





Placental multi-omics integration identifies candidate functional genes for birthweight

Fasil Tekola-Ayele ^{1✉}, Xuehuo Zeng¹, Suvo Chatterjee¹, Marion Ouidir¹, Corina Lesseur ², Ke Hao ³, Jia Chen², Markos Tesfaye⁴, Carmen J. Marsit ⁵, Tsegaselassie Workalemahu⁶ & Ronald Wapner⁷

Abnormal birthweight is associated with increased risk for cardiometabolic diseases in later life. Although the placenta is critical to fetal development and later life health, it has not been integrated into largescale functional genomics initiatives, and mechanisms of birthweight-associated variants identified by genome wide association studies (GWAS) are unclear. The goal of this study is to provide functional mechanistic insight into the causal pathway from a genetic variant to birthweight by integrating placental methylation and gene expression with established GWAS loci for birthweight. We identify placental DNA methylation and gene expression targets for several birthweight GWAS loci. The target genes are broadly enriched in cardiometabolic, immune response, and hormonal pathways. We find that methylation causally influences *WNT3A*, *CTDNEP1*, and *RANBP2* expression in placenta. Multi-trait colocalization identifies *PLEKHA1*, *FES*, *CTDNEP1*, and *PRMT7* as likely functional effector genes. These findings reveal candidate functional pathways that underpin the genetic regulation of birthweight via placental epigenetic and transcriptomic mechanisms. Clinical trial registration; ClinicalTrials.gov, NCT00912132.

¹Division of Population Health Research, Division of Intramural Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA. ²Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ³Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁴Section of Sensory Science and Metabolism (SenSMet), National Institute on Alcohol Abuse and Alcoholism & National Institute of Nursing Research, National Institutes of Health, Bethesda, MD, USA. ⁵Gangarosa Department of Environmental Health, Rollins School of Public Health of Emory University, Atlanta, GA, USA. ⁶Department of Obstetrics and Gynecology, Maternal-Fetal Medicine, University of Utah, Salt Lake City, UT, USA. ⁷Department of Obstetrics and Gynecology, Columbia University, New York, NY, USA. ✉email: ayelef@mail.nih.gov

Individuals with birthweight in the lower or upper bounds of the population distribution are at greater risk of cardiovascular and metabolic diseases during adulthood and mortality during childhood^{1–5}. Therefore, advances in understanding the biological underpinnings of birthweight will have broad relevance to understanding the etiology of complex diseases and developing therapeutics. Genome-wide association studies (GWAS) have identified several genetic variants associated with fetal growth and birthweight^{6–12}. However, most of the GWAS variants map to noncoding regions of the genome, and the functional mechanism through which the variants influence birthweight remains unclear.

Multi-omic molecular features of the placenta, one of the most functionally-relevant organs for fetal growth¹³, can provide a remarkable opportunity to investigate the biological embedding of genetic influences on birthweight. Despite growing evidence supporting the key role of the placenta in regulating fetal growth and subsequent health, data on placental multi-omic molecular features integrated with genetic variation are rare¹⁴. Large-scale tissue-specific gene expression regulation resources such as the Genotype-Tissue Expression Portal (GTEx) do not include placenta¹⁵. As a result, birthweight GWAS signals that may be explained by regulatory effects on placental epigenome and transcriptome are not understood. Establishing that a variant driving the GWAS signal also influences nearby gene expression (i.e., expression quantitative trait locus, *cis*-eQTL) in a relevant tissue suggests a putative regulatory mechanism¹⁶. Recent work has indicated that integrating additional information from DNA methylation (DNAm) quantitative trait locus (*cis*-mQTL) strengthens the evidence for causality^{17–19}. Therefore, the prioritization of candidate functional genes for birthweight can be facilitated by elucidating the coordinated relationships of GWAS variants with placental methylation and gene expression.

In this study, we provide functional mechanistic insight into birthweight by integrating GWAS variants known to be associated with birthweight with placental methylation and gene expression. Our primary analyses integrate placental DNAm, gene expression, and genotype data from participants of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) Fetal Growth Studies–Singleton cohort^{20,21} with GWAS summary statistics of birthweight⁶. First, we determine whether birthweight GWAS variants regulate nearby gene expression (i.e., *cis*-eQTL) and DNAm (i.e., *cis*-mQTL) in the placenta, and assess functional features of the regulated genes. Second, for birthweight GWAS variants found to be both *cis*-eQTL and *cis*-mQTL, we elucidate the directional relationships among the genetic variants, DNAm and gene expression in the placenta using mediation²² and Mendelian Randomization²³ approaches. Third, we apply multitrait colocalization¹⁷ at each GWAS locus found to be both *cis*-eQTL and *cis*-mQTL to investigate whether all three traits (i.e., birthweight, placental DNAm, and gene expression) share the same causal variant and identify candidate genes for functional follow-up. Loci found to be both *cis*-eQTL and *cis*-mQTL are further assessed using an independent dataset from the Rhode Island Child Health Study (RICHS)²⁴.

Results

Overview of the analysis workflow. An overview of the analysis workflow is presented in Fig. 1. *cis*-eQTL and *cis*-mQTL loci were identified using the NICHD Fetal Growth Studies–Singleton dataset. A total of 291 placental samples with genotype and DNAm data were included in *cis*-mQTL analysis. Samples were obtained from pregnant women with self-identified Hispanic ($n = 97$), White ($n = 74$), Black ($n = 71$), and Asian race/ethnicity

($n = 49$); mean age of 27.8 years; mean pre-pregnancy body mass index (BMI) of 25.1 kg/m²; and 46.5% female infants. A subset of 71 samples with genotype and RNA-seq data (22 Hispanic, 21 White, 20 Black, 8 Asian) were included in *cis*-eQTL analysis. The characteristics of the study participants including maternal age, race/ethnicity, pre-pregnancy BMI, gestational duration, parity, and birthweight showed no significant difference between the participants with DNAm data and the subset with RNA-seq data (Supplementary Data 1). Out of 286 single-nucleotide polymorphisms (SNPs) found to be associated with birthweight at $P < 5 \times 10^{-8}$ (i.e., 190 lead SNPs with $P < 6.6 \times 10^{-9}$ and 96 SNPs with $6.6 \times 10^{-9} < P < 5 \times 10^{-8}$) by Warrington et al.⁶, 273 SNPs were available in our dataset (Supplementary Data 2). A total of 7901 unique protein-coding gene mRNAs and 104,053 CpG DNAm sites found within 1 Mb distance from the 273 SNPs were included in eQTL and mQTL analyses, respectively. For loci found to be both eQTL and mQTL, we performed mediation, Mendelian Randomization, and multitrait colocalization analyses. Further evaluation of the co-occurring eQTL and mQTL associations was performed in RICHS ($n = 159$; 77.4% White; mean maternal age of 30.9 years; mean pre-pregnancy BMI of 26.6 kg/m²; 49.7% female infants (Supplementary Data 1)). Functional annotation, regulatory enrichment, and pathway analyses were performed on all eQTL and mQTL target genes.

***cis*-eQTLs and *cis*-mQTLs in the placenta for birthweight GWAS loci.** At a false discovery rate (FDR) adjusted $P < 0.05$, we identified 32 significant eQTL-eGene associations consisting 26/273 eQTL SNPs (vs. 0.04% expected genome-wide²⁵, $P = 1.97 \times 10^{-54}$, hypergeometric test) associated with expression of 29 protein-coding genes (eGenes) at FDR adjusted $P < 0.05$. We also identified 813 significant mQTL-DNAm associations consisting 158/273 mQTL SNPs (vs. 0.2% expected genome-wide²⁵, $P < 1 \times 10^{-300}$, hypergeometric test) associated with 778 DNAm sites (Supplementary Data 3, 4)."

Co-occurring *cis*-eQTL and *cis*-mQTL. We found an overlap between eQTL and mQTL at 23 SNPs (14.3% vs. 3.1% expected genome-wide²⁵, $P = 1.84 \times 10^{-11}$, hypergeometric test), forming a total of 197 eQTL-mQTL triplets (consisting 23 SNPs, 26 genes, and 152 DNAm sites) (Fig. 2a). The majority (88.5%) of *cis*-eQTLs were also *cis*-mQTLs (vs. 18.3% genome-wide²⁵, $P = 8.51 \times 10^{-19}$, hypergeometric test) (Fig. 2b). Among the 197 triplets, more SNPs were associated with gene expression and DNAm in opposite directions than in the same direction (68% vs. 32%), consistent with expectation that methylation near promoter and transcription start sites is related to inhibited transcription^{26,27}. A significant inverse correlation was found between association coefficients of mQTL and eQTL in the placenta ($\rho = -0.22$, $P = 7.90 \times 10^{-157}$); the correlation strengthened when the analysis was restricted to 48 triplet subsets that included DNAm sites located within genes identical to the corresponding eGenes ($\rho = -0.58$, $P = 1.5 \times 10^{-5}$) (Fig. 2c, d). Further, we evaluated whether the co-occurring eQTL-mQTL triplets in the placenta have a similar relationship in blood using 23 triplets available in blood eQTL from GTEx¹⁵ and middle-age blood mQTL from ARIES²⁸. The inverse correlation between mQTL and eQTL coefficients in placenta remained ($\rho = -0.80$, $P = 6.61 \times 10^{-6}$) but no correlation was observed in blood ($\rho = 0.05$, $P = 0.82$) (Fig. 2e), suggesting that placenta may be the main organ in which these regulatory effects are targeted.

Mediation analysis using the causal inference test. For the 197 eQTL-mQTL triplets in which 23 co-occurring SNPs were associated with both placental gene expression and DNAm, we

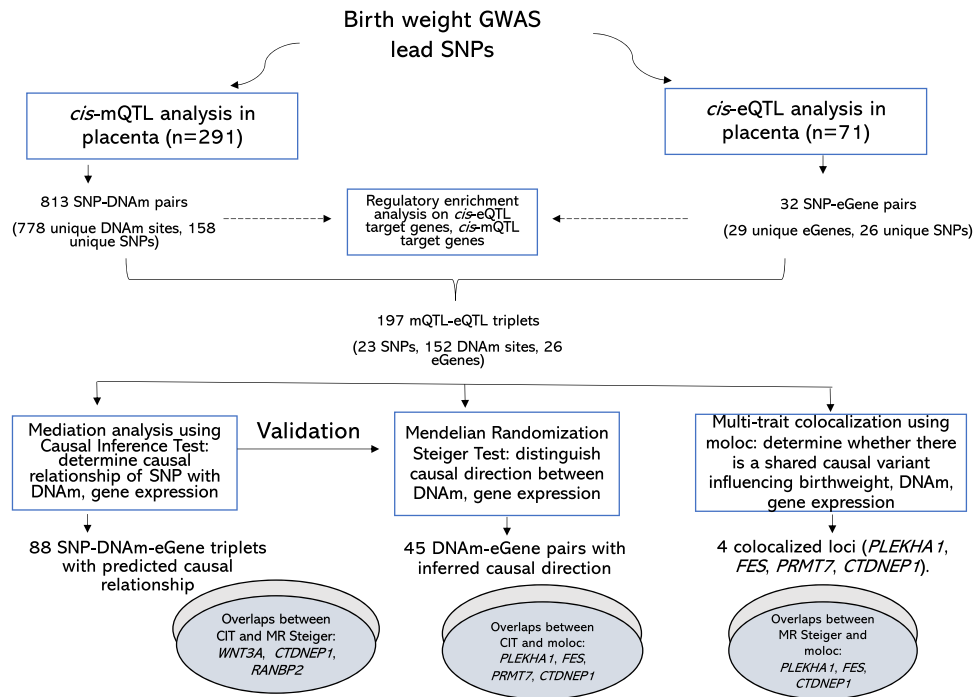


Fig. 1 Overview of study workflow. DNAm denotes DNA methylation, and eGene denotes an eQTL target gene in which gene expression in the placenta is regulated by a nearby genetic variant.

performed a causal inference test (CIT)²² to distinguish among three scenarios in the causal relationship between DNAm and gene expression: (1) a birthweight GWAS SNP alters DNAm in the placenta, which in turn causally influences gene expression (SNP→DNAm→ gene expression, labeled as SME mediation). (2) a birthweight GWAS SNP alters gene expression in the placenta, which causally influences DNAm (SNP→gene expression→ DNAm, labeled as SEM mediation suggesting a passive role of DNAm on genetic regulation of gene expression²⁶, and (3) a birthweight GWAS SNP influences DNAm and gene expression in the placenta through independent pathways, but DNAm and gene expression have no causal relationship (SNP→gene expression or SNP→DNAm, denoted as independent SE/SM). CIT found evidence of a causal relationship in 44.7% (88/197) triplets consisting 15 SNPs, 17 eGenes, and 81 DNAm sites (Supplementary Data 5). Within the 88 predicted causal relationships, 84 (95.5%) were SME, predicting DNAm to causally influence gene expression and 4 (4.5%) were SEM, predicting gene expression to causally influence DNAm (Fig. 3a). There was a significant strong inverse dose-response correlation between the regression coefficients of mQTL and eQTL for triplets with predicted causal relationship (i.e., in the SEM and SME categories) ($\rho = -0.46$, $P = 8.13 \times 10^{-6}$) but not for triplets in the independent SE/SM category ($\rho = -0.15$, $P = 0.15$) (Fig. 3b, c), consistent with the causal inference findings. As expected, the distance between the DNAm site and the corresponding gene transcription start site was shorter for triplets with a predicted causal relationship (mean 134 bp) than independent or unclassified triplets (mean 183 bp and 224 bp, respectively).

Mendelian Randomization analysis for validating direction of causality between DNAm and gene expression. While the CIT has strength in distinguishing causal relationship from horizontal pleiotropy (independent pathways), it has limitations in distinguishing SME from SEM and might misclassify some causal relationships as independent²⁹. To complement these limitations, we assessed the 197 eQTL-mQTL triplets using the Mendelian

Randomization (MR) Steiger directionality test²³. The MR Steiger test provided evidence of significant inferred causal directions between DNAm and gene expression for 45 DNAm-eGene pairs consisting 11 SNPs, 37 DNAm sites, and 12 eGenes ($P < 0.05$; FDR $P < 0.05$ for 13 DNAm-eGene pairs, < 0.1 for 36 DNAm-eGene pairs, < 0.25 for all 45 DNAm-eGene pairs) (Supplementary Data 6). There was significant enrichment of DNAm-eGene pairs with causal directions (45/197 (22.8%) vs. 20.3% in a genome-wide dataset³⁰, $P = 1.17 \times 10^{-2}$, hypergeometric test). As expected, the majority (8/11; 72.7%) of SNPs with MR Steiger's evidence of inferred causal directions were among SNPs in CIT's predicted causal relationships (i.e., SME or SEM), and the remaining (3/11) were among SNPs in CIT's independent SE/SM category. Of the 45 DNAm-eGene pairs, 7 were from DNAm to gene expression (ME), and 38 were from gene expression to DNAm (EM). At the 7 ME DNAm-eGene pairs, the SNPs explained more variation in DNA methylation (ranging from 34.9% for cg01745370 to 74% for cg06106103) than in gene expression (ranging from 12.8% for BTBD16 to 19.2% for PLEKHA1); therefore, the alternative explanation (i.e., reverse causal pathway from gene expression to methylation) is unlikely.

Consistent SME predictions were found by the CIT and MR Steiger test for three triplets (rs708122→cg02991924→WNT3A expression; rs222857→cg04106389 located in DLG4→CTDNEP1 expression; and rs12104672→cg01745370 located in GCC2 promoter region→RANBP2 expression; 3/45 vs. 88/197, $P = 2.96 \times 10^{-5}$, hypergeometric test). WNT3A is expressed in a few tissues only, its highest expression being in the placenta. Further diagnostic plots for inferring causal direction confirmed an SME relationship in rs708122→cg02991924→WNT3A expression (Fig. 4a–f). The three genes with causal directions triangulated by the two methods (i.e., WNT3A, CTDNEP1, and RANBP2) have been implicated by mice studies in placental development and vascularization, embryogenesis, and fetal growth^{31–34}.

Multitrait colocalization. The causal inference frameworks in CIT and MR Steiger tests cannot establish the possibility of

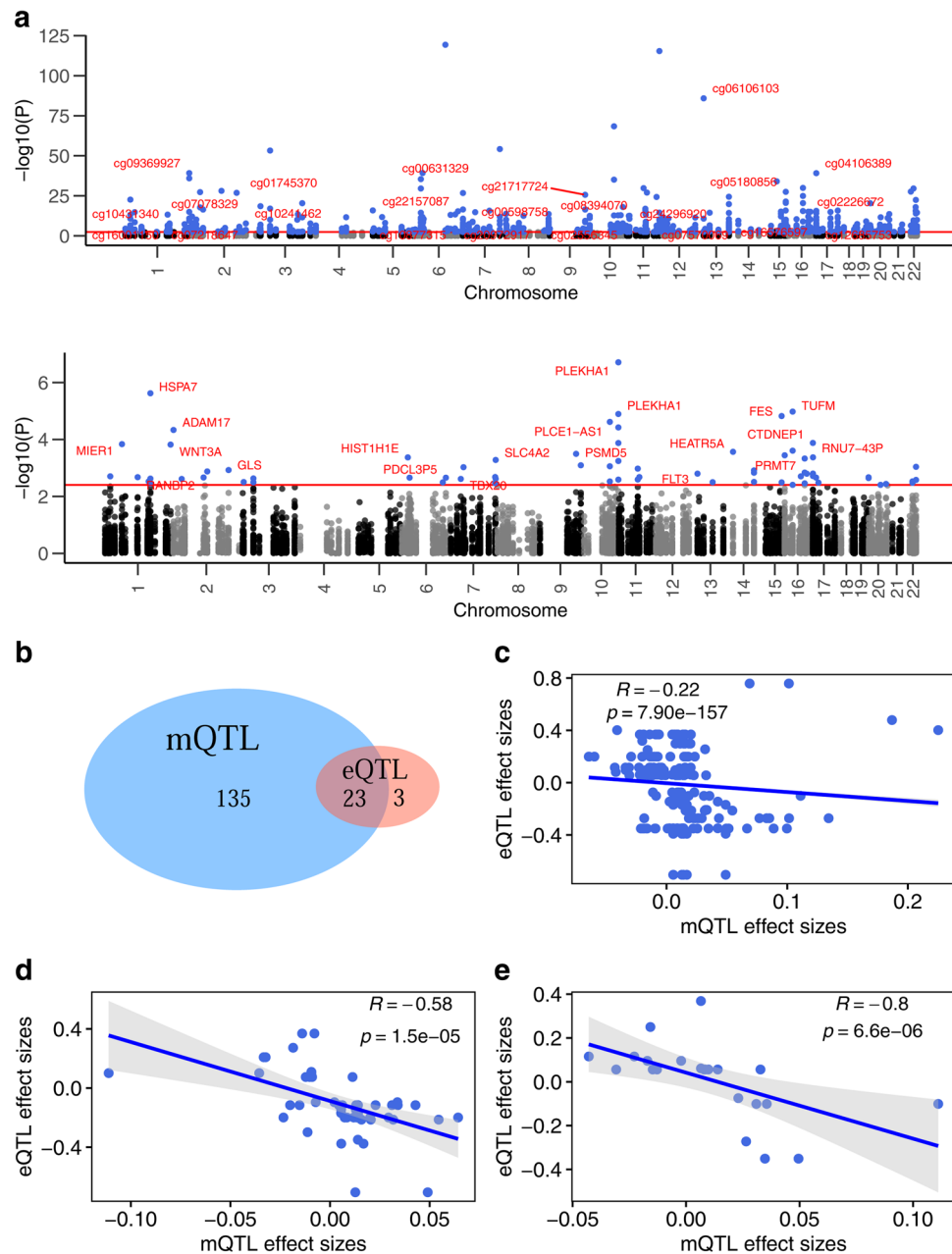


Fig. 2 Birthweight-associated genetic variants with co-occurring *cis*-eQTL and *cis*-mQTL effect. **a** mQTL in the upper panel and eQTL in the lower panel. The red horizontal lines represent the 5% FDR threshold. CpGs and genes with labels correspond to loci in which a genetic variant was associated with both DNA methylation and gene expression. For ease of readability, only one CpG and one gene was labeled per locus. **b** Venn Diagram showing the number of birthweight GWAS loci that were found to be eQTL and mQTL in the placenta. **c–e** P is based on a two-sided Spearman correlation test performed between placental mQTL and eQTL association coefficients for all 197 eQTL–mQTL triplets (**c**), for 48 triplets consisting DNA methylation sites within the eQTL target genes (**d**), and for 23 triplets that overlap with blood eQTL from GTEx and mQTL from ARIES database (**e**). **d**, **e** Error bands in gray represent standard error of the mean.

colocalized causal loci because the association of a variant with multiple traits may be due to distinct causal SNPs that are in linkage disequilibrium (LD). To determine whether there is a shared causal variant and identify functional genes underlying birthweight, DNA methylation and gene expression in the placenta, we performed a multitrait colocalization test using moloc¹⁷. A significant colocalization result implies that the alternative explanation that association signals are due to LD between the associated SNPs and separate causal variants for the traits is unlikely. We found evidence of genetic colocalization (posterior probability of association (PPA) ≥ 0.8) between

placental DNA methylation, gene expression, and birthweight at three loci and narrowly missed the cut-off at one locus: (1) *PLEKHA1* gene expression colocalized with DNAm sites located in *PLEKHA1* and *HTRA1* and birthweight (PPA = 0.85 and 0.88). (2) *FES* gene expression colocalized with nine DNAm sites in *FES* and birthweight (PPA = 0.82–0.91). (3) *PRMT7* gene expression colocalized with a DNAm site in *SMPD3* and birthweight (PPA = 0.8). (4) A fourth locus at chr17p13.1 showed evidence for colocalization of all three traits (*CTDNEP1* gene expression, DNAm sites in *DLG4* and *CLDN7*, and birthweight) that narrowly missed the cut-off (PPA = 0.77–0.78). At the chr17p13.1

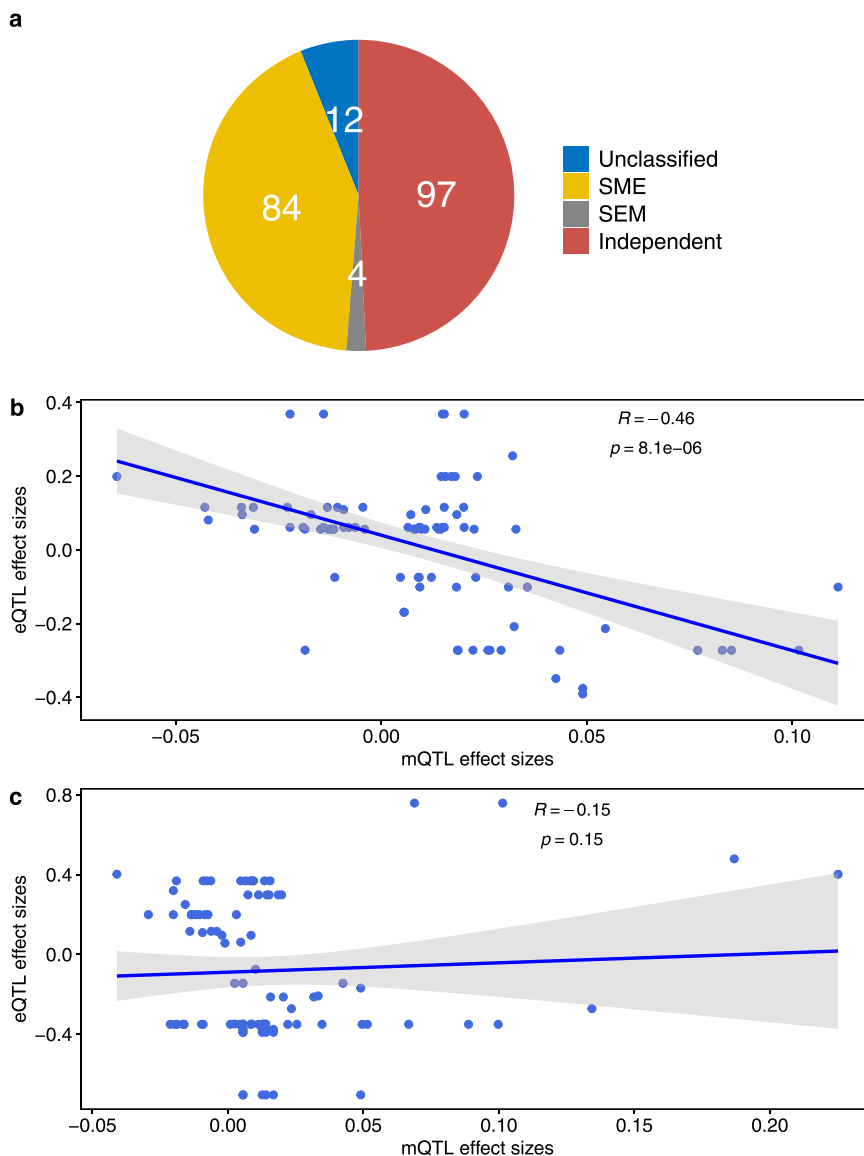


Fig. 3 Causal inference test result. **a** Pie-chart showing the number of eQTL-mQTL triplets predicted as “SEM” mediation representing SNP effect mediated via placental gene expression on nearby DNA methylation, “SME” mediation representing SNP effect mediated via placental DNA methylation on placental gene expression, “independent SE/SM” effect denoting SNP effect on placental DNA methylation and gene expression via independent pathways, and “unclassified” relationship. **b, c** Error bands in gray represent standard error of the mean and *P* is based on a two-sided Spearman correlation test. The dose-response correlation between eQTL and mQTL association coefficients was only significant for triplets with SEM or SME predicted causal relationship (**b**) but not with independent SE/SM relationship (**c**).

locus, there was a strong evidence of colocalization between birthweight and DNAm (PPA = 0.88–0.92), and between DNAm and *EIF5A* gene expression (PPA = 0.88–0.97) (Table 1).

For all four colocalized loci, the best-colocalized SNP identified by moloc was either the lead birthweight GWAS SNP (for *CTDNEP1*) or a nearby SNP in high LD ($r^2 = 0.84$ – 0.98 in 1000 Genomes European samples with the lead SNP) for *PLEKHA1*, *FES*, and *PRMT7*). For all four loci, the direction of association of the best-colocalized SNP with methylation was opposite to its association with gene expression. All four loci with evidence of colocalization overlapped with the triplets that had a predicted causal relationship by CIT and three loci overlapped with the triplets that had an inferred causal direction by MR Steiger test ($M \rightarrow E$ for *CTDNEP1*; $E \rightarrow M$ for *PLEKHA1* and *FES*). The colocalized genes are broadly expressed in several tissues, with the location of protein expression predicted as intracellular, and play key molecular functions including lipid-binding (*PLEKHA1*),

growth factor binding (*HTRA1*), chromatin regulator, kinase, and transferase (*PRMT7*, *FES*) (Supplementary Data 7). The colocalized DNAm sites are predominantly annotated to genomic regions with intron, promoter, and CpG shore features (Supplemental Fig. S1).

Evaluation of colocalization in non-placental tissues. We assessed whether the four loci with evidence of colocalization in the placenta also harbor shared causal variants for blood DNA methylation, gene expression, and birthweight. To do this, we applied moloc using summary statistics data for middle-age blood mQTL from ARIES²⁸, adult blood eQTL from GTEx¹⁵, and GWAS of birthweight⁶. No significant (PPA ≥ 0.8) multitrait colocalization was found. There was suggestive evidence of colocalization between *PLEKHA1* gene expression in blood and birthweight (PPA = 0.76) (Supplementary Data 8). To expand our assessment to non-placenta tissues (with the caveat that

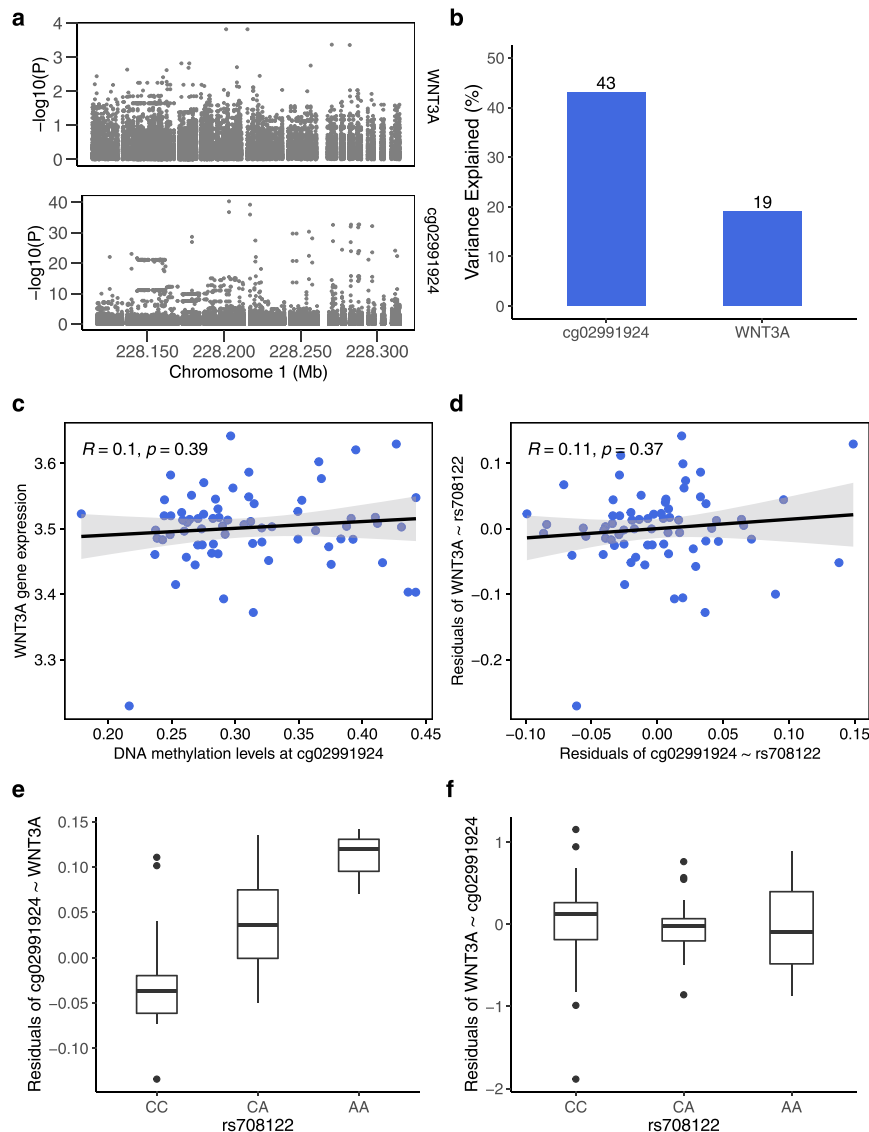


Fig. 4 Direction of the relationship between birthweight-associated SNP, DNA methylation and *WNT3A* gene expression in placenta. **a** Association between all SNPs 100 kb up- and downstream of the birthweight GWAS SNP rs708122 and *WNT3A* gene expression (upper panel) and DNA methylation at cg02991924 (lower panel). The strongest association was found for the index SNP or its nearest linkage disequilibrium (LD) proxies. **b** rs708122 accounts for higher variance in DNA methylation than in *WNT3A* expression based on MR Steiger test. **c, d** Error bands in gray represent standard error of the mean and P is based on a two-sided Spearman correlation test. DNA methylation and *WNT3A* expression were not correlated. **e, f** Box plots indicate median (middle line), 25th, 75th percentile (box) and 5th and 95th percentile (whiskers) as well as outliers (single points). $N = 71$ individuals. rs708122 was associated with DNA methylation after regressing out *WNT3A* expression (**e**) (Kruskal Wallis test $P = 3.06 \times 10^{-7}$) but was not associated with residuals of *WNT3A* after regressing out DNA methylation (**f**) (Kruskal Wallis test $P = 0.57$).

methylation data are not available), we tested colocalization for gene expression and the four birthweight loci in 48 GTEx tissues using coloc³⁵. Whereas no significant colocalization was found for *PLEKHA1* and *CTDNEP1*, evidence of colocalization was found for *PRMT7* and *FES* in some tissues (Supplementary Data 9).

Evaluation of co-occurring *cis*-eQTL and *cis*-mQTLs in the placenta in an independent dataset. Further assessment of the 197 co-occurring eQTL-mQTL pairs identified was done using placental data from the Rhode Island Child Health Study²⁴ (RICHS; $n = 159$). Data on 123 mQTLs and 22 eQTLs were available in RICHS. We found that 61% (75/123) mQTLs tested showed significant associations ($P < 0.05$) (vs. 0.2% genome wide²⁵, $P = 7.41 \times 10^{-169}$, hypergeometric test). Most of the

associated loci overlapped with the four colocalization target genes (*PLEKHA1*, *FES*, *CTDNEP1*, *PRMT7*) and the three genes showing consistent causal predictions by CIT and MR Steiger tests (*WNT3A*, *CTDNEP1*, *RANBP2*). 22.7% (5/22) eQTLs tested showed significant associations ($P < 0.05$) (vs. 0.04% genome-wide²⁵, $P = 5.40 \times 10^{-16}$, hypergeometric test). All eQTLs were also mQTLs in RICHS, mirroring the pattern we found in the NICHD Fetal Growth Studies dataset (Supplementary Data 10). The eQTLs for *PLEKHA1* and *RANBP2* are consistent with a previous report by Warrington et al.⁶

Association between colocalized gene expression and birthweight. We assessed whether placental expression of the colocalized genes was associated with birthweight using linear regression adjusted for fetal sex and gestational age. *PLEKHA1*

Table 1 Multitrait colocalization evidence for loci with causal-variant sharing between birthweight, placental DNAm, and gene expression.

GWAS lead SNP	SNP position (hg19)	eGene	DNAm site (nearest gene)	DNAm site position (hg19)	DNAm site relation to gene	Distance from GWAS SNP to eGene/DNAm site	N SNPs	PPA for G, E, M	PPA for GM, E	PPA for GE, M	PPA for GEM
rs71486610	Chr10:124134803	PLEKHA1	cg02556345 (PLEKHA1)	124181965	Body	Intronic 147.1 kb	726	4.1 × 10 ⁻²	3.7 × 10 ⁻²	2.4 × 10 ⁻²	0.88
			cg06474225 (HTRA1)	124228770	Body	Intronic 93.9 kb	686	4.1 × 10 ⁻²	2.0 × 10 ⁻²	3.2 × 10 ⁻²	0.85
			cg14366292 (HTRA1)	124328787	Body	Intronic 194.0 kb	684	9.3 × 10 ⁻⁵	7.8 × 10 ⁻⁵	1.0 × 10 ⁻⁸	0.92
			cg17033087 (DMB1)	124329533	Body	Intronic 194.7 kb	682	1.2 × 10 ⁻⁴	1.1 × 10 ⁻⁵	0.90	0.88
			cg11976790 (DMB1)	124320064	TSS	Intronic 185.3 kb	686	1.7 × 10 ⁻⁴	1.7 × 10 ⁻²	0.88	0.88
			cg27015047 (DMB1)	124320805	Body	Intronic 186.0 kb	687	2.1 × 10 ⁻⁴	1.4 × 10 ⁻²	0.88	0.88
			cg25647583 (DMB1)	91427184	TSS	Intronic 2.1 kb	507	4.6 × 10 ⁻³	2.2 × 10 ⁻²	1.3 × 10 ⁻⁵	0.91
			cg2092439* (FES)	91429916	Body	Intronic 0.6 kb	510	2.4 × 10 ⁻³	9.6 × 10 ⁻³	1.7 × 10 ⁻²	0.91
			cg06330618 (FES)	91428456	Exon	Intronic 0.8 kb	509	1.9 × 10 ⁻²	1.9 × 10 ⁻²	1.0 × 10 ⁻²	0.90
			cg09397246 (FES)	91427361	TSS	Intronic 1.9 kb	507	6.2 × 10 ⁻³	2.3 × 10 ⁻²	1.7 × 10 ⁻²	0.88
rs1868158 rs222857	Chr16:68398924 Chr17:7164563	PRMT7 CTDNEP1**	cg07718650 (FES)	91434288	Body	Intronic 5.0 kb	527	2.1 × 10 ⁻²	2.1 × 10 ⁻²	1.6 × 10 ⁻²	0.88
			cg26405020 (FES)	91427363	TSS	Intronic 1.9 kb	507	6.6 × 10 ⁻³	2.9 × 10 ⁻²	2.9 × 10 ⁻²	0.86
			cg08724371 (FES)	91442212	-	Intronic 12.9 kb	544	8.4 × 10 ⁻³	2.2 × 10 ⁻²	2.4 × 10 ⁻²	0.85
			cg26899598* (FES)	91429088	Body	Intronic 0.2 kb	509	1.1 × 10 ⁻²	3.4 × 10 ⁻²	2.9 × 10 ⁻²	0.82
			cg18661868* (FES)	91427965	5'UTR	Intronic 1.3 kb	508	1.6 × 10 ⁻²	2.6 × 10 ⁻²	3.1 × 10 ⁻²	0.80
			cg02226672 (SMPD3)	68398533	Body	6.4 kb 0.4 kb	373	3.6 × 10 ⁻²	3.0 × 10 ⁻²	2.5 × 10 ⁻²	0.80
			cg17265693 (CLDN7)	7166490	TSS	9.3 kb 1.9 kb	434	4.5 × 10 ⁻³	2.1 × 10 ⁻²	2.6 × 10 ⁻²	0.78
			cg19697794 (DLG4)	7117125	Body	9.3 kb 47.4 kb	473	5.4 × 10 ⁻³	2.0 × 10 ⁻²	4.0 × 10 ⁻³	0.77
			cg19466160* (DLG4)	7117160	Body	9.3 kb 47.4 kb	473	5.1 × 10 ⁻³	1.9 × 10 ⁻²	4.6 × 10 ⁻³	0.77
			cg17265693 (CLDN7)	7217367	Body	45.8 kb 52.8 kb	434	3.7 × 10 ⁻⁷	2.8 × 10 ⁻⁷	3.6 × 10 ⁻⁵	0.92
EIF5A	Chr17:265693 Chr16:66160 (DLG4) Chr17:579446 (GPS2) Chr15:91429287	cg01697794 (DLG4)	7222668	Body	45.8 kb 58.1 kb	473	6.1 × 10 ⁻⁵	3.4 × 10 ⁻⁷	3.4 × 10 ⁻⁷	3.2 × 10 ⁻⁵	0.88
		cg19466160 (DLG4)	124181965	Body	45.8 kb 47.4 kb	473	3.7 × 10 ⁻⁶	3.9 × 10 ⁻⁷	3.9 × 10 ⁻⁷	0.88	
		cg17579446 (GPS2)	124228770	Body	45.8 kb 52.8 kb	396	4.8 × 10 ⁻¹³	1.1 × 10 ⁻⁷	1.9 × 10 ⁻⁷	3.2 × 10 ⁻⁵	
		cg04514024 (NEURL4)	124328787	Body	45.8 kb 58.1 kb	403	0.97	1.9 × 10 ⁻¹⁹	4.7 × 10 ⁻⁷	3.3 × 10 ⁻²⁰	

Chr. chromosome. Evidence of colocalization for a given scenario was based on the posterior probability of association (PPA) ≥ 0.8. For each locus indexed by the GWAS lead SNP, N SNPs denotes the number of SNPs within 1 Mb from the index SNP. This table presents loci associated with all three traits and have evidence of colocalization in ≥ 2 of the traits birthweight, placental DNA methylation (i.e. PPA ≥ 0.80 for any of the following four scenarios out of 15 Hypotheses tested by moloc): (1) PPA for G, E, M corresponding H10; association for GWAS trait (G), eQTL trait (E) and mQTL trait (M), but different causal variants for {G} and {E}, (2) PPA for GM, E corresponding H12; association for traits G, E and M, but different causal variants for {GM} and {E}, (3) PPA for G, E, M corresponding H11; association for traits G, E and M, but different causal variants for {G,E} and {M}, (4) PPA for GEM corresponding H14; SNP is associated with all 3 traits {G,E,M}. *CpG site has the promoter-associated regulatory feature. TSS denotes transcription start site. **PPA is very close to 0.8. In bold is PPA close to or ≥ 0.8.

and *HTRA1* expression levels were significantly inversely associated with birthweight in the RICHs dataset and the association strengthened in meta-analyses of the results from RICHs and NICHD Fetal Growth Studies. *HTRA1* expression showed suggestive inverse association with placental weight in the NICHD Fetal Growth Studies ($P=0.06$). Moreover, *PRMT7* and *CTDNEP1* expression levels showed directionally consistent inverse associations with birthweight in both datasets, although not reaching statistical significance (Supplementary Data 11). The inverse direction of association between these four genes and birthweight concurred with the eQTL findings such that the alleles with birthweight-lowering effect also had gene expression-increasing effect in the placenta. It is worth noting that gene expression levels are attributed to both genetic and environmental factors, and traits such as birthweight can in turn drive gene expression differences³⁶. As a result, the relationship between *cis*-regulated gene expression and a trait may not be accurately mirrored by the relationship between measured gene expression and a trait. More accurate inference of the portion of gene expression, partitioned into genetic and non-genetic components, causally associated with birthweight warrants further investigation in larger well-characterized cohorts.

Functional annotation, regulatory enrichment, and pathways. Enrichment analysis of the 778 DNAm sites from our significant mQTL-DNAm pairs was performed using eFORGE³⁷. We found significant enrichment in the placenta and depletion in the brain for DNase 1 hypersensitive sites (DHS), and enrichment in the placenta as well as other tissues for 15-state chromatin marks and H3 histone marks (Supplementary Data 12). The DNAm sites were enriched in CpG island shelf regions (FDR-adjusted $P < 0.05$) (Supplementary Data 13) and a striking majority (90%; 700/778) have been reported in epigenome-wide association studies (Supplementary Data 14).

Genes annotating the mQTL SNPs and their target DNAm sites ($n = 402$ genes) were significantly enriched (FDR-adjusted $P < 0.05$) in several ingenuity pathway analysis (IPA) canonical pathways related to immune response and endocrine signaling, in addition to metabolic and developmental pathways previously highlighted by GWAS of birthweight^{6,7} (Supplementary Data 15). Consistent with this finding, gene set enrichment analysis (GSEA) found significant enrichment of upregulated hallmark gene sets associated with low oxygen levels (oxidative stress), immune response, adipocyte differentiation (adipogenesis), and development of skeletal muscle (myogenesis) and pancreatic beta cells (Supplementary Data 16). Significant enrichment was also observed for differential gene expression in GTEx tissues (Bonferroni-adjusted $P < 0.05$), with the strongest enrichment for upregulation of the gene set detected in subcutaneous adipose tissues relative to other GTEx tissues and the strongest enrichment for downregulation of the gene set detected in the brain relative to other GTEx tissues (Supplemental Fig. S2).

Birthweight-associated SNPs found to be eQTL in the placenta were significantly enriched for eQTL in 30 GTEx tissues; birthweight-associated SNPs that were mQTL in placenta were significantly enriched for mQTL in blood samples from the ARIES mQTL database²⁸ (Supplemental Fig. S3a-c).

Discussion

Using integrated multi-omics approaches involving the placenta, an organ essential in fetal development and implicated in later life health, we identified placental DNAm and gene expression regulatory targets for birthweight GWAS loci. We found 23 loci in which the birthweight GWAS lead SNP was concurrently associated with nearby gene expression and DNAm in the placenta.

For these co-regulated molecular traits, two complementary causal inference approaches found evidence of causal relationships between DNAm and gene expression. Multitrait colocalization prioritized four candidate causal genes (*PLEKHA1*, *FES*, *PRMT7*, and *CTDNEP1*) in which variations in placental DNAm, gene expression and birthweight are attributed to a single causal genetic variant. Together, these data suggest that the effect of the GWAS variants on birthweight is possibly mediated via their direct regulatory influence on epigenetic and transcriptomic changes in the placenta. Moreover, our finding that there is enrichment of immune and hormonal pathways among mQTL target genes suggests that the GWAS loci may influence birthweight via these processes, commensurate with recently reported roles of the placenta in immune regulation of maternal–fetal cross-talks^{38–40}, and hormonal regulation of fetoplacental growth^{41–43}.

We observed that SNPs associated with birthweight were more likely to be *cis*-mQTLs than *cis*-eQTLs in the placenta (158/273; 57.9% vs. 26/273; 9.5%) and only 14.6% of *cis*-mQTLs were *cis*-eQTLs. Similarly, in a previous study, less than one-third of DNAm changes associated with birthweight have been correlated with nearby gene expression in the placenta⁴⁴. These observations may partly be explained by the diverse regulatory roles of DNAm beyond gene expression such as DNase I accessibility, chromatin accessibility, and histone modification²⁷. In contrast, the majority of *cis*-eQTLs were *cis*-mQTLs, which is much higher than the genome-wide overlaps between eQTLs and mQTLs in placenta²⁵, as well as other tissues^{26,27,45}. Therefore, the birthweight GWAS loci that regulate placental gene expression may be tightly linked with epigenetic mechanisms in the placenta either as a mediator or a passive consequence.

Multitrait colocalization analysis identified four genes (*PLEKHA1*, *FES*, *PRMT7*, and *CTDNEP1*) in which a genetic variant associated with birthweight colocalizes with variations in DNA methylation and gene expression in the placenta. This evidence suggests that the association between the GWAS SNPs and birthweight at these loci might involve placental DNA methylation and transcription of nearby genes, and transcription may be a mechanism of effect consistent with causality. *PLEKHA1* is a GWAS locus for type 2 diabetes, adult height, and age-related macular degeneration^{46–48}. Previously, evidence of colocalization has been found between *PLEKHA1* expression in the placenta and birthweight⁴⁹. Besides replicating this finding, the present study provided additional evidence for colocalization with DNAm in *PLEKHA1* and *HTRA1*. *HTRA1* has the highest gene expression in human placenta and regulates cell growth by controlling the availability of insulin-like growth factors and TGF- β signaling. Fetal and placental developmental abnormalities have been found in *htra1* knockout mice⁵⁰. We observed that genetic variants associated with higher placental *HTRA1* expression have a birthweight-lowering effect. Similarly, placental expression of *htra1* is significantly increased in preeclampsia, a pregnancy complication associated with low birthweight⁵¹. It has been suggested that *htra1* overexpression in the placenta may disrupt microtubule organization, attenuated cell motility, and endothelial dysfunction^{52,53}, placental trophoblast invasion-related features critical in fetal growth.

Other salient exemplars of colocalized loci identified in our study include protein arginine N-methyltransferase 7 (*PRMT7*) and eukaryotic translation initiation factor 5A (*EIF5A1*). *PRMT7* catalyzes arginine methylation, which is a post-translational modification that may play important physiological roles in DNA damage repair, transcriptional control, and establishment of the *Igf2/H19* imprinting control region methylation⁵⁴. B cell-specific *prmt7*-knockout mice develop impaired immune response⁵⁵ and *Prmt7* was downregulated in the placenta of mothers fed high fat diet⁵⁶.

EIF5A1 is crucial in the establishment of the placental–maternal circulation, placental development, and fetal–maternal immune tolerance^{57,58}.

Accumulating evidence shows considerable shared genetic overlap between fetal growth and cardiometabolic traits in later life^{6,59–61}. Polygenic variants in the fetal genome associated with higher glucose and blood pressure have a birthweight-lowering effect. In contrast, polygenic variants in the maternal genome associated with higher glucose have a birthweight-raising effect possibly due to increased cross-placental transfer of glucose that induces insulin secretion and promotes fetal growth^{59,62,63}. The birthweight-associated SNPs with colocalized effect on placental gene expression and methylation in our study have previously been associated with adult cardiometabolic phenotypes including blood pressure, type 2 diabetes, coronary artery disease, and height^{6,59}. The colocalized genes we identified are broadly expressed in several tissues, but our evaluation in blood-derived eQTL from GTEx and mQTL from ARIES found no statistically significant evidence of multitrait colocalization. We also did not find significant eQTL–birthweight colocalization for *PLEKHA1* and *CTDNEP1* in any GTEx tissue. It is possible that the variants influence birthweight via the effector genes acting primarily in placenta, whereas the same variants may influence cardiometabolic traits in later life via other mechanisms. Multiple tissues with mQTL and eQTL data derived from identical samples will be critical to confirm this.

Our study has limitations. First, the maternal genome can indirectly influence placental molecular traits by altering the intrauterine environment. Some genetic loci with effects that vary by parent-of-origin have profound influence on placental development and function^{64,65} and birthweight⁵⁹. Understanding the genetic regulation of fetal growth is faced with the challenging task in delineating maternal genetic effects from fetal genetic effects, and in determining whether fetal effects depend on the variant's parental origin. We did not consider suggested classifications of the GWAS SNPs via the fetal and maternal genome because cluster assignment remains “unclassified” for several birthweight GWAS loci, and unambiguous assignment of these classifications is a work in progress. For example, some SNPs previously classified to act only via the maternal genome have been found to have additional effect via the fetal genome in a recent larger study with fetal, maternal, and paternal genotypes that improved the accuracy of genotype phasing⁵⁹. Although it is reassuring that the colocalized loci we identified did not overlap with variants that have previously been found to influence birthweight via the maternal, but not fetal, genome^{6,59}, our study cannot distinguish direct fetal genetic effects from indirect maternal genetic effects. Second, given the heterogenous cellular architecture of the placenta⁶⁶, genetic effects that may be cell-type/state-specific need to be refined by future studies. Third, our analysis is likely underpowered to detect causality and colocalization at loci with weak eQTL or mQTL, warranting larger studies when placental datasets with multi-omics profiles become available. The size of our eQTL dataset is smaller than the mQTL dataset due to resource limitations, but is comparable to or larger than some tissue-specific datasets included in GTEx. The absence of significant difference in basic characteristics of the subset of study participants included in the eQTL analysis and those included in the mQTL analysis is reassuring that selection bias had minimal impact in our findings. Fourth, despite our study's strength that the candidate functional genes were identified through corroborating multi-omics evidence and replicated via mQTL in an independent cohort, definitive causal conclusions require future biological experiments. Last, the genetic variants associated with birthweight were discovered by studies involving largely European ancestry individuals; hence functional characterization of the genetic architecture of birthweight can be

improved when new loci are discovered via GWAS in non-European populations.

In conclusion, this study identified functional effector genes that underpin the genetic architecture of birthweight via placental epigenetic and transcriptomic mechanisms. Follow-up of the candidate causal genes identified in our study via experimental studies at the maternal–placental–fetal interface may unravel therapeutic targets to improve fetal growth outcomes and subsequent health in later life.

Methods

The NICHD Fetal Growth study protocol was approved by the institutional review boards of NICHD and each of the participating clinic sites, namely, Columbia University, New York; New York Hospital, Queens, New York; Christiana Care Health System, Delaware; Saint Peter's University Hospital, New Jersey; Medical University of South Carolina, South Carolina; University of Alabama, Alabama; Northwestern University, Illinois; Long Beach Memorial Medical Center, California; University of California, Irvine, California; Fountain Valley Hospital, California; Women and Infants Hospital of Rhode Island, Rhode Island; and Tufts University, Massachusetts.

Dataset. The discovery study included placenta samples obtained at delivery as part of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) Fetal Growth Studies—Singletons in which 2802 pregnant women in the United States were recruited between 8 and 13 gestational weeks and followed through delivery. Women with low risk for adverse pregnancy complications were recruited from four race/ethnic groups (i.e., non-Hispanic White, non-Hispanic Black, Hispanic, and Asian or Pacific Islander) from July 2009 to January 2013. To be enrolled, women had to have no past adverse pregnancy outcomes and no major pre-existing medical conditions including autoimmune diseases, chronic hypertension, diabetes, chronic renal disease, cancer, HIV/AIDS, or psychiatric disorders. Gestational age was determined using the date of the last menstrual period and confirmed by ultrasound between 8 weeks to 13 weeks and 6 days of gestation^{20,21}. Written informed consent was obtained from all study participants.

Placental biopsies measuring 0.5 cm × 0.5 cm × 0.5 cm were taken from the fetal side ($n = 312$) within 1 h of delivery and samples were placed in RNAlater and frozen -80°F for molecular analysis^{25,67}. DNA extracted from the placental biopsies was genotyped using HumanOmni2.5 Beadchip (Illumina Inc., San Diego, CA) and DNAm was profiled on the 312 samples using Illumina's Infinium Human Methylation450 Beadchip (Illumina Inc., San Diego, CA). We removed probes with mean detection $P \geq 0.05$, cross-reactive, non-autosomal, and CpG sites located within 20 base pair from known SNPs. Samples showing discrepancies between phenotypic sex and genotypic sex, outliers from the distribution of the samples' genetic clusters, and with a mismatching sample identifier were excluded. RNA from a subset of the placental samples ($n = 80$) was extracted using TRIZOL reagent (Invitrogen, MA), and RNA sequencing was performed using the Illumina HiSeq2000 system with 100 bp paired-end reads. The reads were mapped to the human reference genome (NCBI/build 37.2) using Tophat version 2.0.4. On the genotype dataset, we removed SNPs with $>5\%$ missing values, minor allele frequency $<0.5\%$, and not in Hardy–Weinberg equilibrium ($P < 10^{-4}$). SNP genotypes were imputed using the Michigan Imputation Server using the 1000 Genomes Phase 3 genotype reference⁶⁸, and filters were applied to remove insertion–deletions, SNPs with minor allele frequency $<0.5\%$ and SNPs with imputation dosage $r^2 < 0.38, 12, 25, 67$. The present analysis included samples that passed quality control, i.e., 291 samples with fetal genotype and DNAm data for mQTL analysis, of which 71 samples also had RNA-seq data for eQTL analysis.

Identification of placental cis-eQTL for GWAS variants associated with birthweight. SNPs known to be associated with birthweight at $P < 5 \times 10^{-8}$ (i.e., 286 SNPs consisting 190 lead SNPs with $P < 6.6 \times 10^{-9}$ and 96 SNPs with $6.6 \times 10^{-9} < P < 5 \times 10^{-8}$) were extracted from a GWAS of birthweight by Warrington et al.⁶ Out of the 286 SNPs, 273 (consisting 180 of the 190 lead SNPs and 93 of the 96 SNPs) were available in our dataset (Supplementary Table S2). A total of 7901 mRNA transcripts within 1 Mb from the 273 SNPs were included in cis-eQTL analysis. cis-eQTL analysis was performed with QTLtools software⁶⁹ using linear regression under an additive genetic model for each SNP, adjusted for fetal sex, self-reported race/ethnicity, top ten genotype principal components (PCs) to account for population structure, and top three RNA-seq PCs. The quantile–quantile (QQ) plot of P -values resulting from the cis-eQTL mapping of the 273 SNPs and all 7901 genes within 1-MB distance from the SNPs showed an absence of inflation ($\lambda = 0.93$) (Supplemental Fig. S4). Statistical significance was considered based on a false discovery rate (FDR) of 0.05.

Identification of placental cis-mQTL for GWAS variants associated with birthweight. A total of 104,053 DNAm sites within 1 Mb from the 273 SNPs were included in cis-mQTL analysis. cis-mQTL analysis was performed using linear

regression under an additive genetic model for each SNP, adjusted for fetal sex, ethnicity, methylation plate, top ten genotype PCs, and top three methylation PCs as implemented in QTLtools⁶⁹. Methylation PCs were included in the model as covariates to remove the hidden batch effects and other potential confounders in the DNAm data. The QQ plot of P -values resulting from the mQTL mapping of the 273 SNPs and all 104,053 CpG DNAm sites within 1-MB distance from the SNPs showed an absence of inflation ($\lambda = 0.99$) (Supplemental Fig. S5). Statistical significance was considered based on FDR of 0.05.

Causal inference test. For birthweight GWAS SNPs that were found to be both eQTL and mQTL in placenta, mediation analysis was performed to determine the direction of causal relationship of the SNP with DNAm and gene expression using the causal inference test (CIT v2.2)²². CIT is a mediation-based approach that performs a series of conditional regression tests to predict a causal relationship from SNP to exposure to outcome. The CIT outputs P -values of a causal model (SNP \rightarrow DNAm \rightarrow gene expression; pCausalCIT) and a reverse causal model (SNP \rightarrow gene expression \rightarrow DNAm; pRevCausalCIT). Predicted causal direction was assigned as SME when pCausalCIT <0.05 and pRevCausalCIT ≥ 0.05 , and as SEM when pCausalCIT ≥ 0.05 and pRevCausalCIT <0.05 . Triplets with pCausalCIT ≥ 0.05 and pRevCausalCIT ≥ 0.05 were considered independent and triplets with pCausalCIT <0.05 and pRevCausalCIT <0.05 were considered unclassified.

Mendelian Randomization (MR) Steiger test to validate the direction of causality between DNA methylation and gene expression. To distinguish the causal direction between DNAm and gene expression for mQTL–eQTL triplets in which a co-occurring SNP was associated with DNAm and gene expression, we performed MR Steiger directionality test²³. The MR Steiger test distinguishes the causal direction between exposure and outcome by testing whether the variance in the outcome explained by the instrumenting SNP is less than the variance in the exposure explained by the instrumenting SNP. Summary statistics from placental cis-mQTL and cis-eQTL were used as input for the MR Steiger test implemented in the TwoSampleMR package²⁹.

Multitrait colocalization. We used a Bayesian multiple-trait colocalization analysis implemented in the R package moloc¹⁷ to investigate whether the same causal variant influenced birthweight, placental methylation at a nearby DNAm site and gene expression at a nearby gene. Input data for moloc were GWAS summary data for “own birthweight” from the recent GWAS of birthweight⁶ (i.e., regression coefficients, their variances, and SNP minor allele frequencies in Europeans) for 23 SNPs with co-occurring cis-eQTL and cis-mQTL effects, placental eQTL data (spanning all genes in which the transcription start site is within 1 MB of a given SNP) and placental mQTL data (spanning all CpG DNAm sites within 1 MB of a given SNP) derived from our analysis. All 23 loci were analyzed because they had 50 or more SNPs with minor allele frequency $>5\%$ within their genomic region in common between all three datasets as recommended¹⁷. Given three traits were evaluated in our analysis, moloc computed 15 possible configurations of the association between genetic variation and traits. For each such locus, colocalization analysis was performed using default parameters and a posterior probability of association $\geq 80\%$ was considered evidence of colocalization as recommended¹⁷. Loci with evidence of colocalization suggest presence of a single causal variant that influences DNA methylation, gene expression, and birthweight.

Next, we assessed whether loci found to have evidence of genetic colocalization in our analysis also harbor shared causal variants for blood DNAm, blood gene expression, and birthweight. To do this, we applied moloc using summary data of blood mQTL at middle-age from the ARIES mQTL database²⁸, blood eQTL from the GTEx Portal¹⁵, and GWAS of birthweight⁶.

Assessment of the eQTL–mQTL pairs in an independent dataset. The 197 eQTL–mQTL pairs identified using the NICHD Fetal Growth Studies dataset were additionally assessed in the Rhode Island Child Health Study (RICHS) dataset. RICHS recruited mother and infant pairs from March 2009 to May 2013 following delivery at the Women and Infants Hospital of Rhode Island. RICHS selected infants born small for gestational age, large for gestational age and controls born appropriate for gestational age matched on sex, gestational age (± 3 days), and maternal age (± 2 years)²⁴. The study protocol was approved by the Institutional Review Boards of Brown University and Women and Infants Hospital of Rhode Island. Placental RNA-seq data from a subset of samples ($n = 200$) were obtained using the Illumina Hi-Seq 2500 platform, aligned to the human reference genome, and RNA transcript abundance was quantified using Salmon⁷⁰. About 20 million single-end RNA-seq reads were generated on each sample⁷¹. Placental DNAm data ($n = 220$) were obtained using the Infinium MethylationEPIC array (Illumina), preprocessed, and normalized with the minfi R package⁷². Genotype data were available for 159 infants obtained using the Illumina MegaEX array and imputed using the Haplotype Reference Consortium reference panel⁷³. eQTL and mQTL analyses were performed on the 159 samples with genotype, RNA-seq, and DNAm data.

Of the 23 SNPs associated with DNAm and gene expression in the 197 eQTL–mQTL pairs in our NICHD Fetal Growth Studies cohort, 22 SNPs were present in RICHS dataset (including rs7753582 used in place of rs7790713 which was

missing in the dataset [$r^2 = 1$ in 1000 Genomes EUR]). Of the 26 eGenes for *cis*-eQTL assessment, 20 eGenes were present in RICHs data after quality control and were *cis* with 18 SNPs. Of the 152 DNAm sites for *cis*-mQTL assessment, 122 DNAm sites were present in RICHs data after quality control, and 121 were *cis* with 22 SNPs. A total of 22 *cis*-eQTL tests were performed between the 18 SNPs and 20 eGenes, and a total of 123 *cis*-mQTL tests were performed between the 22 SNPs and 121 DNAm sites ($n = 159$) using QTLtools software⁶⁹. Analyses were adjusted for sex, self-reported ethnicity, methylation or RNA-seq batch, top ten genotype PCs, and top three expression PCs (for eQTL) or methylation PCs (for mQTL).

Functional annotation and regulatory enrichment. For the 778 unique DNAm sites from the significant mQTL-DNAm pairs, we assessed enrichment and depletion of overlap with tissue-specific or cell-type-specific regulatory features including DNase I hypersensitive sites (DHS), all 15-state chromatin marks, and all five H3 histone marks (i.e., H3K27me3, H3K4me1, H3K4me3, H3K36me3, H3K9me3) using eFORGE v2.0³⁷. The set of 778 DNAm sites was entered as input in eFORGE and tested separately for enrichment and depletion of overlap with each of the three putative functional elements (i.e., DHS, all 15-state chromatin marks, and all five H3 marks) compared to the respective data from consolidated ROADMAP epigenomics. The enrichment analysis was performed in one thousand matched background sets, with each background set consisting 778 DNAm sites matched for gene and CpG island annotation. Regulatory regions were also tested for enrichment in relation to their locations from CpG island genomic regions (island, shelves, shores, and open sea) based on functional information from the Encyclopedia of DNA Elements (ENCODE)⁷⁴ and Roadmap Epigenomics Project⁷⁵. Enrichment or depletion was considered statistically significant at $FDR < 0.05$.

Pathway analysis. Pathway analysis was performed using the Ingenuity Pathway Analysis tool (IPA, QIAGEN, Redwood City, CA, USA, www.qiagen.com/ingenuity). We used IPA for two sets of genes: (1) genes near the SNPs and CpG sites in the significant mQTL-DNAm pairs (a set of 402 unique genes), and (2) genes near the SNPs and the target genes in the significant eQTL-eGene pairs (a set of 50 unique genes). Enrichment in canonical pathways was assessed using the right-tailed Fisher's exact test. Pathways with three or more molecules enriched at $FDR < 0.05$ were considered statistically significant. Furthermore, the abovementioned two sets of genes were annotated in biological context using the GENE2FUNC option using FUMA, a web-based platform that facilitates functional annotation of GWAS results⁷⁶. FUMA provides hypergeometric tests of enrichment of a list of genes in 54 GTEx tissue-specific gene expression sets (GTEx v 8). Gene set enrichment analysis was performed using FUMA to test whether the input genes were enriched in hallmark gene sets, which are biological states displaying coordinated expression as defined by the Molecular Signatures Database (MsigDB v7.2)⁷⁷.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The genotypes, DNA methylation, and gene expression data used in this study from the NICHD Fetal Growth Studies have been deposited in the dbGaP database under accession code [phs001717.v1.p1](https://www.ncbi.nlm.nih.gov/bioproject/5000000). The genotypes and gene expression data from the Rhode Island Child Health Study (RICHs) have been deposited in the dbGaP database under accession code [phs001586.v1.p1](https://www.ncbi.nlm.nih.gov/bioproject/5000000). GWAS summary statistics for birthweight are available via the EGG website (<https://egg-consortium.org/>). The 1000 Genomes Reference Panel datasets are available at <https://www.internationalgenome.org/category/reference/>. The human genome reference is made accessible by the Genome Reference Consortium at https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/. GTEx eQTL data are accessible via the GTEx Portal at <https://gtexportal.org/home/eqtlDashboardPage>. ARIES mQTL database is available for download at <http://www.mqtlldb.org/>. Genotype imputation platform as well as the Haplotype Reference Consortium reference panel can be accessed via the Michigan Imputation Server at <https://imputationserver.readthedocs.io/en/latest/>. The catalog of epigenome-wide association studies can be accessed at <http://ewascatalog.org/>. Regulatory feature annotations of CpG sites are accessible from ENCODE and Roadmap Epigenomics projects at <https://www.encodeproject.org/> and <http://www.roadmapepigenomics.org/data/>. Functional mapping of SNPs and genes via FUMA is accessible at <https://fuma.ctglab.nl/>.

Received: 14 August 2021; Accepted: 11 April 2022;

Published online: 02 May 2022

References

1. Lawn, J. E. et al. Every Newborn: progress, priorities, and potential beyond survival. *Lancet* **384**, 189–205 (2014).

- Risnes, K. R. et al. Birthweight and mortality in adulthood: a systematic review and meta-analysis. *Int. J. Epidemiol.* **40**, 647–661 (2011).
- Harder, T., Rodekamp, E., Schellong, K., Dudenhausen, J. W. & Plagemann, A. Birth weight and subsequent risk of type 2 diabetes: a meta-analysis. *Am. J. Epidemiol.* **165**, 849–857 (2007).
- Hales, C. N. et al. Fetal and infant growth and impaired glucose tolerance at age 64. *BMJ* **303**, 1019–1022 (1991).
- Godfrey, K. M. & Barker, D. J. Fetal nutrition and adult disease. *Am. J. Clin. Nutr.* **71**, 1344S–1352SS (2000).
- Warrington, N. M. et al. Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat. Genet.* **51**, 804–814 (2019).
- Horikoshi, M. et al. Genome-wide associations for birth weight and correlations with adult disease. *Nature* **538**, 248–252 (2016).
- Tekola-Ayele, F. et al. Trans-ethnic meta-analysis of genome-wide association studies identifies maternal ITPR1 as a novel locus influencing fetal growth during sensitive periods in pregnancy. *PLoS Genet.* **16**, e1008747 (2020).
- Beaumont, R. N. 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. *Hum. Mol. Genet.* **27**, 742–756 (2018).
- Horikoshi, M. et al. New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nat. Genet.* **45**, 76–82 (2013).
- Freathy, R. M. et al. Variants in ADCY5 and near CCN1 are associated with fetal growth and birth weight. *Nat. Genet.* **42**, 430–435 (2010).
- Tekola-Ayele, F. et al. Admixture mapping identifies African and Amerindigenous local ancestry loci associated with fetal growth. *Hum. Genet.* **140**, 985–997 (2021).
- Hemberger, M., Hanna, C. W. & Dean, W. Mechanisms of early placental development in mouse and humans. *Nat. Rev. Genet.* **21**, 27–43 (2020).
- Gong, S. et al. The RNA landscape of the human placenta in health and disease. *Nat. Commun.* **12**, 2639 (2021).
- Consortium, G. T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
- Ongen, H. et al. Estimating the causal tissues for complex traits and diseases. *Nat. Genet.* **49**, 1676–1683 (2017).
- Giambartolomei, C. et al. A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics* **34**, 2538–2545 (2018).
- Tak, Y. G. & Farnham, P. J. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin* **8**, 57 (2015).
- Richardson, T. G. et al. Systematic Mendelian randomization framework elucidates hundreds of CpG sites which may mediate the influence of genetic variants on disease. *Hum. Mol. Genet.* **27**, 3293–3304 (2018).
- Grewal, J. et al. Cohort profile: NICHD fetal growth studies-singletons and twins. *Int. J. Epidemiol.* **47**, 25–251 (2018).
- Buck Louis, G. M. et al. Racial/ethnic standards for fetal growth: the NICHD fetal growth studies. *Am. J. Obstet. Gynecol.* **213**, 449 e1–449 e41 (2015).
- Millstein, J., Chen, G. K. & Breton, C. V. cit: hypothesis testing software for mediation analysis in genomic applications. *Bioinformatics* **32**, 2364–2365 (2016).
- Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081 (2017).
- Marsit, C. J., Maccani, M. A., Padbury, J. F. & Lester, B. M. Placental 11-beta hydroxysteroid dehydrogenase methylation is associated with newborn growth and a measure of neurobehavioral outcome. *PLoS ONE* **7**, e33794 (2012).
- Delahaye, F. et al. Genetic variants influence on the placenta regulatory landscape. *PLoS Genet.* **14**, e1007785 (2018).
- Gutierrez-Arcelus, M. et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *eLife* **2**, e00523 (2013).
- Banovich, N. E. et al. Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. *PLoS Genet.* **10**, e1004663 (2014).
- Gaunt, T. R. et al. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.* **17**, 61 (2016).
- Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenotype. *eLife* **7**, e34408 (2018).
- Taylor, D. L. et al. Integrative analysis of gene expression, DNA methylation, physiological traits, and genetic variation in human skeletal muscle. *Proc. Natl Acad. Sci. USA* **116**, 10883–10888 (2019).
- Aslanukov, A. et al. RanBP2 modulates Cox11 and hexokinase I activities and haploinsufficiency of RanBP2 causes deficits in glucose metabolism. *PLoS Genet.* **2**, e177 (2006).
- Tanaka, S. S. et al. Dullard/Ctdnep1 modulates WNT signalling activity for the formation of primordial germ cells in the mouse embryo. *PLoS ONE* **8**, e57428 (2013).

33. Greco, T. L. et al. Analysis of the vestigial tail mutation demonstrates that Wnt-3a gene dosage regulates mouse axial development. *Genes. Dev.* **10**, 313–324 (1996).
34. Zhang, Z. et al. Wnt/beta-catenin signaling pathway in trophoblasts and abnormal activation in preeclampsia (Review). *Mol. Med. Rep.* **16**, 1007–1013 (2017).
35. Panjwani, N. et al. LocusFocus: Web-based colocalization for the annotation and functional follow-up of GWAS. *PLoS Comput. Biol.* **16**, e1008336 (2020).
36. Porcu, E. et al. Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome. *Nat. Commun.* **12**, 5647 (2021).
37. Breeze, C. E. et al. eFORGE: a tool for identifying cell type-specific signal in epigenomic data. *Cell Rep.* **17**, 2137–2150 (2016).
38. Schumacher, A., Sharkey, D. J., Robertson, S. A. & Zenclussen, A. C. Immune cells at the fetomaternal interface: how the microenvironment modulates immune cells to foster fetal development. *J. Immunol.* **201**, 325–334 (2018).
39. PrabhuDas, M. et al. Immune mechanisms at the maternal-fetal interface: perspectives and challenges. *Nat. Immunol.* **16**, 328–334 (2015).
40. Colucci, F. The immunological code of pregnancy. *Science* **365**, 862–863 (2019).
41. Yamasato, K., Tsai, P. S., Davis, J., Yamamoto, S. Y. & Bryant-Greenwood, G. D. Human relaxins (RLNH1, RLNH2), their receptor (RXFP1) and fetoplacental growth. *Reproduction* **154**, 67–77 (2017).
42. Paradis, A. & Zhang, L. Role of endothelin in uteroplacental circulation and fetal vascular function. *Curr. Vasc. Pharmacol.* **11**, 594–605 (2013).
43. Barut, F. et al. Intrauterine growth restriction and placental angiogenesis. *Diagn. Pathol.* **5**, 24 (2010).
44. Tekola-Ayele, F. et al. DNA methylation loci in placenta associated with birthweight and expression of genes relevant for early development and adult diseases. *Clin. Epigenetics* **12**, 78 (2020).
45. Bell, J. T. et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* **12**, R10 (2011).
46. Replication, D. I. G. et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46**, 234–244 (2014).
47. Okada, Y. et al. A genome-wide association study in 19 633 Japanese subjects identified LHX3-QSOX2 and IGF1 as adult height loci. *Hum. Mol. Genet.* **19**, 2303–2312 (2010).
48. Chen, W. et al. Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration. *Proc. Natl Acad. Sci. USA* **107**, 7401–7406 (2010).
49. Peng, S. et al. Genetic regulation of the placental transcriptome underlies birth weight and risk of childhood obesity. *PLoS Genet.* **14**, e1007799 (2018).
50. Hasan, M. Z., Ikawati, M., Tocharus, J., Kawaichi, M. & Oka, C. Abnormal development of placenta in Htra1-deficient mice. *Dev. Biol.* **397**, 89–102 (2015).
51. Ajayi, F. et al. Elevated expression of serine protease Htra1 in preeclampsia and its role in trophoblast cell migration and invasion. *Am. J. Obstet. Gynecol.* **199**, 557 e1–10 (2008).
52. Chien, J. et al. Serine protease Htra1 associates with microtubules and inhibits cell migration. *Mol. Cell Biol.* **29**, 4177–4187 (2009).
53. Zhang, L. et al. High temperature requirement factor A1 (HTRA1) gene regulates angiogenesis through transforming growth factor-beta family member growth differentiation factor 6. *J. Biol. Chem.* **287**, 1520–1526 (2012).
54. Bedford, M. T. & Clarke, S. G. Protein arginine methylation in mammals: who, what, and why. *Mol. Cell* **33**, 1–13 (2009).
55. Ying, Z. et al. Histone arginine methylation by PRMT7 controls germinal center formation via regulating Bcl6 transcription. *J. Immunol.* **195**, 1538–1547 (2015).
56. Gabory, A. et al. Maternal diets trigger sex-specific divergent trajectories of gene expression and epigenetic systems in mouse placenta. *PLoS ONE* **7**, e47986 (2012).
57. Qin, X. et al. Inhibition of eIF5A results in aberrant uterine natural killer cell function and embryo loss in mice. *Am. J. Reprod. Immunol.* **71**, 229–240 (2014).
58. Qin, X. et al. Eukaryotic initiation factor 5A and Ca(2+) /calmodulin-dependent protein kinase 1D modulate trophoblast cell function. *Am. J. Reprod. Immunol.* **80**, e12845 (2018).
59. Juliusdottir, T. et al. Distinction between the effects of parental and fetal genomes on fetal growth. *Nat. Genet.* **53**, 1135–1142 (2021).
60. Tekola-Ayele, F. et al. Genetic overlap between birthweight and adult cardiometabolic diseases has implications for genomic medicine. *Sci. Rep.* **9**, 4076 (2019).
61. Chatterjee, S., Ouidir, M. & Tekola-Ayele, F. Pleiotropic genetic influence on birth weight and childhood obesity. *Sci. Rep.* **11**, 48 (2021).
62. Rahman, M. L. et al. Maternal and offspring genetic risk of type 2 diabetes and offspring birthweight among African ancestry populations. *J. Clin. Endocrinol. Metab.* **104**, 5032–5042 (2019).
63. Sosenko, I. R. et al. The infant of the diabetic mother: correlation of increased cord C-peptide levels with macrosomia and hypoglycemia. *N. Engl. J. Med.* **301**, 859–862 (1979).
64. Hanna, C. W. Placental imprinting: Emerging mechanisms and functions. *PLoS Genet.* **16**, e1008709 (2020).
65. Coan, P. M., Burton, G. J. & Ferguson-Smith, A. C. Imprinted genes in the placenta—a review. *Placenta* **26**(Suppl A), S10–S20 (2005).
66. Tsang, J. C. H. et al. Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. *Proc. Natl Acad. Sci. USA* **114**, E7786–E7795 (2017).
67. Tekola-Ayele, F. et al. Sex differences in the associations of placental epigenetic aging with fetal growth. *Aging* **11**, 5412–5432 (2019).
68. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
69. Delaneau, O. et al. A complete tool set for molecular QTL discovery and analysis. *Nat. Commun.* **8**, 15452 (2017).
70. Love, M. I., Hogenesch, J. B. & Irizarry, R. A. Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. *Nat. Biotechnol.* **34**, 1287–1291 (2016).
71. Peng, S. et al. Expression quantitative trait loci (eQTLs) in human placentas suggest developmental origins of complex diseases. *Hum. Mol. Genet.* **26**, 3432–3441 (2017).
72. Fortin, J. P., Triche, T. J. Jr. & Hansen, K. D. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* **33**, 558–560 (2017).
73. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
74. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
75. Roadmap Epigenomics, C. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
76. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
77. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

Acknowledgements

This work was supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH) including American Recovery and Reinvestment Act funding via contract numbers HHSN275200800013C; HHSN275200800002I; HHSN27500006; HHSN275200800003IC; HHSN275200800014C; HHSN275200800012C; HHSN275200800028C; HHSN275201000009C and HHSN27500008. Additional support was obtained from the NIH Office of the Director, the National Institute on Minority Health and Health Disparities (NIMHD) and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). RICHS is partially supported by NIH-NIEHS R01ES022223 (CJM, JC and KH), NIH-NIEHS R01ES022223-03S1 (CJM), and NIH-NIEHS R24ES028507 (CJM). CL is funded through NIH-NICHD R00HD097286. The authors acknowledge the research teams at all participating clinical centers for the NICHD Fetal Growth Studies, including Christina Care Health Systems, Columbia University, Fountain Valley Hospital, California, Long Beach Memorial Medical Center, New York Hospital, Queens, Northwestern University, University of Alabama at Birmingham, University of California, Irvine, Medical University of South Carolina, Saint Peters University Hospital, Tufts University, and Women and Infants Hospital of Rhode Island. The authors also acknowledge C-TASC and The EMMES Corporations in providing data and imaging support. Genotyping was performed in the Department of Laboratory Medicine and Pathology, University of Minnesota. This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>). The authors are also thankful to the RICHS study participants for their participation, and the study staff at Women and Infants Hospital for their dedication to the project. Data on birthweight GWAS summary statistics have been contributed by the EGG Consortium using the UK Biobank Resource and has been downloaded from www.egg-consortium.org.

Author contributions

F.T.-A. conceived and designed this study and wrote the draft manuscript. X.Z., S.C., and M.O. performed statistical analyses and visualizations. C.L. analyzed the replication dataset M.T. reviewed functional annotations. T.W. contributed to data quality control. K.H. and J.C. contributed the data. C.J.M. and R.W. contributed data and samples. F.T.-A., X.Z., S.C., M.O., C.L., K.H., J.C., M.T., C.J.M., T.W., and R.W. interpreted the results, reviewed the draft manuscript, provided critical intellectual content, and approved the final manuscript.

Funding

Open Access funding provided by the National Institutes of Health (NIH).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-30007-1>.

Correspondence and requests for materials should be addressed to Fasil Tekola-Ayele.

Peer review information *Nature Communications* thanks Janine LaSalle and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2022