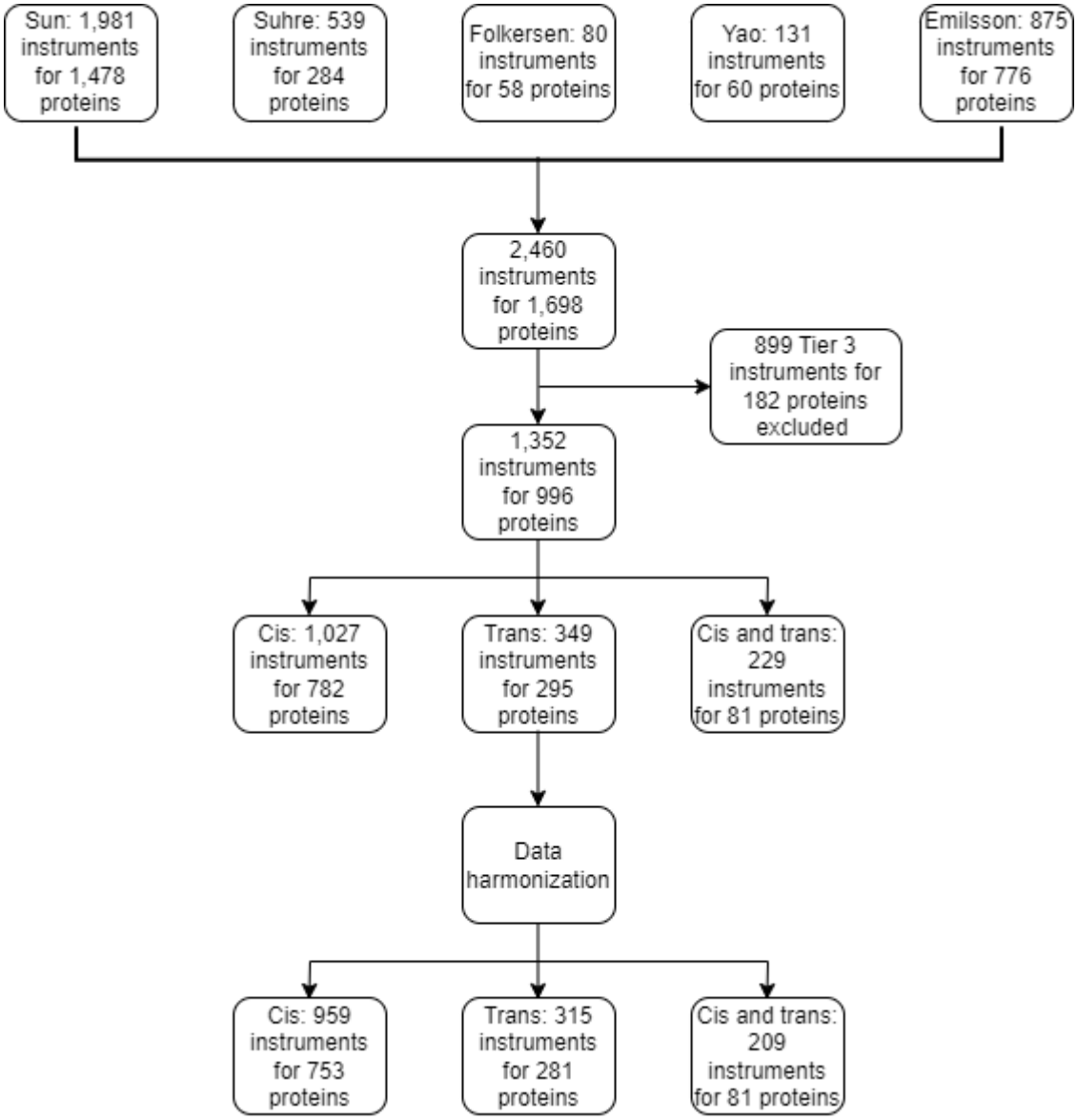
medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

52. Andrikos AA-OX, Andrikos D, Schmidt B, Birdir C, Kimmig R, Gellhaus A, et al. Course of the sFlt-1/PlGF ratio in fetal growth restriction and correlation with biometric measurements, feto-maternal Doppler parameters and time to delivery. (1432-0711 (Electronic)).
53. Chang Y-S, Chen C-N, Jeng S-F, Su Y-N, Chen C-Y, Chou H-C, et al. The sFlt-1/PlGF ratio as a predictor for poor pregnancy and neonatal outcomes. Pediatrics & Neonatology. 2017;58(6):529-33.
54. Fauman EB, Hyde C. An optimal variant to gene distance window derived from an empirical definition of cis and trans protein QTLs. BMC Bioinformatics. 2022;23(1):169.
55. Paltiel L, Anita H, Skjerden T, Harbak K, Bækken S, Nina Kristin S, et al. The biobank of the Norwegian Mother and Child Cohort Study – present status. Norsk Epidemiologi. 2014;24(1-2).
56. Corfield EC, Frei O, Shadrin AA, Rahman Z, Lin A, Athanasiu L, et al. The Norwegian Mother, Father, and Child cohort study (MoBa) genotyping data resource: MoBaPsychGen pipeline v.1. bioRxiv. 2022:2022.06.23.496289.
57. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nature Genetics. 2016;48(10):1279-83.
58. Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, et al. Computationally efficient whole-genome regression for quantitative and binary traits. Nature Genetics. 2021;53(7):1097-103.
59. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics. 2015;47(3):291-5.
60. Liu B, Gloudemans MJ, Rao AS, Ingelsson E, Montgomery SB. Abundant associations with gene expression complicate GWAS follow-up. Nature Genetics. 2019;51(5):768-9.

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

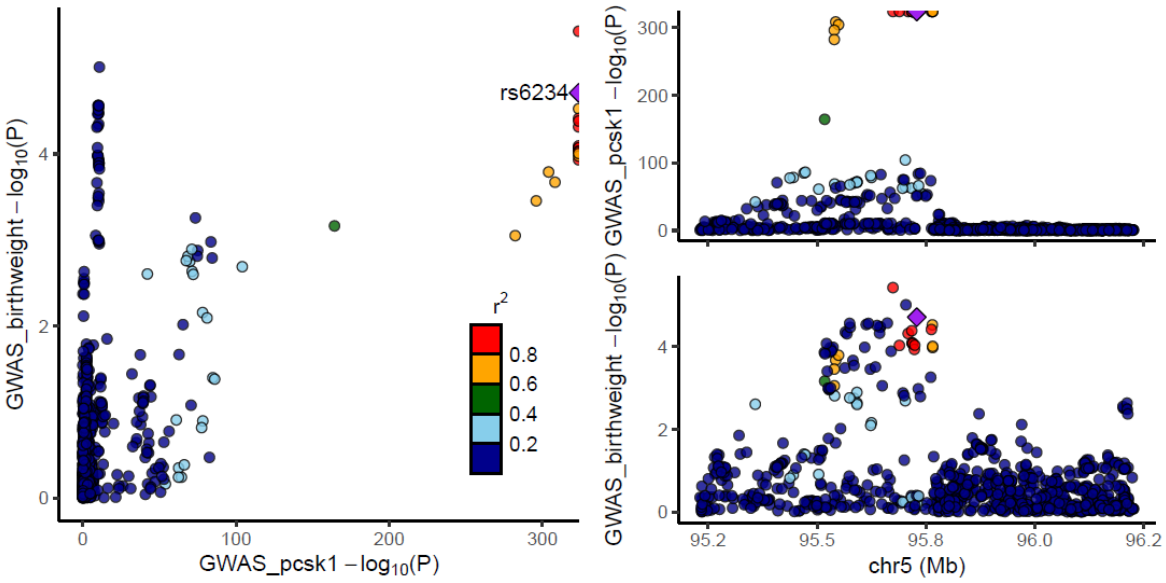
Supplementary



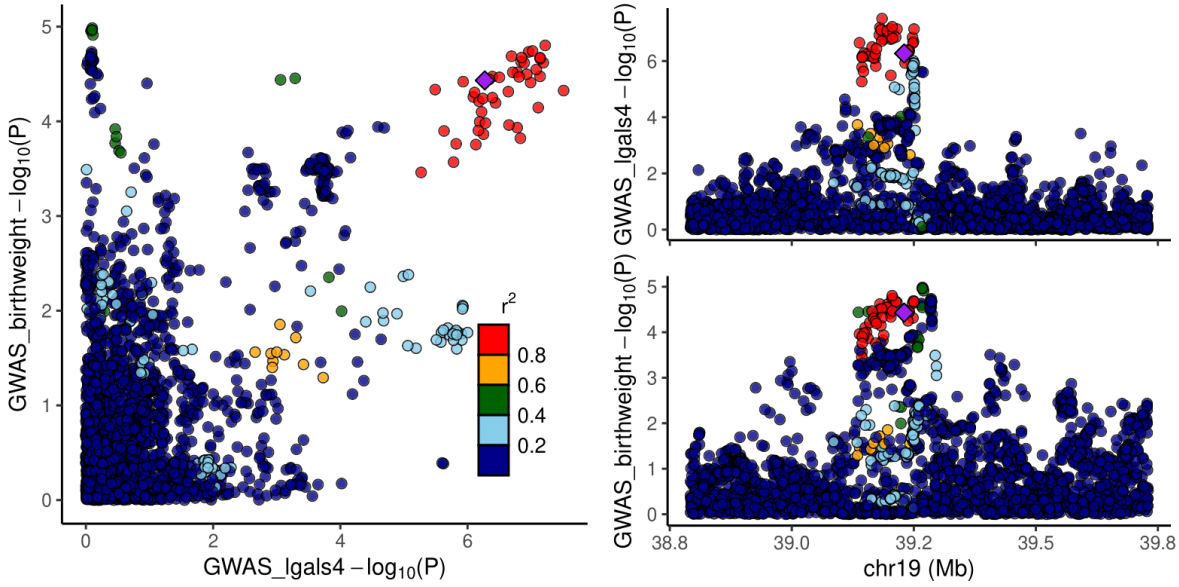
Supplementary figure 1 – a workflow of instruments (pQTL) used in our analysis

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.



Supplementary figure 2 - Genetic associations with PCSK1 and offspring birth weight in maternal analysis. Each data point represents one genetic variant. The 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 left panel, the plot shows the correlation between log 10 p-values for the genetic association with protein levels (x axis) and birth weight (y axis). The right panels show recombination plots for the protein (top panel) and birth weight (bottom panel) with log10 p-values in the y axis and chromosome positions in the x axis.

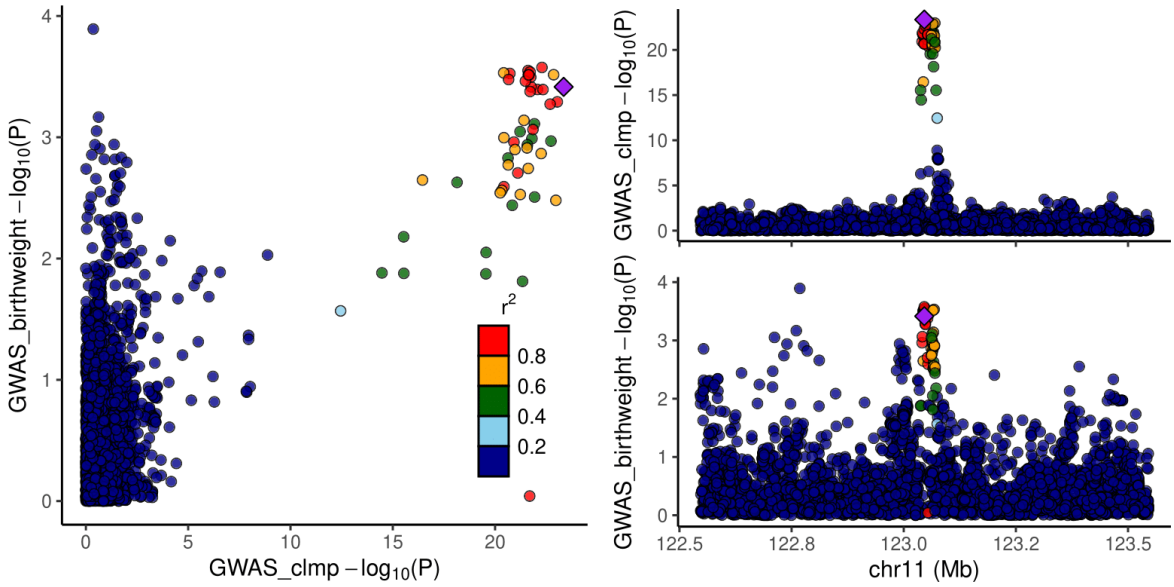


Supplementary figure 3 - Genetic associations with LGALS4 and offspring birth weight in maternal analysis. Each data point represents one genetic variant. The purple diamond represents the selected pQTL. Colours indicate the R^2 values for strength of linkage disequilibrium between the genetic variant and the pQTL. In the left panel, the plot shows the correlation between log 10 p-values for the genetic association with protein levels (x axis) and birth weight (y axis). The right panels show recombination

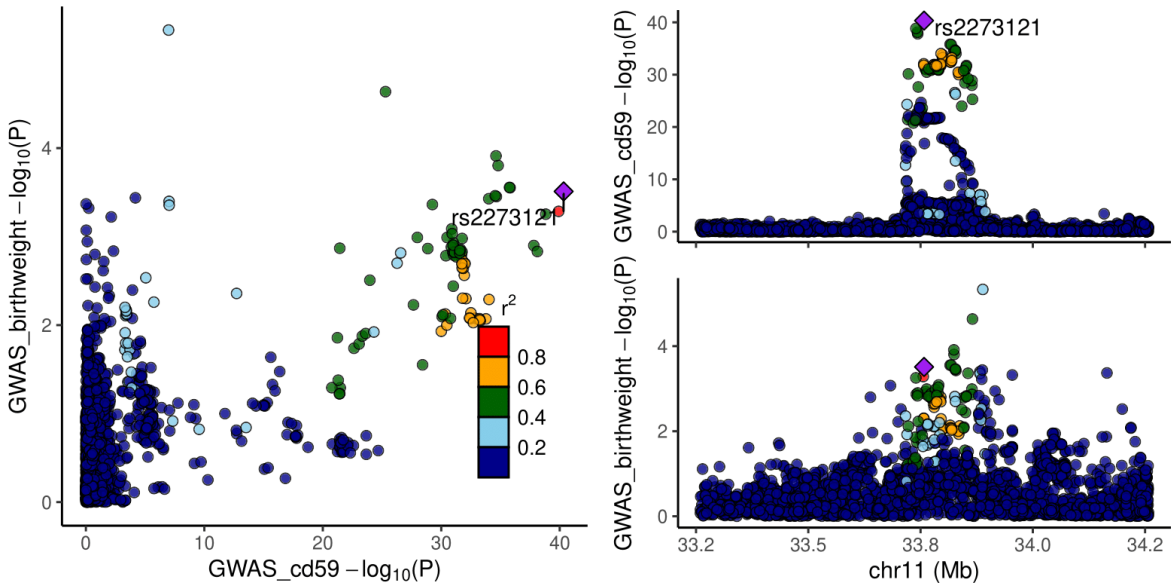
Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

plots for the protein (top panel) and birth weight (bottom panel) with log10 p-values in the y axis and chromosome positions in the x axis.



Supplementary figure 4 - Genetic associations with CLMP and offspring birth weight in maternal analysis. Each data point represents one genetic variant. The 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 left panel, the plot shows the correlation between log 10 p-values for the genetic association with protein levels (x axis) and birth weight (y axis). The right panels show recombination plots for the protein (top panel) and birth weight (bottom panel) with log10 p-values in the y axis and chromosome positions in the x axis.

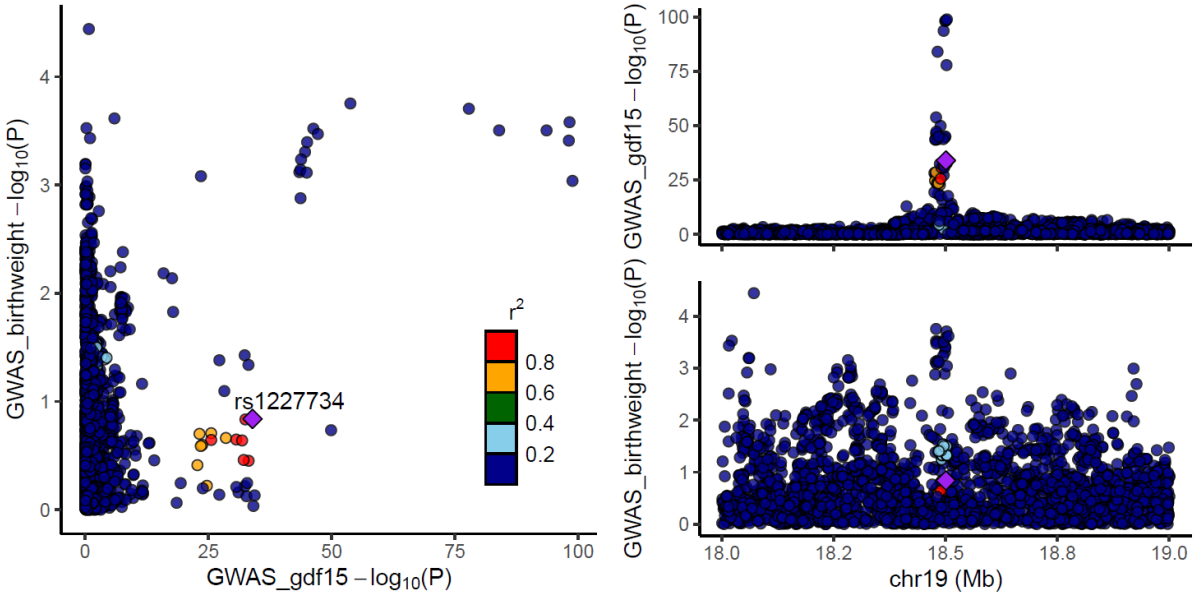


Supplementary figure 5 - Genetic associations with CD59 and offspring birth weight in maternal analysis. Each data point represents one genetic variant. The purple diamond represents the selected pQTL. Colours indicate the R^2 values for strength of

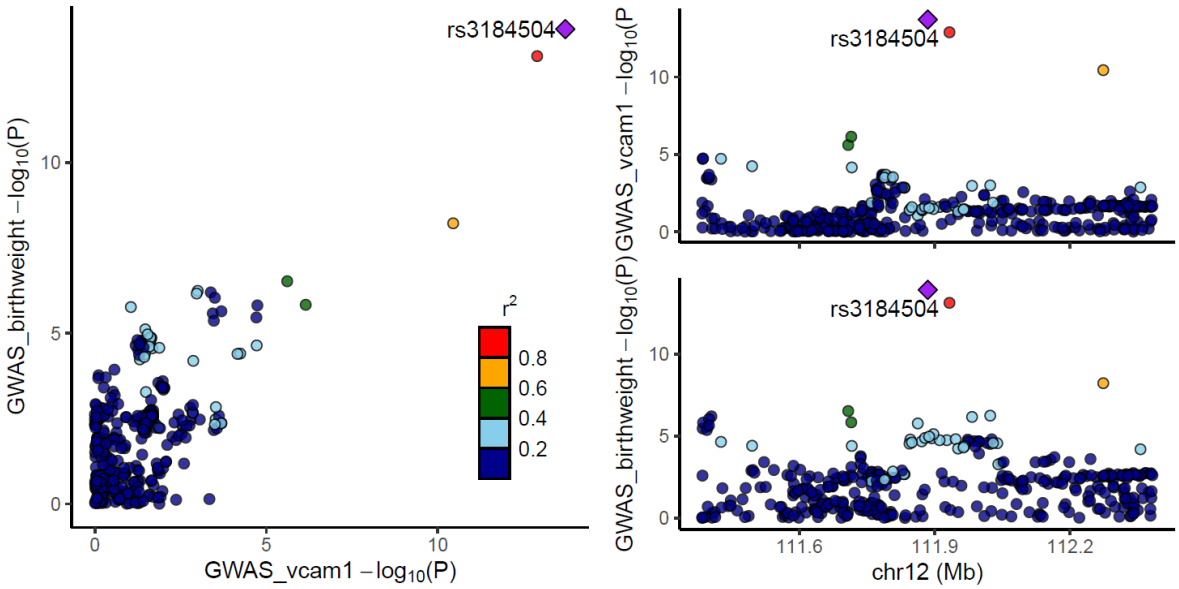
Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

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



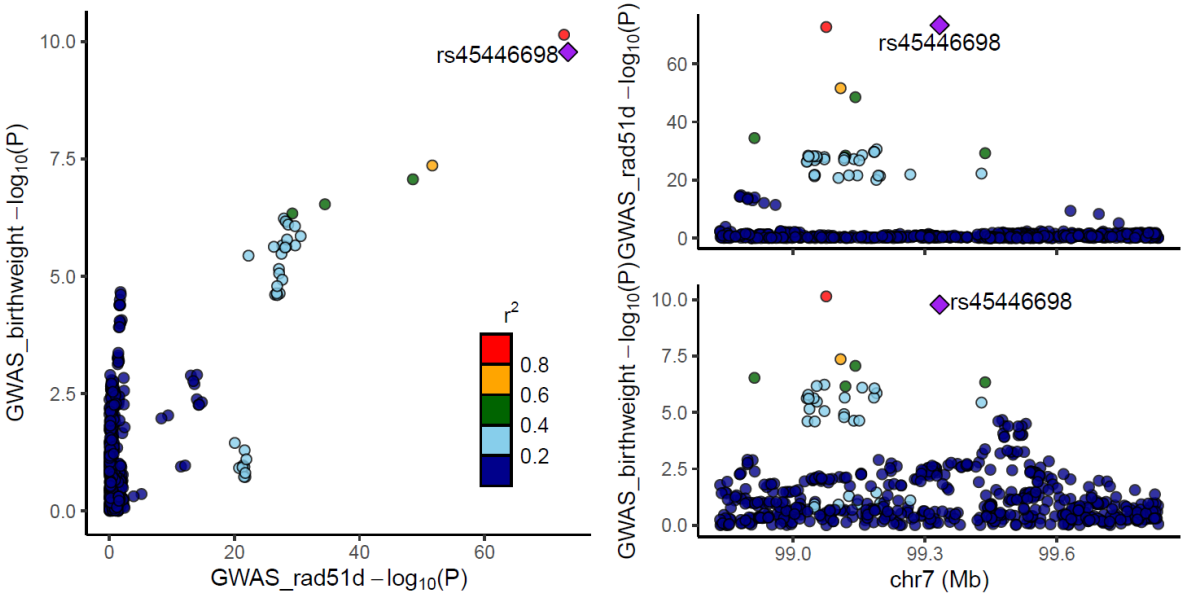
Supplementary figure 6 - Genetic associations with GDF15 and offspring birth weight in maternal analysis. Each data point represents one genetic variant. The purple diamond represents the selected pQTL. Colours indicate the R^2 values for strength of linkage disequilibrium between the genetic variant and the pQTL. In the left panel, the plot shows the correlation between log 10 p-values for the genetic association with protein levels (x axis) and birth weight (y axis). The right panels show recombination plots for the protein (top panel) and birth weight (bottom panel) with log10 p-values in the y axis and chromosome positions in the x axis.



Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

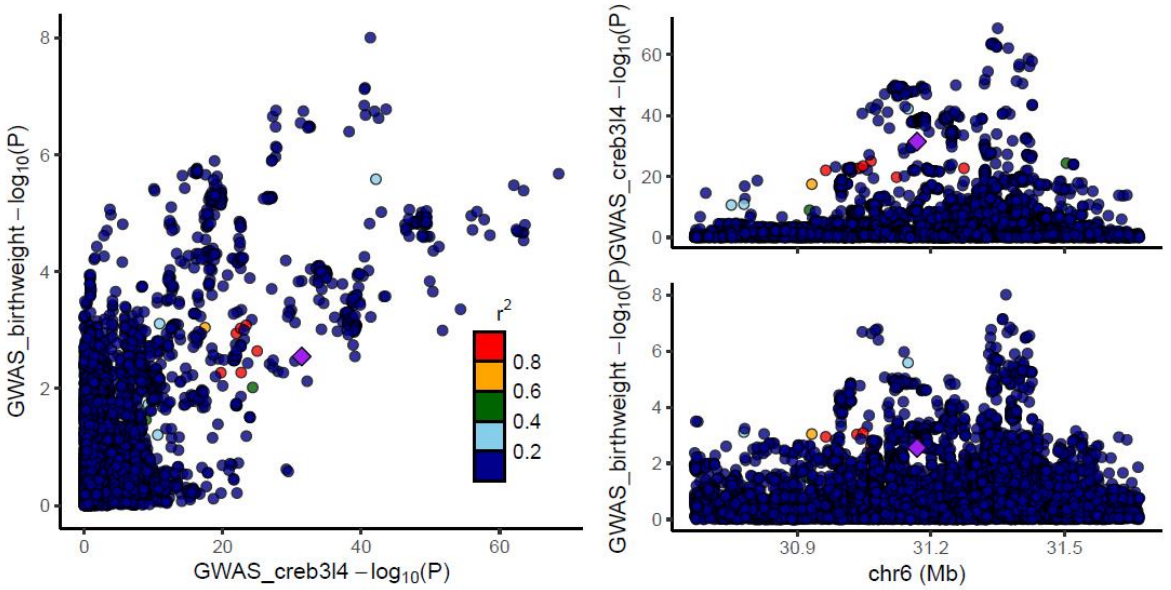
Supplementary figure 7 - Genetic associations with VCAM1 and offspring birth weight in maternal analysis. Each data point represents one genetic variant. The purple diamond represents the selected pQTL. Colours indicate the R^2 values for strength of linkage disequilibrium between the genetic variant and the pQTL. In the left panel, the plot shows the correlation between log 10 p-values for the genetic association with protein levels (x axis) and birth weight (y axis). The right panels show recombination plots for the protein (top panel) and birth weight (bottom panel) with log10 p-values in the y axis and chromosome positions in the x axis.



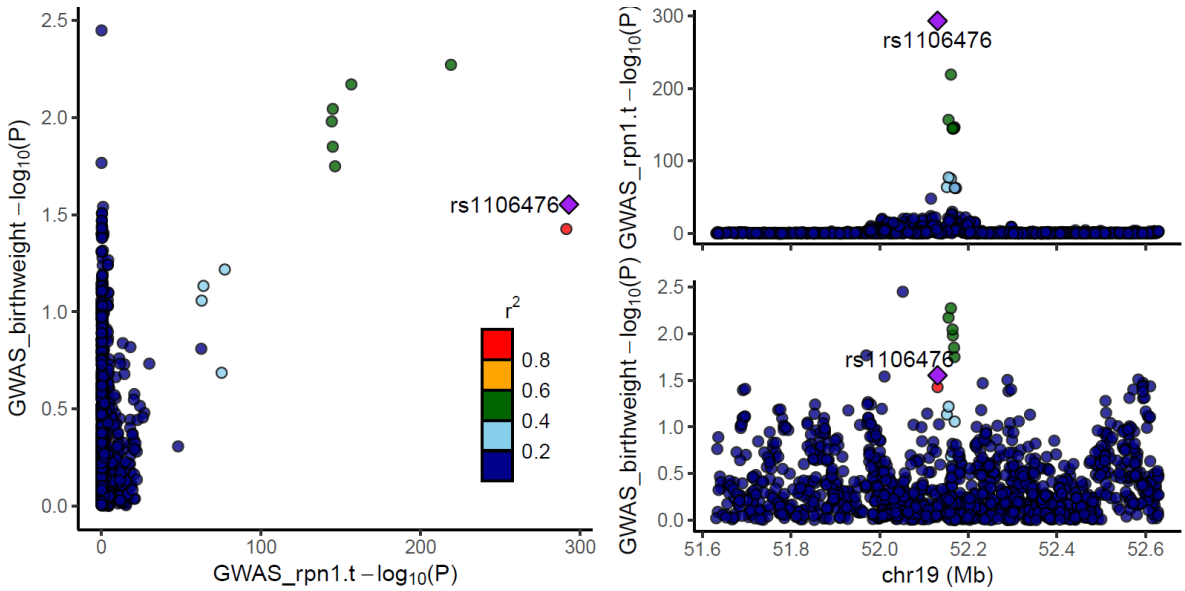
Supplementary figure 8 - Genetic associations with RAD51D and offspring birth weight in maternal analysis. Each data point represents one genetic variant. The purple diamond represents the selected pQTL. Colours indicate the R^2 values for strength of linkage disequilibrium between the genetic variant and the pQTL. In the left panel, the plot shows the correlation between log 10 p-values for the genetic association with protein levels (x axis) and birth weight (y axis). The right panels show recombination plots for the protein (top panel) and birth weight (bottom panel) with log10 p-values in the y axis and chromosome positions in the x axis.

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.



Supplementary figure 9 - Genetic associations with CREB3L4 and offspring birth weight in maternal analysis. Each data point represents one genetic variant. The 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 left panel, the plot shows the correlation between log 10 p-values for the genetic association with protein levels (x axis) and birth weight (y axis). The right panels show recombination plots for the protein (top panel) and birth weight (bottom panel) with log10 p-values in the y axis and chromosome positions in the x axis.

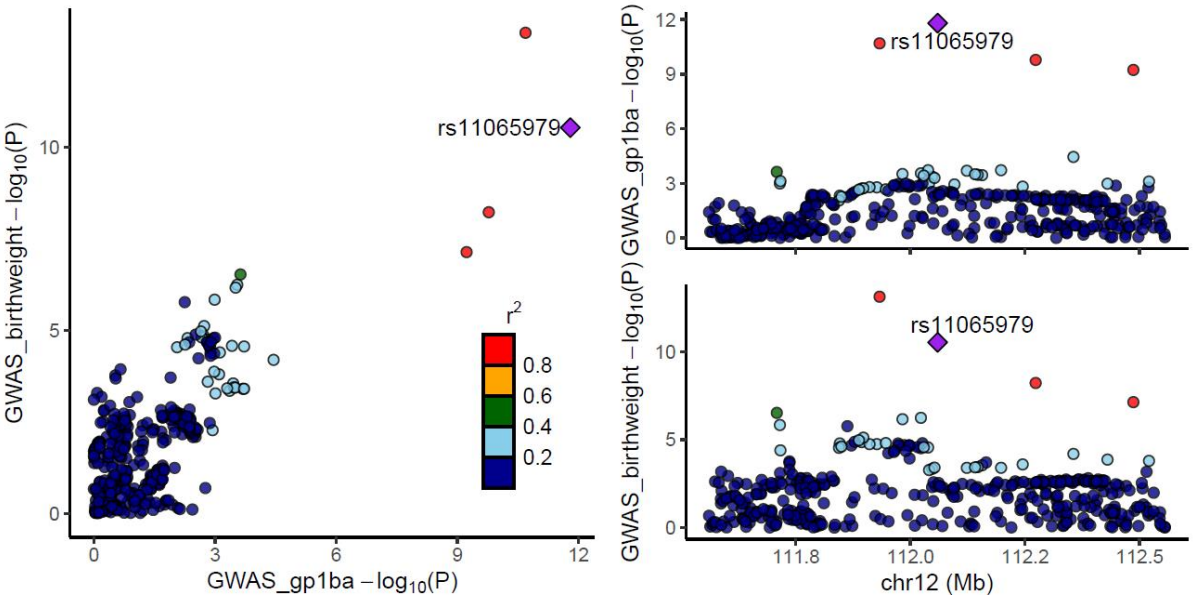


Supplementary figure 10 - Genetic associations with RPN1 in trans, and offspring birth weight in maternal analysis. Each data point represents one genetic variant. The purple diamond represents the selected pQTL. Colours indicate the R^2 values for strength of linkage disequilibrium between the genetic variant and the pQTL. In the left panel, the plot shows the correlation between log 10 p-values for the genetic

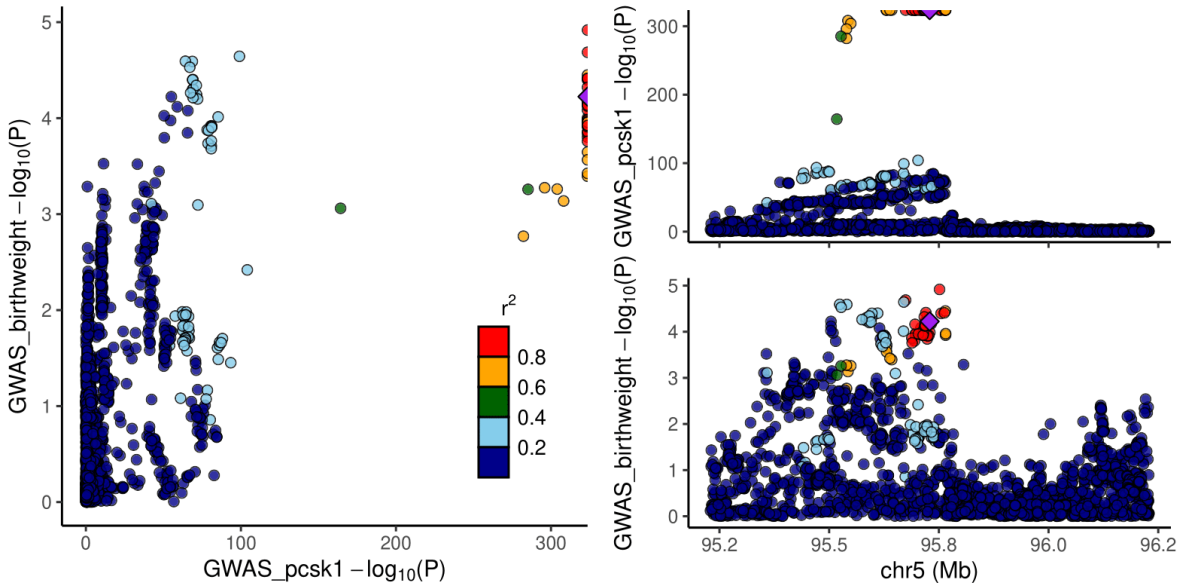
Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

association with protein levels (x axis) and birth weight (y axis). The right panels show recombination plots for the protein (top panel) and birth weight (bottom panel) with log10 p-values in the y axis and chromosome positions in the x axis.



Supplementary figure 11 - Genetic associations with GP1BA and offspring birth weight in maternal analysis. Each data point represents one genetic variant. The 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 left panel, the plot shows the correlation between log 10 p-values for the genetic association with protein levels (x axis) and birth weight (y axis). The right panels show recombination plots for the protein (top panel) and birth weight (bottom panel) with log10 p-values in the y axis and chromosome positions in the x axis.

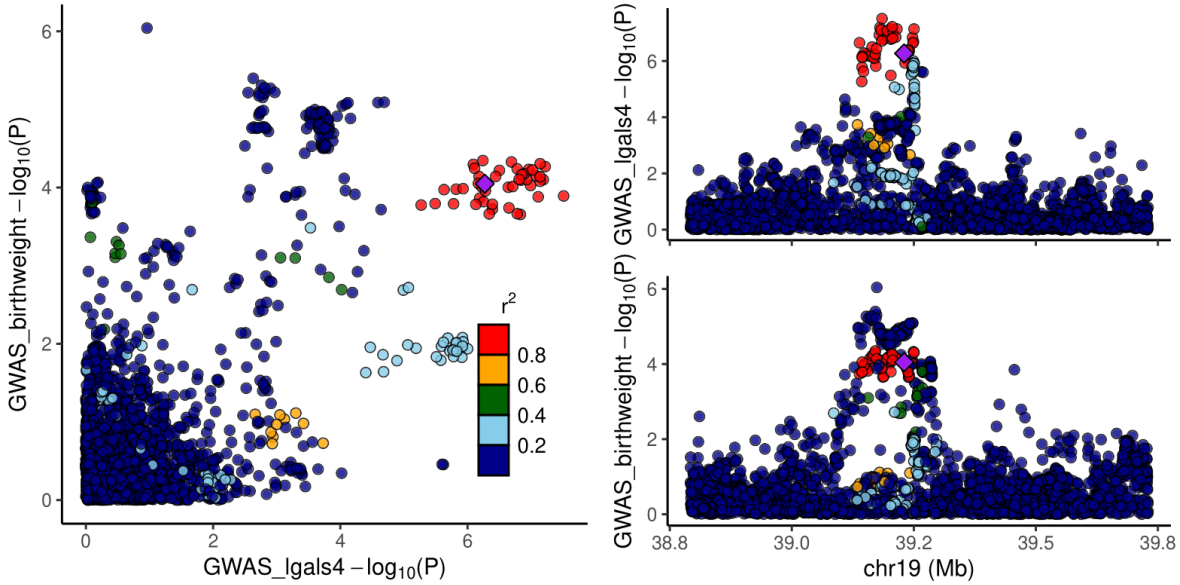


Supplementary figure 12 - Genetic associations with PCSK1 and offspring birth weight in offspring analyses. Each data point represents one genetic variant. The purple diamond represents the selected pQTL. Colours indicate the R^2 values for strength of linkage disequilibrium between the genetic variant and the pQTL. In the left panel, the

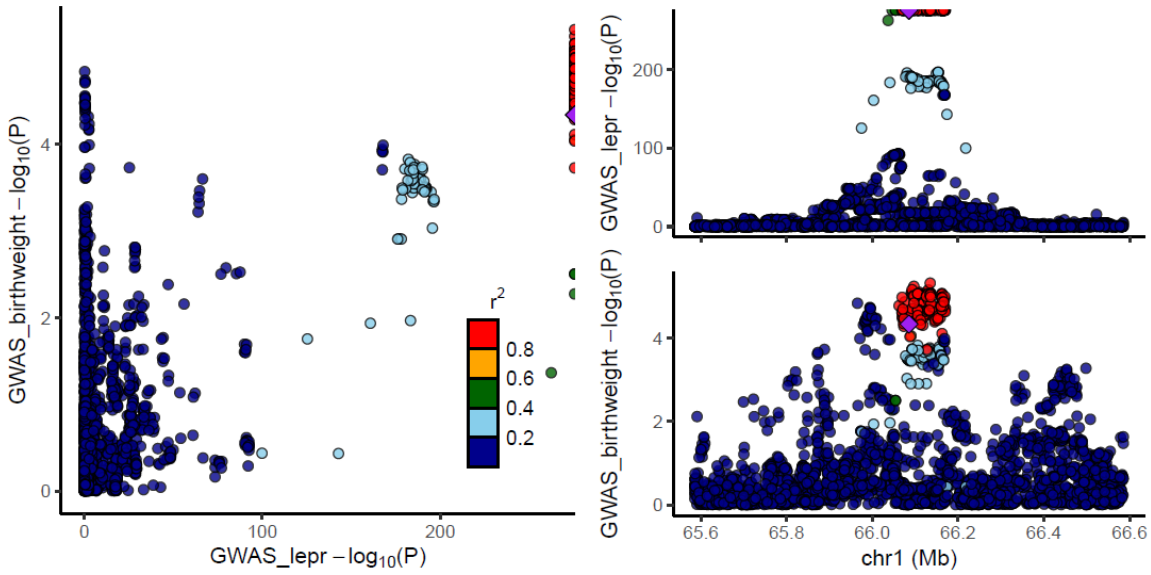
Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

plot shows the correlation between log 10 p-values for the genetic association with protein levels (x axis) and birth weight (y axis). The right panels show recombination plots for the protein (top panel) and birth weight (bottom panel) with log10 p-values in the y axis and chromosome positions in the x axis.



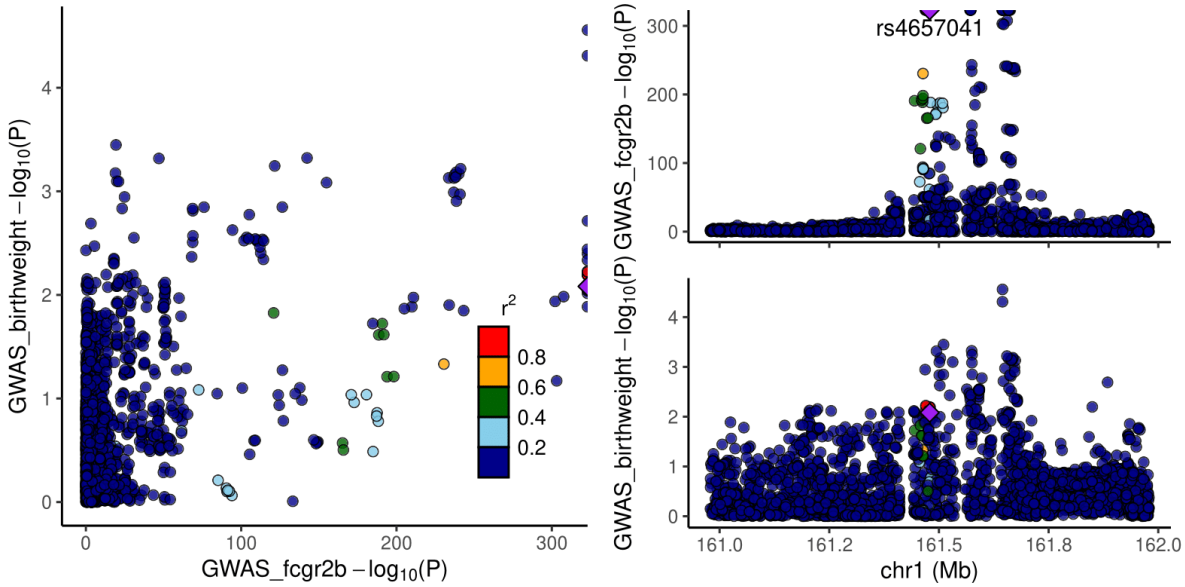
Supplementary figure 13 - Genetic associations with LGALS4 and offspring birth weight in offspring analysis. Each data point represents one genetic variant. The purple diamond represents the selected pQTL. Colours indicate the R^2 values for strength of linkage disequilibrium between the genetic variant and the pQTL. In the left panel, the plot shows the correlation between log 10 p-values for the genetic association with protein levels (x axis) and birth weight (y axis). The right panels show recombination plots for the protein (top panel) and birth weight (bottom panel) with log10 p-values in the y axis and chromosome positions in the x axis.



Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Supplementary figure 14 - Genetic associations with LEPR and offspring birth weight in offspring analysis. Each data point represents one genetic variant. The 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 left panel, the plot shows the correlation between log 10 p-values for the genetic association with protein levels (x axis) and birth weight (y axis). The right panels show recombination plots for the protein (top panel) and birth weight (bottom panel) with log10 p-values in the y axis and chromosome positions in the x axis.



Supplementary figure 15 - Genetic associations with FCGR2B and offspring birth weight in offspring analysis. Each data point represents one genetic variant. The 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 left panel, the plot shows the correlation between log 10 p-values for the genetic association with protein levels (x axis) and birth weight (y axis). The right panels show recombination plots for the protein (top panel) and birth weight (bottom panel) with log10 p-values in the y axis and chromosome positions in the x axis.

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Supplementary figure 16 - Fetal single cell gene expression expressed in the pancreas for PCSK1, where the organs are shown in yellow and the single cells demonstrating expression in navy blue.

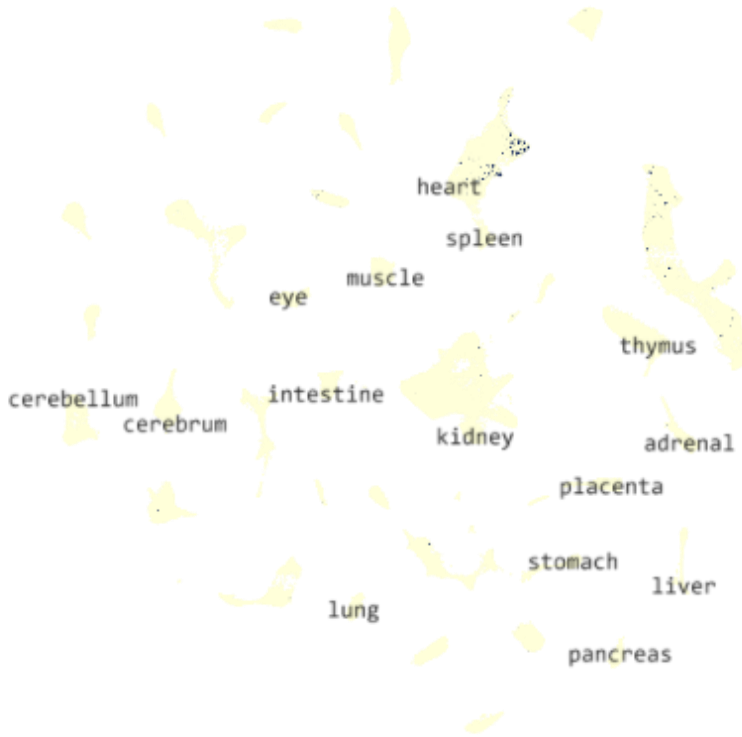
Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Supplementary figure 17 - Fetal single cell gene expression expressed tissue-wide for LGALS4, where the organs are shown in yellow and the single cells demonstrating expression in navy blue.

Public

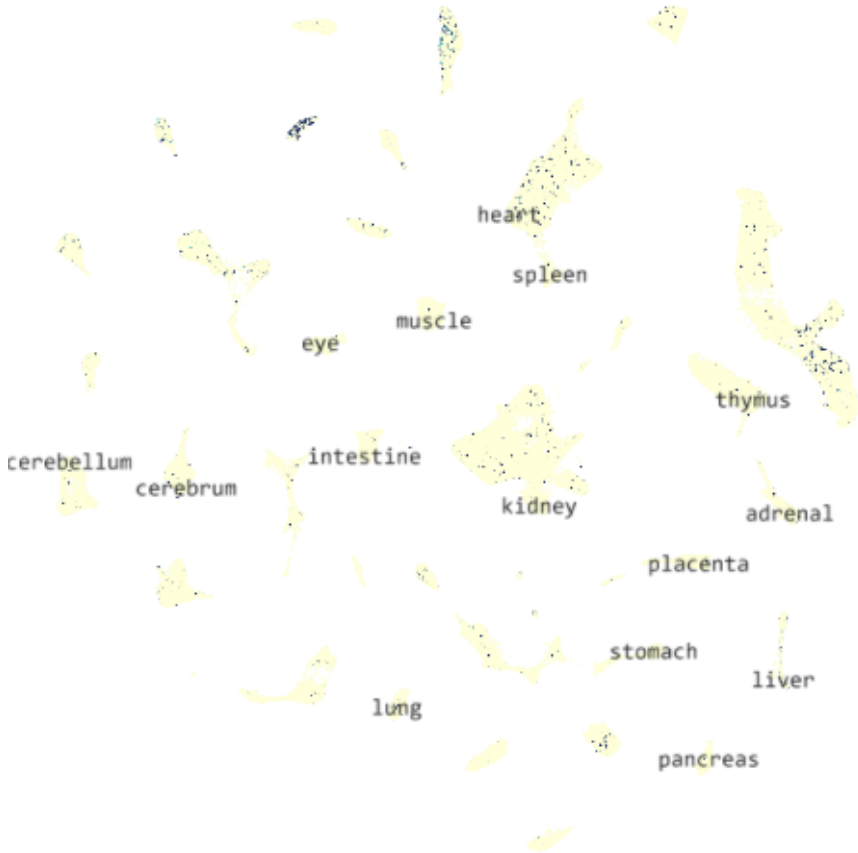
medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.



Supplementary figure 18 - Fetal single cell gene expression expressed in the heart, kidney and thymus for FCGR2B, where the organs are shown in yellow and the single cells demonstrating expression in navy blue.

Public

medRxiv preprint doi: https://doi.org/10.1101/2023.10.20.23297135; this version posted October 21, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.



Supplementary figure 19 - Fetal single cell gene expression expressed tissue-wide for LEPR, where the organs are shown in yellow and the single cells demonstrating expression in navy blue.

Public