Genetic control of fetal placental genomics contributes to development of health and disease

Arjun Bhattacharya[1,2]*, Anastasia N. Freedman[3], Vennela Avula[3], Rebeca Harris[4], Weifang Liu[5], Calvin Pan[6], Aldons J. Lusis[6,7,8], Robert M. Joseph[9], Lisa Smeester[3,10,11], Hadley J. Hartwell[3], Karl C.K. Kuban[12], Carmen J. Marsit[13], Yun Li[5,14,15],T. Michael O'Shea[16], Rebecca C. Fry[3,10,11]*†, Hudson P. Santos Jr[4,10]*†

†These authors contributed equally to this manuscript.

*Correspondence to Arjun Bhattacharya (abtbhatt@ucla.edu), Rebecca C. Fry (rfry@email.unc.edu), or Hudson P. Santos, Jr (hsantosj@email.unc.edu)

[1]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, 90095, USA
[2]Institute for Quantitative and Computational Biosciences, David Geffen School of Medicine, University of California, Los Angeles, CA, 90095, USA
[3]Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, 27514, USA
[4]Biobehavioral Laboratory, School of Nursing, University of North Carolina, Chapel Hill, NC, 27514, USA
[5]Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, 27514, USA
[6]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, 90095, USA
[7]Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, 90095, USA
[8]Department of Microbiology, Immunology and Molecular Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, 90095, USA
[9]Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, MA, 02118, USA
[10]Institute for Environmental Health Solutions, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, 27514, USA
[11]Curriculum in Toxicology and Environmental Medicine, University of North Carolina, Chapel Hill, NC, 27514, USA
[12]Department of Pediatrics, Division of Pediatric Neurology, Boston University Medical Center, Boston, MA, 02118, USA
[13]Gangarosa Department of Environmental Health, Rollins School of Public Health Emory University, Atlanta, GA, 30322, USA
[14]Department of Genetics, University of North Carolina, Chapel Hill, NC, 27514, USA
[15]Department of Computer Science, University of North Carolina, Chapel Hill, NC, 27514, USA
[16]Department of Pediatrics, School of Medicine, University of North Carolina, Chapel Hill, NC, 27514, USA

ABSTRACT

As the master regulator *in utero*, the placenta is core to the Developmental Origins of Health and Disease (DOHaD) hypothesis but is historically understudied. To identify placental gene-trait associations (GTAs) across the life course, we performed distal mediator-enriched transcriptome-wide association studies (TWAS) for 40 traits, integrating placental multi-omics from the Extremely Low Gestational Age Newborn Study. At $P < 2.5 \times 10^{-6}$, we detected 248 GTAs, mostly for neonatal and metabolic traits, across 176 genes, enriched for cell growth and immunological pathways. In aggregate, genetic effects mediated by placental expression significantly explained 4 early-life traits but no later-in-life traits. 89 GTAs showed significant mediation through distal genetic variants, identifying hypotheses for distal regulation of GTAs. Investigation of one hypothesis in human placenta-derived choriocarcinoma cells showed that knockdown of mediator gene *EPS15* upregulated predicted targets *SPATA13* and *FAM214A*, both associated with waist-hip ratio in TWAS, and multiple genes involved in metabolic pathways. These results suggest profound health impacts of placental genomic regulation in developmental programming across the life course.

KEYWORDS

placental biology, in utero development, developmental origins of health and disease, developmental programming, multi-omics integration, transcriptome-wide association study, *trans*-regulation

INTRODUCTION

The placenta serves as the master regulator of the intrauterine environment via nutrient transfer, metabolism, gas exchange, neuroendocrine signaling, growth hormone production, and immunologic surveillance[1–3]. Due to strong influences on postnatal health, the placenta is central to the Developmental Origins of Health and Disease (DOHaD) hypothesis – that the *in utero* experience has lifelong impacts on child health by altering developmental programming and influencing risk of common, noncommunicable health conditions[4]. For example, physiological characteristics of the placenta have been linked to neuropsychiatric, developmental, and metabolic diseases or health traits (collectively referred to as traits) that manifest throughout the life course, either early- or later-in-life (**Figure 1**)[1,5–8]. Despite its long-lasting

influences on health, the placenta is understudied in large consortia studies of multi-tissue gene regulation[9,10]. Studying regulatory mechanisms in the placenta underlying biological processes in developmental programming could provide novel insight into health and disease etiology.

The complex interplay between genetics and placental transcriptomics and epigenomics has strong effects on gene expression that may explain variation in gene-trait associations (GTAs). Quantitative trait loci (QTL) analyses have identified a strong influence of *cis*-genetic variants on both placental gene expression and DNA methylation[11]. Furthermore, there is growing evidence that the placental epigenome influences gene regulation, often distally (more than 1-3 Megabases away in the genome)[12], and that placental DNA methylation and microRNA (miRNA) expression are associated with health traits in children[13]. Dysfunction of transcription factor regulation in the placenta has also shown profound effects on childhood traits[14]. Although combining genetics, transcriptomics, and epigenomics lends insight into the influence of placental genomics on complex traits[15], genome-wide screens for GTAs that integrate different molecular profiles and generate functional hypotheses require more sophisticated computational methods.

To this end, advances in transcriptome-wide association studies (TWAS) have allowed for integration of genome-wide association studies (GWAS) and eQTL datasets to boost power in identifying GTAs, specific to a relevant tissue[16,17]. However, traditional methods for TWAS largely overlook genetic variants distal to genes of interest, ostensibly mediated through regulatory biomarkers, like transcription factors, miRNAs, or DNA methylation sites[18]. Not only may these distal biomarkers explain a significant portion of both gene expression heritability and trait heritability on the tissue-specific expression level[19,20], they may also influence tissue-specific trait associations for individual genes. Due to the strong interplay of regulatory elements in placental gene regulation, we sought to systematically characterize portions of gene expression that are influenced by these distal regulatory elements.

Here, we set out to identify the following: (1) which genes show associations between their placental genetically-regulated expression (GReX) and various traits across the life course, (2) which traits along

the life course can be explained by placental GReX, in aggregate, and (3) which transcription factors, miRNAs, or CpG sites potentially regulate trait-associated genes in the placenta (**Figure 1**). We leveraged multi-omic data from fetal-side placenta tissue from the Extremely Low Gestational Age Newborn (ELGAN) Cohort Study[21] to train predictive models of gene expression enriched for distal SNPs using MOSTWAS, a recent TWAS extension that integrates multi-omic data[22]. Using 40 GWAS of European-ancestry subjects from large consortia[23–27], we performed a series of TWAS for non-communicable health traits and disorders that may be influenced by the placenta to identify GTAs and functional hypotheses for regulation (**Figure 2**). To our knowledge, this is the first distal mediator-enriched TWAS of health traits that integrates multi-omic data from the placenta.

RESULTS

Overview of analytic framework

We conduct a series of distal mediator-enriched transcriptome-wide association studies (TWAS) for a variety of complex traits by integrating GWAS data with placental eQTL data from ELGAN. First, we use a recent methodology, MOSTWAS[22], to train predictive models of gene expression using both local- and distal-SNPs to genes (**Figure 2A**). Next, we employ these models to conduct TWAS for these traits using GWAS summary statistics to identify genes with placental genetically-regulated expression (GReX) associated with different traits across the life course (**Figure 2B**)[17]. We then estimate the extent to which placental genetically-regulated expression across all trait-associated genes can explain the variability in a trait and correlations between traits (**Figure 2C**)[17,28]. Next, to provide more biological context, for genes estimated to have placental GTAs, we run multiple follow-up analyses (**Figure 2C**): gene ontology enrichment analyses[29], probabilistic fine-mapping of overlapping loci[30], phenome-wide analyses for select genes, and prioritization of functional hypotheses for upstream distal regulation[22]. Lastly, for one particular functional hypothesis with strong computational support, we conduct an in-vitro assay in human placenta-derived cell lines to validate the predicted mediator-TWAS gene relationship and the transcriptomic consequences of this mediator (**Figure 2D**).

Complex traits are genetically heritable and correlated

We curated GWAS summary statistics from subjects of European ancestry for 40 non-communicable traits and disorders across five health categories to identify potential links to genetically-regulated placental expression (traits and cohorts for each GWAS are summarized in **Supplemental Table S1**, sample sizes are provided in **Supplemental Table S2**). These five categories of traits (autoimmune/autoreactive disorders, metabolic traits, cardiovascular disorders, early childhood outcomes, and neuropsychiatric traits) have been linked previously to placental and fetal biology and morphology[1–8].These 40 traits, derived from 5 different consortia (**Supplemental Table S1**), comprise of 3 autoimmune/autoreactive disorders, 8 body size/metabolic traits, 4 cardiovascular disorders, 14 neonatal/early childhood traits, and 11 neuropsychiatric traits/disorders[23–27]. The 26 traits that are not categorized as neonatal/early childhood traits are measured exclusively in adults. In addition, these 40 GWAS are not derived from the same samples of patients.

To quantify the total genetic contribution to each trait and the genetic associations shared between traits, using linkage disequilibrium (LD) score regression with LD scores generated for individuals of European ancestry from the 1000 Genomes projects[31,32], we estimated the SNP heritability (h^2) and genetic correlation (r_g) of these traits, respectively (**Supplemental Figure S1 and S2**). Of the 40 traits, 37 showed significantly positive SNP heritability and 18 with $\hat{h}^2 > 0.10$ (**Supplemental Figure S1, Supplemental Table S1**), with the largest heritability for childhood BMI ($\hat{h}^2 = 0.69, SE = 0.064$). As expected, we observed strong, statistically significant genetic correlations between traits of similar categories (i.e., between neuropsychiatric traits or between metabolic traits) (**Supplemental Figure S2**; **Supplemental Table S3**). At Benjamini-Hochberg FDR-adjusted $P < 0.05$, we also observed significant correlations between traits from different categories: diabetes and angina ($\hat{r}_g = 0.51$, FDR-adjusted $P = 6.53 \times 10^{-33}$), Tanner scale (in children) and BMI ($\hat{r}_g = 0.42$, FDR-adjusted $P = 1.06 \times 10^{-3}$), and BMI and obsessive compulsive disorder ($\hat{r}_g = -0.28$, FDR-adjusted $P = 1.79 \times 10^{-9}$), for example. Given strong and potentially shared genetic influences across these traits, we examined whether genetic associations with these traits are mediated by the placental transcriptome.

Multiple placental gene-trait associations detected across the life course

In the first step of our TWAS (**Figure 2A**), we leveraged MOSTWAS[22], a recent TWAS extension that includes distal variants in transcriptomic prediction, to train predictive models of placental expression. As large proportions of total heritable gene expression are explained by distal-eQTLs local to regulatory hotspots[18,20], MOSTWAS uses data-driven approaches to identify mediating regulatory biomarkers or distal-eQTLs mediated through local regulatory biomarkers to increase predictive power for gene expression and power to detect GTAs (**Supplemental Figure S3**)[22]. In this analysis, these regulatory biomarkers include potential regulatory protein (RP) encoding genes (as curated by TFcheckpoint[33]), miRNAs, and CpG methylation sites from the ELGAN Study. we assume that these RP genes, miRNAs, and genes and other regulatory features local to these CpG methylation sites have distal effects on the transcription of genes of interest and thus potentially mediate distal-eQTLs to the gene of interest (**Methods**).

Using genotypes from umbilical cord blood[34] and mRNA expression, CpG methylation, and miRNA expression data from fetal-side placenta[15] from the ELGAN Study[21] for 272 infants born pre-term, we built genetic models to predict RNA expression levels for genes in the fetal placenta (demographic summary in **Supplemental Table S4**). Out of a total of 12,020 genes expressed across all samples in ELGAN, we successfully built significant models for 2,994 genes, with positive SNP-based expression heritability (nominal $P < 0.05$) and five-fold McNemar's adjusted cross-validation (CV) $R^2 \geq 0.01$ (**Figure 3A [Step 1]; Methods**). Only these 2,994 models are used in subsequent TWAS steps. Mean SNP heritability for these genes was 0.39 (25% quantile = 0.253, 75% quantile = 0.511), and mean CV R^2 was 0.031 (quantiles: 0.014, 0.034). For out-sample validation, we imputed expression into individual-level genotypes from the Rhode Island Child Health Study (RICHS; $N = 149$)[35,36], showing strong portability across studies: of 2,005 genes with RNA-seq expression in RICHS, 1,131 genes met adjusted $R^2 \geq 0.01$, with mean $R^2 = 0.011$ (quantiles: 7.71×10^{-4}, 0.016) (**Figure 3B; Supplemental Table S5**). Summary statistics of demographic and clinical variables for the RICHS show similar distributions of race, though RICHS excluded all pre-term babies, a clear difference in these two cohorts (**Supplemental Table S4**).

We integrated GWAS summary statistics for 40 traits from European-ancestry subjects with placental gene expression using our predictive models. Using the weighted burden test with the 1000Genomes European ancestry LD matrix as a reference[17], we detected 932 GTAs (spanning 686 unique genes) at $P < 2.5 \times 10^{-6}$, a transcriptome-wide significance threshold consistent with previous TWAS[17,28] (**Figure 3A [Step 2]**). As many of these loci carry significant signal because of strong SNP-trait associations, we employed Gusev *et al*'s permutation test to assess how much signal is added by the SNP-expression weights and confidently conclude that integration of expression data significantly refines association with the trait[17]. At FDR-adjusted $P < 0.05$ and spanning 176 unique genes, we detected 248 such GTAs, with 11 autoimmune/autoreactive, 136 body size/metabolic, 32 cardiovascular, 39 neonatal/childhood, and 30 neuropsychiatric GTAs (**Figure 3A [Step 3]**, **Supplemental Table S2** and **S6;** Miami plots of TWAS Z-scores in **Supplemental Figures S4-S9**).

The 39 GTAs detected with adult BMI included *LARS2* ($Z = 11.4$) and *CAST* ($Z = -4.61$). These two GTAs have been detected using *cis*-only TWAS in different tissues[17,28]. In addition, one of the 30 genes identified in association with waist-hip ratio (in adults) was prioritized in other tissues by TWAS: *NDUFS1* ($Z = -5.38$)[28]. We cross-referenced susceptibility genes with a recent *cis*-only TWAS of fetal birthweight, childhood obesity, and childhood BMI by Peng *et al* using placental expression data from RICHS[8]. Of the 19 birthweight-associated genes they identified, we could only train significant expression models for two in ELGAN: *PLEKHA1* and *PSG8*. We only detected a significant association between *PSG8* and fetal birthweight ($Z = -7.77$). Similarly, of the 6 childhood BMI-associated genes identified by Peng *et al*, only 1 had a significant model in ELGAN and showed no association with the trait; there were no overlaps with childhood obesity-associated genes[8]. We hypothesize that minimal overlap with susceptibility genes identified by Peng *et al* is due to differing phenotypes and eQTL architectures in the datasets and different inclusion criteria for significant gene expression models.

Next, we tested for horizontal pleiotropic effects of the SNPs employed in the models for TWAS-prioritized genes; if SNPs affect the outcome through a pathway independent of expression of the gene, the TWAS association may be biased[37,38]. Here, using PMR-Summary-Egger[38], we test the magnitude of this null

hypothesis for each of the 248 TWAS-prioritized GTAs. At FDR-adjusted $P < 0.05$, only three GTAs showed significant horizontal pleiotropic effects: *MOV10*, *SLC35G2*, and *HLA-A*, all associated with adult waist-hip ratio (**Supplemental Table S6**). These three genes may have upwardly biased TWAS associations, as the SNPs used to construct their GReX may influence the outcome through a different molecular pathway.

As these GTAs indicate trait association and do not reflect causality, we used FOCUS[30], a Bayesian fine-mapping approach. For TWAS-significant genes with overlapping genetic loci, FOCUS estimates posterior inclusion probabilities (PIP) in a credible set of genes that explains the association signal at the locus. We found 8 such overlaps and estimated a 90% credible set of genes explaining the signal for each locus (**Supplemental Table S9**). For example, we identified 3 genes associated with triglycerides in adults at the 12q24.13 chromosomal region (*ERP29*, *RPL6*, *BRAP*), with *ERP29* defining the region's 90% credible set with approximately 95% PIP. Similarly, we detected 3 genes associated with adultBMI at 10q22.2 (*AP3M1*, *SAMD8*, *MRPS16*), with *AP3M1* defining the region's 90% credible set with approximately 99% PIP.

We conducted over-representation analysis for biological process, molecular function, and PANTHER gene pathway ontologies for TWAS-detected susceptibility genes (**Figure 3D**, **Supplemental Table S7**)[29]. Overall, considering all 176 TWAS-identified genes, we observed enrichments for nucleic acid binding and immune or cell growth signaling pathways (e.g., B-cell/T-cell activation and EGF receptor, interleukin, PDGF, and Ras signaling pathways). By trait, we found related pathways (sphingolipid biosynthesis, cell motility, etc) for TWAS genes for metabolic and morphological traits (e.g., BMI and childhood BMI); for most traits, we were underpowered to detect ontology enrichments. We also assessed the overlap of TWAS genes with GWAS signals. A total of 112 TWAS genes did not overlap with GWAS loci ($P < 5 \times 10^{-8}$) within a 500 kilobase interval around any SNPs (local and distal) included in predictive models (**Table 1**).

Genetically-regulated placental expression mediates trait heritability and genetic correlations

To assess how genetically-regulated placental expression explains trait variance, we computed trait heritability on the placental expression level (h^2_{GE}) using all examined and all TWAS-prioritized susceptibility genes using a linkage disequilibrium (LD) score regression approach[17,31]. Overall, we found 4/14 neonatal traits (childhood BMI, head circumference, total puberty growth, and pubertal growth start) with significant $\hat{h}^2_{GE} > 0$ (FDR-adjusted $P < 0.05$ for jack-knife test of significance)[28]; none of the 26 traits outside the neonatal category were appreciably explained by placental GReX (**Supplemental Figure S10**). **Figure 4A** shows that mean $\hat{h}^2_{GE}$ is higher in neonatal traits than other groups. In fact, placenta expression-mediated genetic heritability explains a larger proportion of total SNP heritability of neonatal traits, compared to traits from other categories (**Figure 4B**). A comparison of the number of GWAS-significant SNPs and TWAS-significant genes also shows that neonatal traits are enriched for placental TWAS associations, even though significant genome-wide GWAS architecture cannot be inferred for these traits (**Supplemental Figure S11**). These observations suggest that placental GReX affects neonatal traits more profoundly, as a significantly larger proportion of neonatal traits showed significant heritability on the placental GReX level than later-in-life traits.

Using RHOGE[28], we assessed genetic correlations (r_{GE}) between traits at the level of placental GReX (**Supplemental Figure S12**). We found several known correlations: between cholesterol and triglycerides, both in adults, ($\hat{r}_{GE} = 0.99, P = 1.44 \times 10^{-118}$) and childhood BMI and adult BMI ($\hat{r}_{GE} = 0.55, P = 3.67 \times 10^{-8}$). Interestingly, we found correlations between traits across categories (**Figure 4C**): IQ and diastolic blood pressure, both in adults,($\hat{\rho}_{GE} = -0.55, P = 2.44 \times 10^{-5}$) and age of asthma diagnosis and adult glucose levels ($\hat{\rho}_{GE} = 0.86, P = 3.05 \times 10^{-6}$). These traits have been linked in morphological analyses of the placenta, but our results suggest possible genomic contributions[39]. Overall, these correlations suggest shared genetic pathways for these pairs of traits or for etiologic antecedents of these traits; these shared pathways could be either at the susceptibility genes or through shared distal loci, mediated by RPs, miRNAs, or CpG methylation sites.

Genes with multiple GTAs have phenome-wide associations in early- and later-life traits

We noticed that multiple genes were identified in GTAs with multiple traits, leading us to examine potential horizontally pleiotropic genes. Of the 176 TWAS-prioritized genes, we identified 50 genes associated with multiple traits, many of which are genetically correlated (**Table 2**). Nine genes showed more than 3 GTAs across different categories. For example, *IDI1*, a gene involved in cholesterol biosynthesis[40], showed associations with 3 metabolic and 2 neuropsychiatric traits: body fat percentage ($Z = 15.57$), HDL ($Z = 26.48$), triglycerides ($Z = -7.53$), fluid intelligence score ($Z = 6.37$), and schizophrenia ($Z = -5.56$), with all five traits measured in adults. A link between cholesterol-related genes and schizophrenia has been detected previously, potentially due to coregulation of myelin-related genes[41]. Mediated by CpG site cg01687878 (found within *PITPNM2*), predicted expression of *IDI1* was also computed using distal SNPs within Chromosome 12q24.31, a known GWAS risk loci for hypercholesteremia[42]; the inclusion of this locus may have contributed to the large TWAS associations. Similarly, *SAMD4A* also shows associations with 4 adult body size/metabolic - body fat percentage ($Z = 6.70$), cholesterol ($Z = -6.76$), HDL ($Z = -6.78$), triglycerides ($Z = -5.30$) - and 1 adult cardiovascular trait (diastolic blood pressure with $Z = -5.29$). These associations also pick up on variants in Chromosome 12q24.31 local to CpG sites cg05747134 (within *MMS19*) and cg04523690 (within *SETD1B*). Another gene with multiple trait associations is *CMTM4*, an angiogenesis regulator[43], showing associations with body fat percentage ($Z = 6.17$), hypertension ($Z = 5.24$), and fetal birthweight ($Z = 8.11$). *CMTM4* shows evidenced risk of intrauterine growth restriction due to involvement with endothelial vascularization[44], potentially suggesting that *CMTM4* has a more direct effect *in utero*, which mediates its associations with body fat percentage and hypertension.

We further studied the 9 genes with 3 or more distinct GTAs across different categories (**Figure 5A**). Using UK Biobank[23] GWAS summary statistics, we conducted TWAS for a variety of traits, measured in adults, across 8 groups, defined generally around ICD code blocks (**Figure 5A, Supplemental Figure S13**); here, we grouped metabolic and cardiovascular traits into one category for ease of analysis. At FDR-adjusted $P < 0.05$, *ATPAF2*, *RPL6*, and *SEC11A* showed GTA enrichments for immune-related traits, *ATAPF2* for neonatal traits, *IDI1* for mental disorders, and *RPS25* for musculoskeletal traits. Across these 8 trait groups, *RPL6* showed multiple strong associations with circulatory, respiratory, immune-

related, and neonatal traits (**Figure 5A**). Examining specific GTAs for *ATPAF2*, *IDI1*, *RPS25*, and *SEC11A* reveals associations with multiple biomarker traits (**Supplemental Figure S13**). For example, at $P < 2.5 \times 10^{-6}$, *ATPAF2* and *IDI1*'s immune GTA enrichment includes associations with eosinophil, monocyte, and lymphocyte count and IGF-1 concentration. *ATPAF* and *RPS25* show multiple associations with platelet volume and distribution and hematocrit percentage. In addition, *IDI1* was associated with multiple mental disorders (obsessive compulsive disorder, anorexia nervosa, bipolar disorder, and general mood disorders), consistent with its TWAS associations with fluid intelligence and schizophrenia (**Supplemental Figure S13**). As placental GReX of these genes correlates with biomarkers, these results may not necessarily signify shared genetic associations across multiple traits. Rather, this may point to more fundamental effects of these TWAS-identified genes that manifest in complex traits later in life.

We next examined whether placental GReX of these 9 genes correlate with fundamental traits at birth. We imputed expression into individual-level ELGAN genotypes ($N = 729$). Controlling for race, sex, gestational duration, inflammation of the chorion, and maternal age, as described in **Methods and Materials**, we tested for associations for 6 representative traits measured at birth or at 24 months: neonatal chronic lung disease, birth head circumference Z-score, fetal growth restriction, birth weight Z-score, necrotizing enterocolitis, and Bayley II Mental Development Index (MDI) at 24 months[15]. Shown in **Figure 5B** and **Supplemental Table S10**, at FDR-adjusted $P < 0.05$, we detected negative associations between *SEC11A* GReX and birthweight Z-score (effect size: -0.248, 95% adjusted CI: [-0.434,-0.063]) and GReX of *ATPAF2* and head circumference Z-score (-0.173, [-0.282,-0.064]). Furthermore, we detected negative associations between MDI and GReX of *RPL6* (-2.636, [-4.251,-1.02]) and *ERP29* (-3.332, [-4.987,-1.677]). As many of these genes encode for proteins involved in core processes (i.e., *RPL6* is involved in *trans*-activation of transcription and translation, and *SEC11A* has roles in cell migration and invasion)[45,46], understanding how the placental GReX of these genes affects neonatal traits may elucidate the potential long-lasting impacts of placental dysregulation.

Body size and metabolic placental GTAs show trait associations in mice

To further study functional consequences for selected TWAS-identified genes, we evaluated the 109 metabolic trait-associated genes in the Hybrid Mouse Diversity Panel (HMDP) for correlations with obesity-related traits[47]. This panel includes 100 inbred mice strains with extensive collection of obesity-related phenotypes from over 12,000 genes, with expression measured in a variety of adult tissues. Of the 109 genes, 73 were present in the panel and 36 showed significant cis-GReX associations with at least one obesity-related trait at FDR-adjusted P < 0.10 (**Supplemental Table S11**). For example, *EPB41L1* (*Epb4.1l1* in mice), a gene that mediates interactions in the erythrocyte plasma membrane, was associated with cholesterol and triglycerides in TWAS and showed 22 GReX associations with cholesterol, triglycerides, and HDL in mouse liver, adipose, and heart, with R^2 ranging between 0.09 and 0.31. Similarly, *UBC* (*Ubc* in mice), a ubiquitin maintaining gene, was associated with waist-hip ratio in the placental TWAS and showed 27 GReX associations with glucose in adults, insulin, and cholesterol in mouse aorta, liver, and adipose tissues in HMDP, with R^2 ranging between 0.08 and 0.14. Though generalizing these functional results from non-placental tissue in mice to humans is tenuous, we believe these 36 individually significant genes in the HMDP are fruitful targets for follow-up studies.

MOSTWAS reveals functional hypotheses for distal placental regulation of GTAs

An advantage of MOSTWAS's methodology is in functional hypothesis generation by identifying potential mediators that affect TWAS-identified genes. Using the distal-SNPs added-last test from MOSTWAS[22], we interrogated distal loci incorporated into expression models for trait associations, beyond the association at the local locus. For 88 of 248 associations, predicted expression from distal SNPs showed significant associations at FDR-adjusted $P < 0.05$ (**Figure 3A [Step 4]**, **Supplemental Table S6**). For each significant distal association, we identified a set of biomarkers that potentially affects transcription of the TWAS gene: a total of 9 regulatory protein-encoding genes (RPs) and 159 CpG sites across all 89 distal associations. Particularly, we detected two RPs, *DAB2* (distal mediator for *PAPPA* and diastolic blood pressure, distal $Z = -3.98$) and *EPS15*, both highly expressed in placenta[48,49]. Mediated through *EPS15* (overall distal $Z = 7.11$ and 6.33, respectively), distally predicted expression of *SPATA13* and *FAM214A* showed association with waist-hip ratio. *EPS15* itself showed a TWAS association for waist-hip ratio (**Supplemental Table S6**), and the direction of the *EPS15* GTA was opposite to those of *SPATA13*

and *FAM214A*. Furthermore, *RORA*, a gene encoding a transcription factor involved in inflammatory signaling[50], showed a negative association with transcription of *UBA3*, a TWAS gene for fetal birthweight. Low placental *RORA* expression was previously shown to be associated with lower birthweight[51]. Aside from functions related to transcription regulation, the 9 RPs (*CUL5, DAB2, ELL, EPS15, RORA, SLC2A4RG, SMARCC1, NFKBIA, ZC3H15*) detected by MOSTWAS were enriched for several ontologies (**Supplemental Table S12**), namely catabolic and metabolic processes, response to lipids, and multiple nucleic acid-binding processes[29].

As we observed strong correlations between expressions of RP-TWAS gene pairs in ELGAN (**Supplemental Figure S14**), we then examined the associations between TWAS-identified genes and the GReX of any predicted mediating RPs in an external dataset. Using RICHS, we conducted a gene-based *trans*-eQTL scan using Liu *et al*'s Gene-Based Association Testing (GBAT) method[52] to computationally validate RP-TWAS gene associations. We predicted GReX of the RPs using *cis*-variants through leave-one-out cross-validation and scanned for associations with the respective TWAS genes (**Figure 4C**, **Supplemental Table S13**). We found a significant association between predicted *EPS15* and *FAM214A* expressions (effect size -0.24, FDR-adjusted $P = 0.019$). In addition, we detected a significant association between predicted *NFKBIA* and *HNRNPU* (effect size -0.26, FDR-adjusted $P = 1.9 \times 10^{-4}$). We also considered an Egger regression-based Mendelian randomization framework[53] in RICHS to estimate the causal effects of RPs on the associated TWAS genes (**Methods and Materials**) using, as instrumental variables, *cis*-SNPs correlated to the RP and uncorrelated with the TWAS genes. We estimated significant causal effects for two RP-TWAS gene pairs (**Figure 5C**, **Supplemental Table S14**): *EPS15* on *FAM214A* (causal effect estimate -0.58; 95% CI [0.21, 0.94]) and *RORA* on *UBA3* (0.58; [0.20, 0.96]). These GBAT and MR estimates between *EPS15* and *FAM214A* are in opposite directions of the simple correlations presented in **Supplemental Figure S14**. However, as discussed in previous TWAS and MR studies[17,53], correlations between GReX and a phenotype are not equivalent to correlations between full expression and the phenotype, as full expression is subject multiple post-transcriptional process, while GReX is not.

We also examined the CpG methylation sites MOSTWAS marked as potential mediators for expression of TWAS genes for overlap with *cis*-regulatory elements in the placenta from the ENCODE Project Phase II[10], identifying 34 CpG sites (mediating 29 distinct TWAS genes) that fall in *cis*-regulatory regions (**Supplemental Table S15**). Interestingly, one CpG site mediating (cg15733049, Chromosome 1:2334974) *FAM214A* is found in low-DNase activity sites in placenta samples taken at various timepoints; additionally, cg15733049 is local to *EPS15*, the RP predicted to mediate genetic regulation of *FAM214A*. Furthermore, expression of *LARS2*, a TWAS gene for adult BMI, is mediated by cg04097236 (found within *ELOVL2*), a CpG site found in low DNase or high H3K27 activity regions; *LARS2* houses multiple GWAS risk SNPs for type 2 diabetes[54] and has shown adult BMI TWAS associations in other tissues[17,28]. Results from these external datasets add more evidence that these mediators play a role in gene regulation of these TWAS-identified genes and should be investigated experimentally in future studies.

In-vitro assays reveal widespread transcriptomic consequences of EPS15 knockdown

Based on our computational results, we experimentally studied whether the inverse relationship between RP *EPS15* and its two prioritized target TWAS genes, *SPATA13* and *FAM214A*, is supported *in vitro*. We used a FANA oligonucleotide targeting *EPS15* to knock down *EPS15* expression in human placenta-derived JEG-3 choriocarcinoma cells and assessed the gene expression of the targets in no-addition controls, scramble oligo controls, and the knockdown variant via qRT-PCR. JEG-3 cells were selected for study based on their know first trimester-like phenotypes, including the synthesis and secretion of hCG, human placenta lactogen, progesterone, estrone, and estradiol[55,56]. Addition of FANA-EPS15 to JEG-3 cells decreased *EPS15* gene expression, while increasing the expression of *SPATA13* and *FAM214A* (50% decrease in *EPS15* expression, 795% and 377% increase in *SPATA13* and *FAM214A* expression, respectively). At FDR-adjusted $P < 0.10$, changes in gene expression of *EPS15* and downstream targets from the scramble were statistically significant against the knockdown oligo. Similarly, changes in gene expression between the control mRNA and RP and target mRNA were statistically significant (**Figure 6A**).

To further investigate the transcriptomic consequences of *EPS15* knockdown *in vitro*, we measured transcriptome-wide gene expression in the choriocarcinoma cell lines via RNA-seq and conducted differential gene expression analysis across the knockdown cells and scramble oligo controls[57–59]. Due to small sample sizes, we define a differentially expression gene with absolute $\log_2$-fold change greater than 0.5 at $P < 1.32 \times 10^{-6}$, a Bonferroni correction across all assayed genes (**Methods**). We detected 650 genes down-regulated and 838 genes up-regulated in the *EPS15* knockdown cells, validating the negative correlations between *EPS15* and *SPATA13* and *FAM214A* observed in qRT-PCR (**Figure 6B**, **Supplemental Table S16-S17**). In particular, these down-regulated genes were enriched for cell cycle, cell proliferation, or replication ontologies, while up-regulated genes were enriched for multiple different pathways, including lipid-related processes, cell movement, and extracellular organization (**Figure 5C, Supplemental Table S18-S19**). Enrichments for cellular, molecular, and disease pathway ontologies support these enrichments (**Supplemental Figure S15**, **Supplemental Table S18-S19**). Though we could not study the effects of these three genes on body size-related traits, cis-GReX correlation analysis from the HMDP did reveal a negative cis-GReX correlation ($r = -0.31$, FDR-adjusted P = 0.07) between *Eps15* (mouse analog of human gene *EPS15*) and free fatty acids in mouse liver (**Supplemental Table S11**). These results prioritize *EPS15* for further study in larger cell line or animal studies as a potential regulator for multiple downstream genes, perhaps for genes affecting cell proliferation and replication in the placenta, like *SPATA13*[60].

DISCUSSION

The placenta has been understudied in large multi-tissue consortia efforts that study tissue-specific regulatory mechanisms[9,10] relevant to complex trait etiology. To address this gap, we systematically categorized placental gene-trait associations relevant to the DOHaD hypothesis using MOSTWAS, a method for enriching TWAS with distal genetic variants[22]. We detected 176 genes (enriched for cell growth and immune pathways) with transcriptome-wide significant associations, with the majority of GTAs linked to metabolic and neonatal/childhood traits. Furthermore, we could only estimate significantly positive placental GReX-mediated heritability for 4 neonatal traits but not for later-in-life traits. Many of these TWAS-identified genes, especially those with neonatal GTAs, showed multiple GTAs across trait

categories (9 genes with 3 or more GTAs). We examined phenome-wide GTAs for these 9 genes in UKBB and found enrichments for traits affecting in immune and circulatory system (e.g., immune cell, erythrocyte, and platelet counts). We followed up with selected early-life traits in ELGAN and found associations with neonatal body size and infant cognitive development. These results suggest that placental expression, mediated by fetal genetics, is most likely to have large effects on early-life traits, but these effects may persist later-in-life as etiologic antecedents for complex traits.

MOSTWAS also generates hypotheses for regulation of TWAS-detected genes, through distal mediating biomarkers, like transcription factors, miRNAs, or products downstream of CpG methylation islands[22]. Our computational results prioritized 89 GTAs with strong distal associations. We interrogated one such functional hypothesis: *EPS15*, a predicted RP-encoding gene in the EGFR pathway, regulates two TWAS genes positively associated with waist-hip ratio - *FAM214A*, a gene of unknown function, and *SPATA13*, a gene that regulates cell migration and adhesion[60–62]. In fact, *EPS15* itself showed a negative TWAS association with waist-hip ratio. In particular, *EPS15*, mainly involved in endocytosis, is a maternally imprinted gene and predicted to promote offspring health[49,63–65]. There is ample literature that implicates the protein product of *EPS15* as a direct or indirect transcription regulator. The protein Eps15 is an adaptor protein that regulates intracellular trafficking and has been detected in the nucleus of mammalian cells[66]. Once in the nucleus, Eps15 has shown to positively modulate transcription in a GAL4 transactivation assay[67]. Furthermore, Eps15 and its binding partner intersectin activate the Elk-1 transcription factor, pointing to Eps15's function in regulating gene expression in the nucleus[68]. Specific to the placenta, it has been proposed, through mouse models, that Eps15's interactions with multiple proteins suggest a role in cell adhesion of trophoblast to endothelial cells through biogenesis of exosomes and extracellular vesicles, a critical part of placental and fetal development[69–71].

In placental-derived choriocarcinoma epithelial cells, knockdown of *EPS15* showed increased expression of both *FAM214A* and *SPATA13*, as well as multiple genes involved in metabolic and hormone-related pathways. Though not implicating a direct causal effect, *EPS15*'s inverse association with *SPATA13* and *FAM214A* could provide more context to its full influence in placental developmental programming,

perhaps by affecting cell proliferation or adhesion pathways. *In vivo* animal experiments, albeit limited in scope and generalizability, can be employed to further investigate GTAs, building off results from the HMDP showing cis-GReX correlations between *EPS15* mouse analog and fatty acid levels. Although these cis-GReX correlations from HMDP cannot be generalized from mice to humans, our *in vitro* assay provides valuable evidence for *EPS15* genomic regulation in the placenta. Our results also support the potential of MOSTWAS to build mechanistic hypotheses for upstream regulation of TWAS genes that hold up to experimental rigor.

We conclude with limitations of this study and future directions. First, our analysis considers only placental tissue. Though many of our GTAs leverage distal-eQTL architecture which tend to be tissue-specific, the QTLs we leverage in TWAS may not be placenta-specific. A similar analysis across developmental and adult tissues could reveal more widespread genetic signals associated with these traits. Second, the ELGAN Study gathered molecular data from infants born extremely pre-term. If unmeasured confounders affect both prematurity and a trait of interest, GTAs could be subject to backdoor collider confounding[72]. However, significant TWAS genes did not show associations for gestational duration, suggesting minimal bias from this collider effect. An extensive comparison of genome-wide eQTL architecture between ELGAN and RICHS, highlighting differences in genetic effects on gene expression across pre-term status, could be of particular scientific importance. An interesting future endeavor could include negative control variables to account for unmeasured confounders in predictive models to allow for more generalizability of predictive models[73,74]. Fourth, though we did scan neonatal traits in ELGAN using individual-level genotypes, as the sample size is small, larger GWAS with longitudinal traits could allow for rigorous Mendelian randomization studies that investigate relationships between traits across the life course, in the context of placental regulation. Fifth, we curated a list of regulatory proteins to include as potential mediators but use RNA expression of the genes that code for these proteins as a proxy for abundance. We contend that RNA abundance of the gene is a noisy estimate of the protein abundance. An interesting extension of this analysis could consider a proteome-wide association study, using the MOSTWAS framework to identify protein interactions that are disease-related. Lastly, due to small sample sizes of other ancestry groups in ELGAN, we could only credibly

impute expression into samples from European ancestry and our TWAS only considers GWAS in populations of European ancestry[75]. We emphasize acquisition of larger genetic and genomic datasets from understudied and underserved populations, especially related to early-in-life traits.

Our findings reveal functional evidence for the fundamental influence of placental genetic and genomic regulation on developmental programming of early- and later-in-life traits, identifying placental gene-trait associations and testable functional hypotheses for upstream placental regulation of these genes. Future large-scale tissue-wide studies should emphasize the placenta as a core tissue for learning about the developmental origins of health and disease.

ONLINE METHODS

Data acquisition and quality control

Genotype data

Genomic DNA was isolated from umbilical cord blood and genotyping was performed using Illumina 1 Million Quad and Human OmniExpression-12 v1.0 arrays[34,76]. Prior to imputation, from the original set of 731,442 markers, we removed SNPs with call rate < 90% and MAF < 1%. We only consider genetic variants on autosomes. We did not use deviation from Hardy-Weinberg equilibrium as an exclusion criterion since ELGAN is an admixed population. This resulted in 700,845 SNPs. We removed 4 individuals out of 733 with sample-level missingness > 10% using PLINK[77]. We first performed strand-flipping according to the TOPMed Freeze 5 reference panel and using eagle and minimac4 for phasing and imputation[78–80]. Genotypes were coded as dosages, representing 0, 1, and 2 copies of the minor allele. The minor allele was coded in accordance with the NCBI Database of Genetic Variation[81]. Overall, after QC and normalization, we considered a total of 6,567,190 SNPs. We obtained processed genetic data from the Rhode Island Children's Health Study, as described before[36].

Expression data

mRNA expression was determined using the Illumina QuantSeq 3' mRNA-Seq Library Prep Kit, a method with high strand specificity[82]. mRNA-sequencing libraries were pooled and sequenced (single-end 50 bp)

on one lane of the Illumina HiSeq 2500. mRNA were quantified through pseudo-alignment with salmon[57] mapped to the GENCODE Release 31 (GRCh37) reference transcriptome. miRNA expression profiles were assessed using the HTG EdgeSeq miRNA Whole Transcriptome Assay (HTG Molecular Diagnostics, Tucson, AZ). miRNA were aligned to probe sequences and quantified using the HTG EdgeSeq System[83].

Genes and miRNAs with less than 5 counts for each sample were filtered, resulting in 12,020 genes and 2,047 miRNAs for downstream analysis. We only consider autosomal genes and miRNAs. Distributional differences between lanes were first upper-quartile normalized[84,85]. Unwanted technical and biological variation (e.g. tissue heterogeneity) was then estimated using RUVSeq[86], where we empirically defined transcripts not associated with outcomes of interest as negative control housekeeping probes[87]. One dimension of unwanted variation was removed from the variance-stabilized transformation of the gene expression data using the limma package[59,86–88]. We obtained pre-processed RNA expression data from the Rhode Island Children's Health Study, as described before[36]. Pre-processing steps for RNA expression data from the RICHS are different from those employed here in the ELGAN study.

Methylation data

Extracted DNA sequences were bisulfate-converted using the EZ DNA methylation kit (Zymo Research, Irvine, CA) and followed by quantification using the Infinium MethylationEPIC BeadChip (Illumina, San Diego, CA), which measures CpG loci at a single nucleotide resolution, as previously described[89–92]. Quality control and normalization were performed resulting in 856,832 CpG probes from downstream analysis, with methylation represented as the average methylation level at a single CpG site (β-value)[90,93–96]. DNA methylation data was imported into R for pre-processing using the minfi package[94,95]. Quality control was performed at the sample level, excluding samples that failed and technical duplicates; 411 samples were retained for subsequent analyses.

Functional normalization was performed with a preliminary step of normal-exponential out-of band (noob) correction method[97] for background subtraction and dye normalization, followed by the typical functional

normalization method with the top two principal components of the control matrix[94,95]. Quality control was performed on individual probes by computing a detection P-value and excluded 806 (0.09%) probes with non-significant detection (P > 0.01) for 5% or more of the samples. A total of 856,832 CpG sites were included in the final analyses. Lastly, the ComBat function was used from the sva package to adjust for batch effects from sample plate[98]. In addition, to account for cell-type heterogeneity, 5 surrogate values were estimated and removed from the data to account using the sva package, as previously described[15,90,98]. The data were visualized using density distributions at all processing steps. Each probe measured the average methylation level at a single CpG site. Methylation levels were calculated and expressed as β values, with

$$\beta = \frac{M}{U + M + 100},$$

where M is the intensity of the methylated allele and U is the intensity of the unmethylated allele. β-values were logit transformed to M values for statistical analyses[99]. Overall, after QC and normalization, we considered 846,233 CpG sites, only on autosomes.

GWAS summary statistics

Summary statistics were downloaded from the following consortia: the UK Biobank[23], Early Growth Genetics Consortium[24], Genetic Investigation of Anthropometric Traits[25], Psychiatric Genomics Consortium[26], and the Complex Trait Genetics Lab[27] (**Supplemental Table 1**). Genomic coordinates were transformed to the hg38 reference genome using liftOver[100,101]. SNP heritability for each trait and genetic correlations for all pairwise combinations of traits were estimated using LD score regression with the European ancestry sample from the 1000 Genomes Project as a reference for LD scores[31,32].

QTL mapping

The first step in the MOSTWAS pipeline is to scan for associations between SNPs and genes (genome-wide eQTL analysis) and between mediators and genes. We conducted genome-wide eQTL mapping between all genotypes and all genes in the transcriptome using a standard linear regression in MatrixeQTL[102]. Here, we ran an additive model with gene expression as the outcome, SNP dosage as the primary predictor of interest, with covariate adjustments for 20 genotype PCs (for population stratification),

sex, gestational duration, maternal age, maternal smoking status, and 10 expression PEER factors[103]. Mediators here are defined as RNA expression of genes that code for regulatory proteins (curated in TFcheckpoint[33]), miRNAs, and monomorphic CpG methylation sites. In sum, we call the expression or methylation of a mediator its intensity. We also conducted genome-wide mediator-QTL mapping with the intensity of mediators as the outcome with the same predictors as in the eQTL mapping. Lastly, we also assessed associations between mediators and gene expression using the same linear models, with mediator intensity as the main predictor. All intensities were scaled to zero mean and unit variance.

Estimation of SNP heritability of gene expression

An important step in a TWAS pipeline is estimation of SNP heritability of expression, as SNP heritability is a strong determinant of TWAS study power[17,104]. Heritability using genotypes within 1 Megabase of the gene of interest and any prioritized distal loci was estimated using the GREML-LDMS method, proposed to estimate heritability by correction for bias in LD in estimated SNP-based heritability[105]. Analysis was conducted using GCTA v1.93.1[106]. Briefly, Yang *et al* shows that estimates of heritability are often biased if causal variants have a different minor allele frequency (MAF) spectrums or LD structures from variants used in analysis. They proposed an LD and MAF-stratified GREML analysis, where variants are stratified into groups by MAF and LD, and genetic relationship matrices (GRMs) from these variants in each group are jointly fit in a multi-component GREML analysis.

Gene expression models

We used MOSTWAS to train predictive models of gene expression from germline genetics, including distal variants that were either close to associated mediators (transcription factors, miRNAs, CpG sites) or had large indirect effects on gene expression[22] (**Supplemental Figure S1**). Our assumption here is that distal-eQTLs of a gene that are local to transcription factor-encoding genes, miRNAs, or regulatory features local to CpG methylation sites may be potentially mediated by cis-QTLs to these local features. This assumption has been employed by multiple studies previously to identify trans-eQTLs in multiple tissues[107–110]. For CpG methylation sites, we used the maxprobes R package to filter out cross-reactive or polymorphic probes, which may induce bias[111–113]. MOSTWAS contains two methods of predicting

expression: (1) mediator-enriched TWAS (MeTWAS) and (2) distal-eQTL prioritization via mediation analysis. For MeTWAS, we first identified mediators strongly associated with genes through correlation analyses between all genes of interest and a set of distal mediators (FDR-adjusted $P < 0.05$). We then trained local predictive models (using SNPs within 1 Mb) of each mediator using either elastic net or linear mixed model, used these models to impute the mediator in the training sample, and included the imputed values for mediators as fixed effects in a regularized regression of the gene of interest. For DePMA, we first conducted distal eQTL analysis to identify all distal-eQTLs at $P < 10^{-6}$ and then local mediator-QTL analysis to identify all mediator-QTLs for these distal-eQTLs at FDR-adjusted $P < 0.05$. We tested each distal-eQTL for their absolute total mediation effect on the gene of interest through a permutation test and included eQTLs with significantly large effects in the final expression model. Full mathematical details are provided in Bhattacharya *et al* [22]. We considered only genes with significantly positive heritability at nominal $P < 0.05$ using a likelihood ratio test and five-fold McNemar's adjusted cross-validation $R^2 \geq 0.01$, a cross-validation cutoff used by many previous TWAS analyses[16,17,28,36,75,114,115]. McNemar's adjustment to the traditional R^2 is computed as

$$R^2_{adjusted} = 1 - (1 - R^2)\frac{n-1}{n-\nu-1},$$

where n is the sample size and ν is the number of predictors in this linear model. Since this R^2 is computed only between the observed and predicted expression values, $\nu = 1$.

TWAS tests of association

Overall TWAS test

In an external GWAS panel, if individual SNPs are available, model weights from either MeTWAS or DePMA can be multiplied by their corresponding SNP dosages to construct the Genetically Regulated eXpression (GReX) for a given gene. This value represents the portion of expression (in the given tissue) that is directly predicted or regulated by germline genetics. We run a linear model or test of association with phenotype using this GReX value for the eventual TWAS test of association.

If individual SNPs are not available, then the weighted burden Z-test, proposed by Gusev *et al*, can be employed using summary statistics[17]. Briefly, we compute

$$\tilde{Z} = \frac{w_G Z}{\left(w_G \boldsymbol{\Sigma}_{s,s} w_G^T\right)^{1/2}}.$$

Here, Z is the vector of Z-scores of SNP-trait associations for SNPs used in predicting expression. The vector w_G represents the vector of SNP-gene effects from MeTWAS or DePMA and $\boldsymbol{\Sigma}_{s,s}$ is the LD matrix (correlation matrix between genotypes) between the SNPs represented in w_G. The test statistic $\tilde{Z}$ can be compared to the standard Normal distribution for inference.

Permutation test

We implement a permutation test, condition on the GWAS effect sizes, to assess whether the same distribution of SNP-gene effect sizes could yield a significant associations by chance[17]. We permute w_G 1,000 times without replacement and recompute the weighted burden test to generate a null distribution for $\tilde{Z}$. This permutation test is only conducted for overall associations at $P < 2.5 \times 10^{-6}$.

Distal-SNPs added-last test

Lastly, we also implement a test to assess the information added from distal-eSNPs in the weighted burden test beyond what we find from local SNPs. This test is analogous to a group added-last test in regression analysis, applied here to GWAS summary statistics. Let Z_l and Z_d be the vector of Z-scores from GWAS summary statistics from local and distal-SNPs identified by a MOSTWAS model. The local and distal-SNP effects from the MOSTWAS model are represented in w_l and w_d. Formally, we test whether the weighted Z-score $\tilde{Z}_d = w_d Z_d$ from distal-SNPs is significantly larger than 0 given the observed weighted Z-score from local SNPs $\tilde{Z}_l = w_l Z_l$. We draw from the assumption that $(\tilde{Z}_d, \tilde{Z}_l)$ follow a bivariate Normal distribution. Namely, we conduct a two-sided Wald-type test for the null hypothesis:

$$H_0: w_d Z_d | w_l Z_l = \tilde{Z}_l = 0.$$

We can derive a null distribution using conditional of bivariate Normal distributions; see Bhattacharya *et al*[22].

Genetic heritability and correlation estimation

At the genome-wide genetic level, we estimated the heritability of and genetic correlation between traits via summary statistics using LD score regression[31]. On the predicted expression level, we adopted approaches from Gusev *et al* and Mancuso *et al* to quantify the heritability (h^2_{GE}) of and genetic correlations (ρ_{GE}) between traits at the predicted placental expression level[17,28]. We assume that the expected χ^2 statistic under a complex trait is a linear function of the LD score[31]. The effect size of the LD score on the χ^2 is proportional to h^2_{GE}:

$$E[\chi^2] = 1 + \left(\frac{N_{Tl}}{M}\right) h^2_{GE} + N_T a,$$

where N_T is the GWAS sample size, M is the number of genes, l is the LD scores for genes, and a is the effect of population structure. We estimated the LD scores of each gene by predicting expression in European samples of 1000 Genomes and computing the sample correlations and inferred h^2_{GE} using ordinary least squares. We employed RHOGE to estimate and test for significant genetic correlations between traits at the predicted expression level[28].

Multi-trait scans in UKBB and ELGAN

For 9 genes with 3 or more associations across traits of different categories, we conducted multi-trait TWAS scans in UK Biobank. Here, we used the weighted burden test in UKBB GWAS summary statistics from samples of European ancestry for 296 traits grouped by ICD code blocks (circulatory, congenital malformations, immune, mental disorders, musculoskeletal, neonatal, neurological, and respiratory). We also imputed expression for these genes in ELGAN using 729 samples with individual genotypes and conducted a multi-trait scan for 6 neonatal traits: neonatal chronic lung disease, head circumference Z-score, fetal growth restriction, birth weight Z-score, necrotizing enterocolitis, and Bayley II Mental Development Index (MDI) at 24 months. For continuous traits (head circumference Z-score, birth weight Z-score, and mental development index), we used a simple linear regression with the GReX of the gene as the main predictor, adjusting for race, sex, gestational duration (in days), inflammation of the chorion, and maternal age. For binary traits, we used a logistic regression with the same predictors and covariates. These covariates have been previously used in placental genomic studies of neonatal traits

because of their strong correlations with the outcomes and with placental transcriptomics and methylomics[15,90,116].

Validation analyses in RICHS

Using genotype and RNA-seq expression data from RICHS[36], we attempted to validate RP-TWAS gene associations prioritized from the distal-SNPs added last test in MOSTWAS. We first ran GBAT, a *trans*-eQTL mapping method from Liu *et al*[52] to assess associations between the loci around RPs and the expression of TWAS genes in RICHS. GBAT tests the association between the predicted expression of a RP with the expression of a TWAS gene, improving power of trans-eQTL mapping[117]. We also conduct directional Egger regression-based Mendelian randomization to estimate and test the causal effects of the expression of the RP on the expression of the TWAS gene[118].

Human Mouse Diversity Panel

To provide some functional evidence of gene associations with metabolic traits, we evaluated the 109 metabolic trait-associated genes from our human placental TWAS in the Hybrid Mouse Diversity Panel (HMDP) for correlations with obesity-related traits in mice[47]. This panel includes 100 inbred mice strains with extensive collection of obesity-related phenotypes (e.g., cholesterol, body fat percentage, insulin, etc) from over 12,000 genes, with expression measured in a variety of adult tissues (liver, adipose, aorta). We note that the HMDP only considers adult tissues and does not include placental gene expression. In the HMDP, we consider both trait correlation to tissue-specific gene expression and cis-GReX (genetically-regulated expression controlled by cis-eQTLs).

In-vitro functional assays

Cell culture and treatment

The JEG-3 choriocarcinoma cells were purchased from the American Type Culture Collection (Manassas, VA). Cells were grown in Gibco RMPI 1640, supplemented with 10% fetal bovine serum (FBS), and 1% penicillin/streptomycin at 37°C in 5% CO_2. Cells were plated at 2.1 x 10^6 cells per 75 cm^3 flask and incubated under standard conditions until achieving roughly 90% confluence. To investigate the effects of

gene silencing, we used AUMsilence FANA oligonucleotides for mRNA knockdown of *EPS15* (AUM Bio Tech, Philadelphia, PA) and subsequent analysis of predicted downstream target genes *SPATA13* and *FAM214A*. On the day of treatment, cells were seeded in a 24-well culture plate at 0.05 x 10^6 cells per well. Cells were plated in biological duplicate. FANA oligos were dissolved in nuclease-free water to a concentration of 500µM, added to cell culture medium to reach a final concentration of 20µM and incubated for 24 hours at 37°C in 5% CO_2.

mRNA expression by quantitative Real-Time Polymerase Chain Reaction and RNA Sequencing

Treated and untreated JEG-3 cells were harvested in 350µL of buffer RLT plus. Successive RNA extraction was performed using the AllPrep DNA/RNA/miRNA Universal Kit according to the manufacturer's protocol. RNA was quantified using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA). RNA was then converted to cDNA, the next step toward analyzing gene expression. Next, mRNA expression was measured for *EPS15*, *SPATA13*, and *FAM214A* using real-time qRT-PCR and previously validated primers. Samples were run in technical duplicate. Real-time qRT-PCR Ct values were normalized against the housekeeping gene B-actin (*ACTB*), and fold changes in expression were calculated based on the ΔΔCT method[119]. Each sample was prepared in biological duplicate and technical duplicate. These samples were pooled together for sequencing to yield data representing four samples per exposure group. Fold change calculations using the Delta Delta CT method was calculated for each sample individually:

$$\text{Delta CT}_{\text{treated}} = \text{CT}_{\text{GOI, treated}} - \text{CT}_{\text{House, treated}}.$$

Treated and untreated samples of JEG-3 RNA previously extracted using the AllPrep DNA/RNA/miRNA Universal Kit were submitted to the High Throughput Sequencing Facility at UNC Chapel Hill for RNA sequencing. Total RNA samples were submitted for sequencing using the HS4000 HO platform. Samples were sequenced in duplicate, and libraries were prepped with the Kapa Stranded mRNA-Seq kit from Illumina Platforms. Sequencing was performed after all samples passed QAQC, with a paired-end read type, with a read length of 2x75.

Statistical analysis

Statistical analysis was performed using a one-way ANOVA (with nominal significance level $\alpha = 0.05$). Post-hoc pairwise t-tests (3 degrees of freedom for biological and technical duplicate) were utilized to investigate direct comparisons within sample groups.

Differential expression analysis

RNA-seq quantified counts (transcripts per kilobase million) were imported using tximeta[58] and summarized to the gene-level. Differential expression analysis between *EPS15* knockdown samples and scramble oligo controls was conducted using DESeq2[59]. Although false positive rates are well-controlled even at low sample sizes[120], true positive rates at such a low sample size are low for smaller thresholds of log-transformed fold changes. Thus, guided by Schurch et al's analysis, due to very limited sample size, we considered a gene to be differentially expressed if the absolute $\log_2$-fold change is greater than 1 and P < 0.05/37,788 = 1.32×10^{-6}. This P-value threshold is a strict Bonferroni threshold across 37,788 quantified genes.

AUTHOR CONTRIBUTIONS

Conceptualization: AB, TMO, RCF, HPS; Data curation: AB, ANF, VA, WL, YL, CP, CJM, TMO, RCF, HPS; Formal Analysis: AB, ANF, WL, HPS; Funding Acquisition: AJL, YL, RMJ, LS, KCKK, CJM, TMO, RCF, HPS; Investigation: AB, ANP, HJH, RCF, HPS; Methodology: AB, YL, HPS; Project administration: AB, RCF, HPS; Resources: TMO, RCF, HPS; Software: AB, WL, YL; Supervision: AB, YL, RCF, HPS; Validation: AB, CJM; Visualization: AB, ANF, RH; Writing – original draft: AB, RCF, HPS; Writing – review & editing: AB, AJL, ANF, VA, RH, WL, YL, RMJ, LS, HJH, KCKK, CJM, TMO, RCF, HPS

ACKNOWLEDGEMENTS

We thank Michael Love, Kanishka Patel, Michael Gandal, Chloe Yap, Bogdan Pasaniuc, and Jon Huang for their thoughts about the research. We also thank the following consortia and research groups for their publicly available GWAS summary statistics, eQTL datasets, and/or epigenomic annotations: the UK Biobank and the Neale Lab, the Genetic Investigation of Anthropometric Traits Consortium, the

Psychiatric Genetics Consortium, the Early Growth Genetics Consortium, the Complex Trait Genetics Lab, the Rhode Island Child Health Study, and the ENCODE Project.

FUNDING

This study was supported by grants from the National Institutes of Health (NIH), specifically the National Institute of Neurological Disorders and Stroke (U01NS040069; R01NS040069), the Office of the NIH Director (UG3OD023348), the National Institute of Environmental Health Sciences (T32-ES007018; P30ES019776; R24ES028597), the National Heart, Lung and Blood Institute (R01HL47883, R01HL148577), the National Institute of Nursing Research (K23NR017898; R01NR019245), and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01HD092374; R03HD101413; P50HD103573).

COMPETING INTERESTS

The authors declare that they have no competing interests.

CODE AVAILABILITY

Sample scripts for analysis are provided at https://github.com/bhattacharya-a-bt/dohad_twas. The MOSTWAS software is accessible at https://bhattacharya-a-bt.github.io/MOSTWAS/articles/MOSTWAS_vignette.html.

DATA AVAILABILITY

ELGAN mRNA, miRNA, and CpG methylation data can be accessed from the NCBI Gene Expression Omnibus GSE154829 and GSE167885. ELGAN genotype data is protected, as subjects are still enrolled in the study; any inquiries or data requests must be made to RCF and HPS. GWAS summary statistics can be accessed at the following links: UK Biobank (http://www.nealelab.is/uk-biobank), GIANT consortium (https://portals.broadinstitute.org/collaboration/giant/index.php/Main_Page), PGC (https://www.med.unc.edu/pgc/download-results/), EGG consortium (https://egg-consortium.org/), and CTG Lab (https://ctg.cncr.nl/software/). The RICHS eQTL dataset can be accessed via dbGaP accession

number phs001586.v1.p1 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001586.v1.p1). Placental epigenomic annotations from the ENCODE Project are available from https://www.encodeproject.org/, with specific accession numbers in Supplemental Table S13. All models and full TWAS results can be accessed at https://doi.org/10.5281/zenodo.4618036[121]. RNA-seq data generated in placental JEG-3 cells are available at GSE185071.

CODE AVAILABILITY

The MOSTWAS software is accessible, with example workflows, at https://bhattacharya-a-bt.github.io/MOSTWAS/articles/MOSTWAS_vignette.html.

REFERENCES

1. McKay, R. Developmental biology: Remarkable role for the placenta. *Nature* vol. 472 298–299 (2011).
2. Baron-Cohen, S. *et al.* Foetal oestrogens and autism. *Mol. Psychiatry* 1–9 (2019) doi:10.1038/s41380-019-0454-9.
3. Thornburg, K. L., O'Tierney, P. F. & Louey, S. The Placenta is a Programming Agent for Cardiovascular Disease. *Placenta* **31**, S54 (2010).
4. Gillman, M. W. Developmental origins of health and disease. *New England Journal of Medicine* vol. 353 1848–1850 (2005).
5. Ursini, G. *et al.* Convergence of placenta biology and genetic risk for schizophrenia article. *Nat. Med.* **24**, 792–801 (2018).
6. Tedner, S. G., Örtqvist, A. K. & Almqvist, C. Fetal growth and risk of childhood asthma and allergic disease. *Clinical and Experimental Allergy* vol. 42 1430–1447 (2012).
7. Bronson, S. L. & Bale, T. L. The Placenta as a Mediator of Stress Effects on Neurodevelopmental Reprogramming. *Neuropsychopharmacology* vol. 41 207–218 (2016).
8. Peng, S. *et al.* Genetic regulation of the placental transcriptome underlies birth weight and risk of childhood obesity. *PLOS Genet.* **14**, e1007799 (2018).

9. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (80-.).* **369**, 1318–1330 (2020).

10. Abascal, F. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).

11. Delahaye, F. *et al.* Genetic variants influence on the placenta regulatory landscape. *PLoS Genet.* **14**, e1007785 (2018).

12. Marsit, C. J. Placental Epigenetics in Children's Environmental Health. *Semin. Reprod. Med.* **34**, 36–41 (2016).

13. Paquette, A. G. *et al.* Regions of variable DNA methylation in human placenta associated with newborn neurobehavior. *Epigenetics* **11**, 603–613 (2016).

14. Aplin, J. D., Myers, J. E., Timms, K. & Westwood, M. Tracking placental development in health and disease. *Nature Reviews Endocrinology* vol. 16 479–494 (2020).

15. Santos Jr, H. P. *et al.* Evidence for the placenta-brain axis: multi-omic kernel aggregation predicts intellectual and social impairment in children born extremely preterm. *Mol. Autism* **11**, 97 (2020).

16. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).

17. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).

18. Pierce, B. L. *et al.* Co-occurring expression and methylation QTLs allow detection of common causal variants and shared biological mechanisms. *Nat. Commun.* **9**, 1–12 (2018).

19. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).

20. Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022-1034.e6 (2019).

21. O'Shea, T. M. *et al.* The ELGAN study of the brain and related disorders in extremely low gestational age newborns. *Early Hum. Dev.* **85**, 719–725 (2009).

22. Bhattacharya, A., Li, Y. & Love, M. I. MOSTWAS: Multi-Omic Strategies for Transcriptome-Wide Association Studies. *PLOS Genet.* **17**, e1009398 (2021).

23. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

24. Middeldorp, C. M. *et al.* The Early Growth Genetics (EGG) and EArly Genetics and Lifecourse Epidemiology (EAGLE) consortia: design, results and future prospects. *Eur. J. Epidemiol.* **34**, 9 (2019).

25. Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**, 25–34 (2009).

26. Sullivan, P. F. *et al.* Psychiatric Genomics: An Update and an Agenda. *Am. J. Psychiatry* **175**, 15 (2018).

27. Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919 (2018).

28. Mancuso, N. *et al.* Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. *Am. J. Hum. Genet.* **100**, 473–487 (2017).

29. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, 199–205 (2019).

30. Mancuso, N. *et al.* Probabilistic fine-mapping of transcriptome-wide association studies. *Nat. Genet.* **51**, 675–682 (2019).

31. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

32. Auton, A. *et al.* A global reference for human genetic variation. *Nature* vol. 526 68–74 (2015).

33. Chawla, K., Tripathi, S., Thommesen, L., Lægreid, A. & Kuiper, M. TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. *Bioinformatics* **29**, 2519–2520 (2013).

34. Ådén, U. *et al.* Candidate gene analysis: Severe intraventricular hemorrhage in inborn preterm neonates. *J. Pediatr.* **163**, (2013).

35. Deyssenroth, M. A. *et al.* Whole-transcriptome analysis delineates the human placenta gene network and its associations with fetal growth. *BMC Genomics* **18**, 520 (2017).

36. 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).

37. Wainberg, M. *et al.* Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* **51**, 592–599 (2019).

38. Yuan, Z. *et al.* Testing and controlling for horizontal pleiotropy with probabilistic Mendelian randomization in transcriptome-wide association studies. *Nat. Commun.* **11**, 1–14 (2020).

39. Misra, D. P., Salafia, C. M., Charles, A. K. & Miller, R. K. Placental measurements associated with intelligence quotient at age 7 years. *J. Dev. Orig. Health Dis.* **3**, 190–197 (2012).

40. Nakamura, K. *et al.* Isopentenyl diphosphate isomerase, a cholesterol synthesizing enzyme, is localized in Lewy bodies. *Neuropathology* **35**, 432–440 (2015).

41. Nagarajan, R., Le, N., Mahoney, H., Araki, T. & Milbrandt, J. Deciphering peripheral nerve myelination by using Schwann cell expression profiling. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 8998–9003 (2002).

42. Lee, C. J. *et al.* CETP, LIPC, and SCARB1 variants in individuals with extremely high high-density lipoprotein-cholesterol levels. *Sci. Rep.* **9**, 1–7 (2019).

43. Chrifi, I. *et al.* CMTM4 regulates angiogenesis by promoting cell surface recycling of VE-cadherin to endothelial adherens junctions. *Angiogenesis* **22**, 75–93 (2019).

44. Kokkinos, M. I., Murthi, P., Wafai, R., Thompson, E. W. & Newgreen, D. F. Cadherins in the human placenta - Epithelial-mesenchymal transition (EMT) and placental development. *Placenta* **31**, 747–755 (2010).

45. Kenmochi, N. *et al.* A map of 75 human ribosomal protein genes. *Genome Res.* **8**, 509–523 (1998).

46. Oue, N. *et al.* Signal peptidase complex 18, encoded by SEC11A, contributes to progression via TGF-α secretion in gastric cancer. *Oncogene* **33**, 3918–3926 (2014).

47. Lusis, A. J. *et al.* The hybrid mouse diversity panel: A resource for systems genetics analyses of metabolic and cardiovascular traits. *Journal of Lipid Research* vol. 57 925–942 (2016).

48. Tao, W., Moore, R., Smith, E. R. & Xu, X. X. Endocytosis and physiology: Insights from disabled-2 deficient mice. *Frontiers in Cell and Developmental Biology* vol. 4 129 (2016).

49. Nelissen, E. C. M., van Montfoort, A. P. A., Dumoulin, J. C. M. & Evers, J. L. H. Epigenetics and

the placenta. *Hum. Reprod. Update* **17**, 397–417 (2011).

50. Oh, S. K. *et al.* RORα is crucial for attenuated inflammatory response to maintain intestinal homeostasis. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 21140–21149 (2019).

51. Everson, T. M. *et al.* Cadmium-associated differential methylation throughout the placental genome: Epigenome-wide association study of two U.S. birth cohorts. *Environ. Health Perspect.* **126**, (2018).

52. Liu, X. *et al.* GBAT: a gene-based association test for robust detection of trans-gene regulation. *Genome Biol.* **21**, 211 (2020).

53. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* **32**, 377–389 (2017).

54. Reiling, E. *et al.* Genetic association analysis of LARS2 with type 2 diabetes. *Diabetologia* **53**, 103–110 (2010).

55. Abou-Kheir, W., Barrak, J., Hadadeh, O. & Daoud, G. HTR-8/SVneo cell line contains a mixed population of cells. *Placenta* **50**, 1–7 (2017).

56. Drwal, E., Rak, A. & Gregoraszczuk, E. Co-culture of JEG-3, BeWo and syncBeWo cell lines with adrenal H295R cell line: an alternative model for examining endocrine and metabolic properties of the fetoplacental unit. *Cytotechnology* **70**, 285 (2018).

57. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).

58. Love, M. I. *et al.* Tximeta: Reference sequence checksums for provenance identification in RNA-seq. *PLOS Comput. Biol.* **16**, e1007664 (2020).

59. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

60. Jean, L. *et al.* Activation of Rac by Asef2 promotes myosin IIdependent contractility to inhibit cell migration on type I collagen. *J. Cell Sci.* **126**, 5585–5597 (2013).

61. Bristow, J. M. *et al.* The Rho-family GEF Asef2 activates Rac to modulate adhesion and actin dynamics and thereby regulate cell migration. *J. Cell Sci.* **122**, 4535–4546 (2009).

62. Kawasaki, Y. *et al.* Identification and characterization of Asef2, a guanine-nucleotide exchange

factor specific for Rac1 and Cdc42. *Oncogene* **26**, 7620–7627 (2007).

63. Piedrahita, J. A. The role of imprinted genes in fetal growth abnormalities. *Birth Defects Research Part A - Clinical and Molecular Teratology* vol. 91 682–692 (2011).

64. Diplas, A. I. *et al.* Differential expression of imprinted genes in normal and IUGR human placentas. *Epigenetics* **4**, 235–240 (2009).

65. Marsit, C. J. *et al.* Placenta-imprinted gene expression association of infant neurobehavior. *J. Pediatr.* **160**, 854 (2012).

66. Van Bergen En Henegouwen, P. M. P. Eps15: A multifunctional adaptor protein regulating intracellular trafficking. *Cell Communication and Signaling* vol. 7 24 (2009).

67. Vecchi, M. *et al.* Nucleocytoplasmic Shuttling of Endocytic Proteins. *J. Cell Biol.* **153**, 1511–1518 (2001).

68. Adams, A., Thorn, J. M., Yamabhai, M., Kay, B. K. & O'Bryan, J. P. Intersectin, an Adaptor Protein Involved in Clathrin-mediated Endocytosis, Activates Mitogenic Signaling Pathways *. *J. Biol. Chem.* **275**, 27414–27420 (2000).

69. B, C. *et al.* Comparative systems biology of human and mouse as a tool to guide the modeling of human placental pathology. *Mol. Syst. Biol.* **5**, (2009).

70. Burkova, E. E., Sedykh, S. E. & Nevinsky, G. A. Human Placenta Exosomes: Biogenesis, Isolation, Composition, and Prospects for Use in Diagnostics. *Int. J. Mol. Sci. 2021, Vol. 22, Page 2158* **22**, 2158 (2021).

71. Mulcahy, L. A., Pink, R. C. & Carter, D. R. F. Routes and mechanisms of extracellular vesicle uptake. *https://doi.org/10.3402/jev.v3.24641* **3**, (2014).

72. Paternoster, L., Tilling, K. & Davey Smith, G. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges. *PLOS Genet.* **13**, e1006944 (2017).

73. VanderWeele, T. J., Tchetgen, E. J. T., Cornelis, M. & Kraft, P. Methodological challenges in Mendelian randomization. *Epidemiology* **25**, 427 (2014).

74. X, S., W, M., JC, N. & EJT, T. Multiply robust causal inference with double-negative control adjustment for categorical unmeasured confounding. *J. R. Stat. Soc. Series B. Stat. Methodol.* **82**,

521–540 (2020).

75. Bhattacharya, A. *et al.* A framework for transcriptome-wide association studies in breast cancer in diverse study populations. *Genome Biol.* **21**, 42 (2020).

76. Yasuno, K. *et al.* Genome-wide association study of intracranial aneurysm identifies three new risk loci. *Nat. Genet.* **42**, 420–425 (2010).

77. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet* **81**, 559–575 (2007).

78. Kowalski, M. H. *et al.* Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet.* **15**, e1008500 (2019).

79. Loh, P. R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).

80. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

81. Sherry, S. T. *et al.* DbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).

82. A Eaves, L. *et al.* A role for microRNAs in the epigenetic control of sexually dimorphic gene expression in the human placenta. *Epigenomics* **12**, 1543–1558 (2020).

83. Qi, Z. *et al.* Reliable Gene Expression Profiling from Small and Hematoxylin and Eosin–Stained Clinical Formalin-Fixed, Paraffin-Embedded Specimens Using the HTG EdgeSeq Platform. *J. Mol. Diagnostics* **21**, 796–807 (2019).

84. Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**, 94 (2010).

85. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).

86. Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor analysis

of control genes or samples. *Nat. Biotechnol.* **32**, 896–902 (2014).

87. Gagnon-Bartsch, J. A. & Speed, T. P. Using control genes to correct for unwanted variation in microarray data. *Biostatistics* **13**, 539–552 (2012).

88. Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann. Appl. Stat.* **10**, 946–963 (2016).

89. Addo, K. A. *et al.* Acetaminophen use during pregnancy and DNA methylation in the placenta of the extremely low gestational age newborn (ELGAN) cohort. *Environ. Epigenetics* **5**, (2019).

90. Santos, H. P. *et al.* Epigenome-wide DNA methylation in placentas from preterm infants: association with maternal socioeconomic status. *Epigenetics* **14**, 751–765 (2019).

91. Bulka, C. M. *et al.* Placental CpG methylation of inflammation, angiogenic, and neurotrophic genes and retinopathy of prematurity. *Investig. Ophthalmol. Vis. Sci.* **60**, 2888–2894 (2019).

92. Clark, J. *et al.* Associations between placental CpG methylation of metastable epialleles and childhood body mass index across ages one, two and ten in the Extremely Low Gestational Age Newborns (ELGAN) cohort. *Epigenetics* **14**, 1102–1111 (2019).

93. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).

94. Fortin, J.-P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* **15**, 503 (2014).

95. Fortin, J. P., Triche, T. J. & Hansen, K. D. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* **33**, 558–560 (2017).

96. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).

97. Triche, T. J., Weisenberger, D. J., Van Den Berg, D., Laird, P. W. & Siegmund, K. D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res.* **41**, e90 (2013).

98. Leek, J. T. & Storey, J. D. Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. *PLoS Genet.* **3**, e161 (2007).

99. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by

microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).

100. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).

101. Lawrence, M., Carey, V. & Gentleman, R. liftOver. (2020).

102. Shabalin, A. A. Gene expression Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012).

103. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7**, 500–507 (2012).

104. Cao, C. *et al.* Power analysis of transcriptome-wide association study: Implications for practical protocol choice. *PLoS Genet.* **17**, e1009405 (2021).

105. Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* **47**, 1114–1120 (2015).

106. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

107. Pierce, B. L. *et al.* Mediation Analysis Demonstrates That Trans-eQTLs Are Often Explained by Cis-Mediation: A Genome-Wide Analysis among 1,800 South Asians. *PLoS Genet.* **10**, (2014).

108. Hawe, J. S. *et al.* Network reconstruction for trans acting genetic loci using multi-omics data and prior information. *bioRxiv* 2020.05.19.101592 (2020) doi:10.1101/2020.05.19.101592.

109. Yang, F., Wang, J., Pierce, B. L. & Chen, L. S. Identifying cis-mediators for trans-eQTLs across many human tissues using genomic mediation analysis. *Genome Res.* **27**, 1859–1871 (2017).

110. Yang, F. *et al.* CCmed: cross-condition mediation analysis for identifying robust trans-eQTLs and assessing their effects on human traits. *bioRxiv* 803106 (2019) doi:10.1101/803106.

111. Pidsley, R. *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* **17**, 1–17 (2016).

112. McCartney, D. L. *et al.* Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genomics Data* **9**, 22–24 (2016).

113. Chen, M. markgene/maxprobes: Methylation Array Cross-Reactive Probes. (2016).

114. Gusev, A. *et al.* Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nat. Genet.* **50**, 538–548 (2018).

115. Mancuso, N. *et al.* Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. *Nat. Commun.* **9**, (2018).

116. Santos, H. P. *et al.* Discrimination exposure and DNA methylation of stress-related genes in Latina mothers. *Psychoneuroendocrinology* **98**, 131–138 (2018).

117. Mefford, J. *et al.* Efficient Estimation and Applications of Cross-Validated Genetic Predictions to Polygenic Risk Scores and Linear Mixed Models. *J. Comput. Biol.* **27**, 599–612 (2020).

118. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).

119. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* **25**, 402–408 (2001).

120. Schurch, N. J. *et al.* How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* **22**, 839 (2016).

121. Bhattacharya, A. & Santos Jr, H. Distal-mediator enriched placental gene expression models and TWAS results for DOHaD-related traits. (2021) doi:10.5281/ZENODO.4618037.

FIGURE LEGENDS

Figure 1: ***Overview of the DOHaD Hypothesis and study scheme.*** The placenta facilitates important functions *in utero*, including nutrient transfer, metabolism, gas exchange, neuroendocrine signaling, growth hormone production, and immunologic control. Accordingly, it is a master regulator of the intrauterine environment and is core to the Developmental Origins of Health and Disease (DOHaD) hypothesis. Placental genomic regulation is influenced by both genetic and environmental factors and affects placental developmental programming. In turn, this programming has been shown to have profound impacts on a variety of disorders and traits, both early- and later-in-life.

Figure 2: ***Overview of analytic pipeline***. (A) Predictive models of placental expression are trained from germline genetics, enriched for mediating biomarkers using MOSTWAS[22] and externally validated in RICHS[36] eQTL data. **(B)** Predictive models are integrated with GWAS for 40 traits to detect placental gene-trait associations (GTAs). **(C)** GTAs are followed up with gene ontology analyses, probabilistic fine-mapping, and phenome-wide scans of genes with multiple GTAs. Relationships between identified distal mediators and TWAS genes are investigated further in RICHS and ENCODE[10]. Expression-mediated genetic heritability of and correlations between traits are estimated. **(D)** In-vitro validation of prioritized transcription factor-TWAS gene pairs are conducted using placenta-derived choriocarcinoma cells by gene silencing and qRT-PCR to measure TWAS-gene expression.

Figure 3: ***Placental MOSTWAS prediction and association test results.*** **(A)** Overview of TWAS association testing pipeline with number of gene-trait associations (GTAs) across unique genes over various levels of TWAS tests. **(B)** Kernel density plots of in- (through cross-validation in ELGAN, red) and out-sample (external validation in RICHS, blue) McNemar's adjusted R^2 between predicted and observed expression. Dotted and solid lines represent the mean and median of the respective distribution, respectively. **(C)** Bar graph of numbers of TWAS GTAs at overall TWAS $P < 2.5 \times 10^{-6}$ and permutation FDR-adjusted $P < 0.05$ (X-axis) across traits (Y-axis). The total number of GTAs per trait are labeled, colored by the category of each trait. The bar is broken down by numbers of GTAs with (orange) and without (green) significant distal expression-mediated associations, as indicated by FDR-adjusted $P < 0.05$ for the distal-SNPs added-last test. **(D)** Enrichment plot of over-representation in 176 TWAS genes of PANTHER pathways (Y-axis) with -$\log_{10}$ FDR-adjusted P-value (X-axis). The size of the point gives the relative enrichment ratio for the given pathway.

Figure 4: ***Trait genetic heritability and correlations mediated by placental expression.*** (**A**) Box plot of expression-mediated trait heritability (h^2_{GE}) (Y-axis) by category (X-axis), with labels if $\hat{h}^2_{GE}$ is significantly greater than 0 using jack-knife test of significance. (**B**) Box plot of expression-mediated trait heritability (h^2_{GE}) standardized by SNP heritability (h^2) (Y-axis) by category (X-axis), with labels if $\hat{h}^2_{GE}$ is significantly

greater than 0 using jack-knife test of significance. (**C**) Forest plot of significant placenta expression-mediated genetic correlations and 95% FDR-adjusted confidence intervals between traits from different categories.

Figure 5: ***Computational follow-up analyses of TWAS-prioritized genes.*** (**A**) Boxplot of -log10 FDR-adjusted P-value of multi-trait scans of GTAs in UKBB, grouped by 8 ICD code blocks across 9 genes with multiple TWAS GTAs across different trait categories. The red dotted line represents FDR-adjusted P = 0.05. (**B**) Forest plot of GTA association estimates and 95% FDR-adjusted confidence intervals for 6 neonatal traits in ELGAN for 9 genes with multiple TWAS GTAs across categories. The red line shows a null effect size of 0, and associations are colored blue for associations at FDR-adjusted P < 0.05. (**C**) Follow-up GBAT and Mendelian randomization (MR) analysis results using RICHS data. On the left, effect size and 95% adjusted confidence intervals from GBAT (X-axis) between GReX of RP-encoding genes and TWAS gene associations (pairs given on Y-axis). On the right, MR effect size and 95% adjusted confidence interval (X-axis) of RP-gene on TWAS gene (pairs on Y-axis). The red line shows a null effect size of 0, and associations are colored blue for associations at FDR-adjusted P < 0.05.

Figure 6: ***In vitro experiments in EPS15-knockdown human placenta-derived choriocarcinoma epithelial cells.*** (**A**) Bar graph of the gene expression fold-changes from the qRT-PCR from JEG-3 RNA. Nominal P-values of pairwise t-tests are shown, with an asterisk if Benjamini-Hochberg *FDR*-adjusted $P < 0.10$. Note differences in Y-axis scales. **(B)** Volcano plot of $\log_2$ fold change (X-axis) of differential expression across EP15 knockdown cells and scramble oligo nucleotide against $-\log_{10}$ FDR-adjusted P-value (Y-axis). Up-regulated genes are in red and down-regulated genes in blue. Top up- and down-regulated genes by P-value are labeled, as well as *EPS15*, *SPATA13*, and *FAM214A*. **(C)** Enrichment plot of over-representation of down- (blue) and up-regulated (red) genes of PANTHER and KEGG pathways (Y-axis) with $-\log_{10}$ FDR-adjusted P-value (X-axis). The size of the point gives the enrichment ratio.

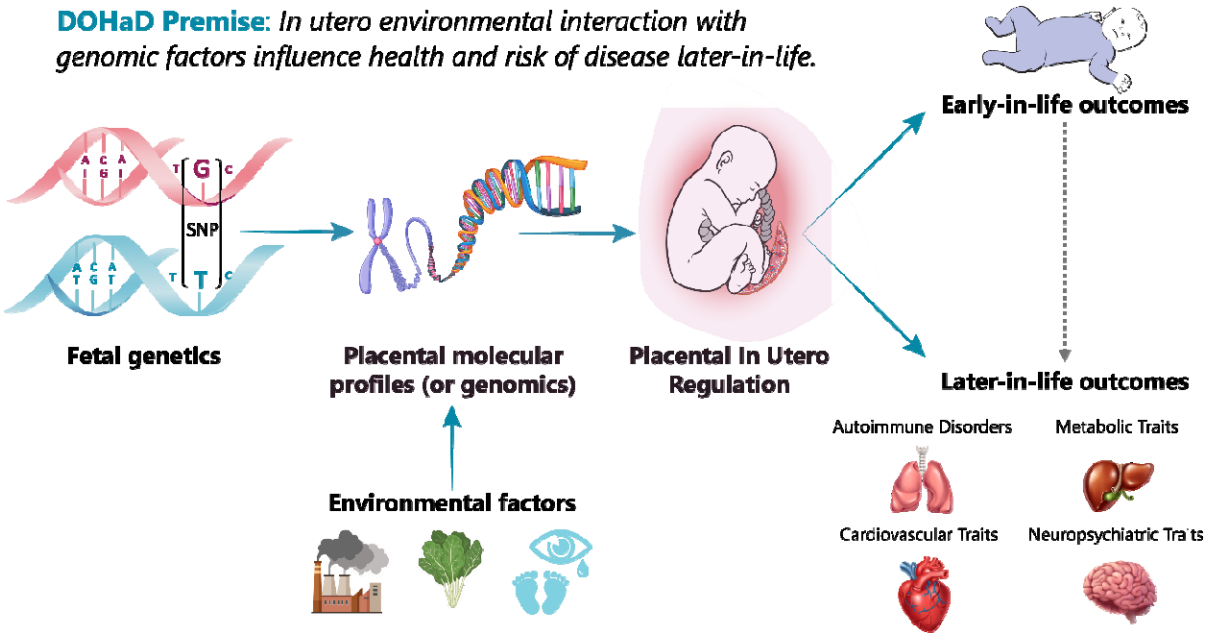

Figure 1: **Overview of the DOHaD Hypothesis and study scheme.** (**A**) The placenta facilitates important functions *in utero*, including nutrient transfer, metabolism, gas exchange, neuroendocrine signaling, growth hormone production, and immunologic control. Accordingly, it is a master regulator of the intrauterine environment and is core to the Developmental Origins of Health and Disease (DOHaD) hypothesis. Placental genomic regulation is influenced by both genetic and environmental factors and affects placental developmental programming. In turn, this programming has been shown to have profound impacts on a variety of disorders and traits, both early- and later-in-life.

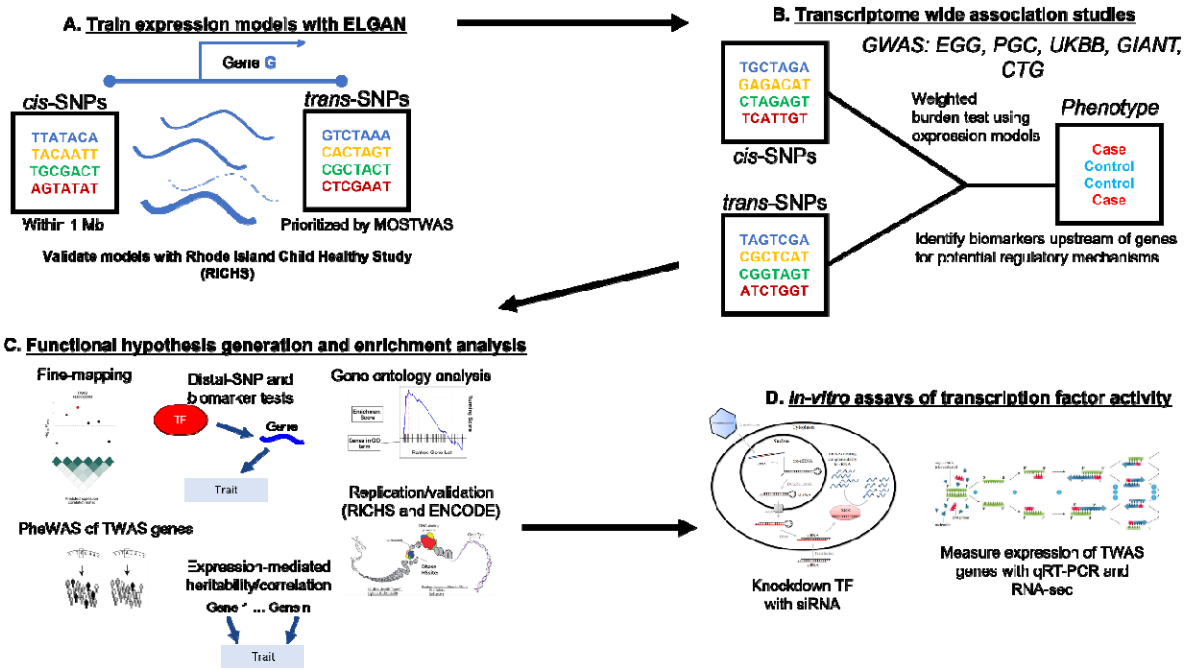

Figure 2: **Overview of analytic pipeline**. (A) Predictive models of placental expression are trained from germline genetics, enriched for mediating biomarkers using MOSTWAS[22] and externally validated in RICHS[36] eQTL data. **(B)** Predictive models are integrated with GWAS for 40 traits to detect placental gene-trait associations (GTAs). **(C)** GTAs are followed up with gene ontology analyses, probabilistic fine-mapping, and phenome-wide scans of genes with multiple GTAs. Relationships between identified distal mediators and TWAS genes are investigated further in RICHS and ENCODE[10]. Expression-mediated genetic heritability of and correlations between traits are estimated. **(D)** In-vitro validation of prioritized transcription factor-TWAS gene pairs are conducted using placenta-derived choriocarcinoma cells by gene silencing and qRT-PCR to measure TWAS-gene expression.

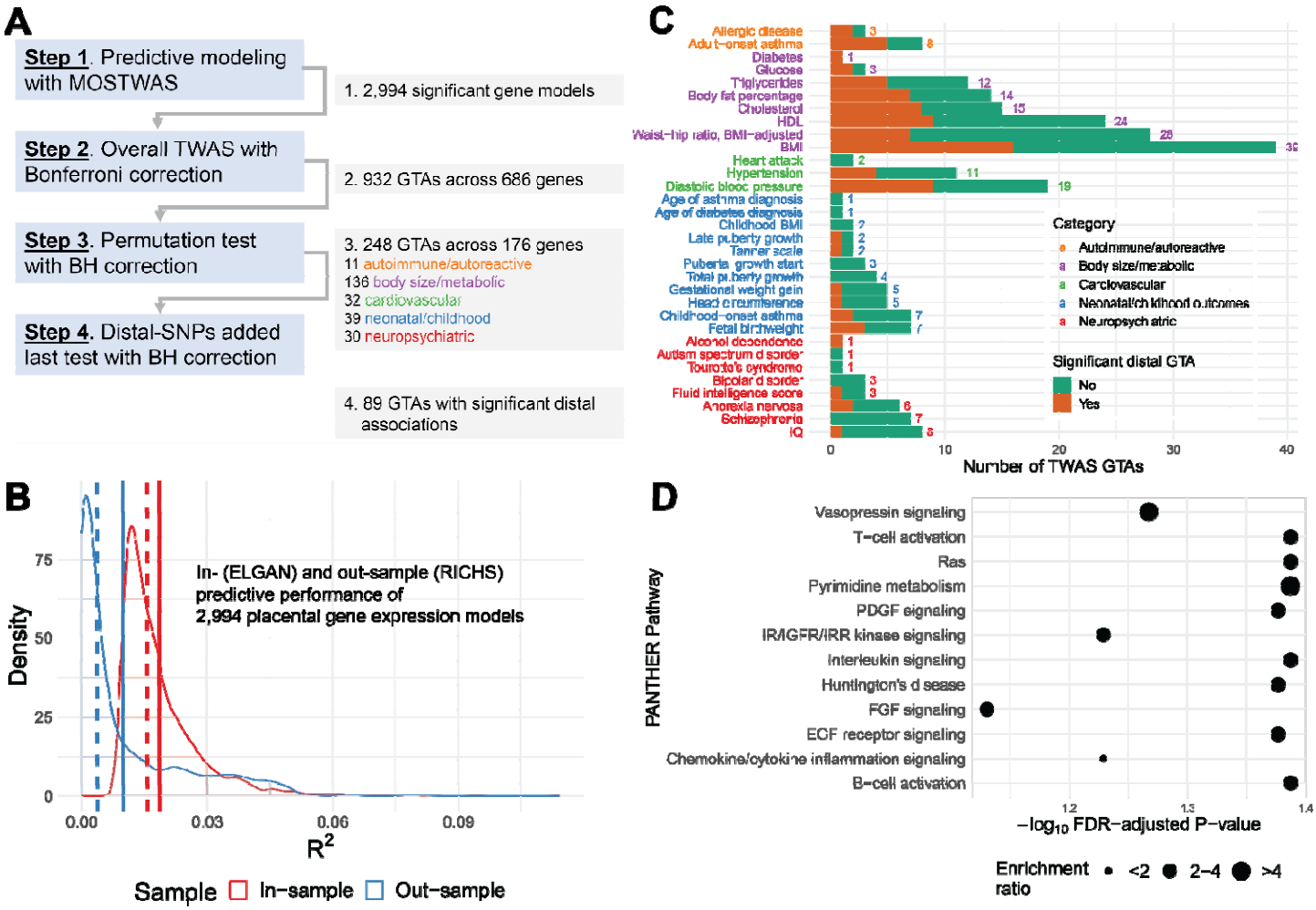

Figure 3: Placental MOSTWAS prediction and association test results. (**A**) Overview of TWAS association testing pipeline with number of gene-trait associations (GTAs) across unique genes over various levels of TWAS tests. (**B**) Kernel density plots of in- (through cross-validation in ELGAN, red) and out-sample (external validation in RICHS, blue) McNemar's adjusted between predicted and observed expression. Dotted and solid lines represent the mean and median of the respective distribution, respectively. (**C**) Bar graph of numbers of TWAS GTAs at overall TWAS and permutation FDR-adjusted (X-axis) across traits (Y-axis). The total number of GTAs per trait are labeled, colored by the category of each trait. The bar is broken down by numbers of GTAs with (orange) and without (green) significant distal expression-mediated associations, as indicated by FDR-adjusted for the distal-SNPs added-last test. (**D**) Enrichment plot of over-representation in 176 TWAS genes of PANTHER pathways (Y-axis) with -$\log_{10}$ FDR-adjusted P-value (X-axis). The size of the point gives the relative enrichment ratio for the given pathway.

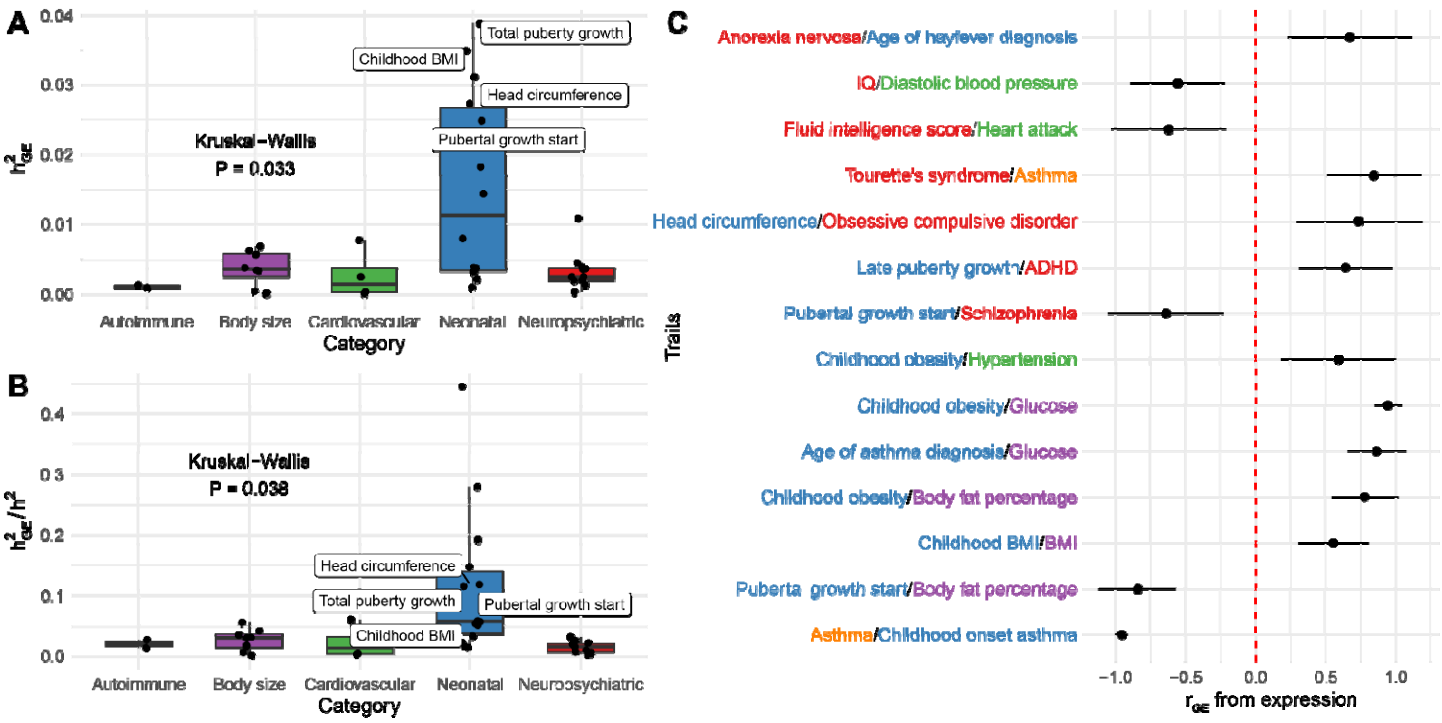

Figure 4: Trait genetic heritability and correlations mediated by placental expression. (**A**) Box plot of expression-mediated trait heritability ( ) (Y-axis) by category (X-axis), with labels if is significantly greater than 0 using jack-knife test of significance. (**B**) Box plot of expression-mediated trait heritability ( ) standardized by SNP heritability ( ) (Y-axis) by category (X-axis), with labels if is significantly greater than 0 using jack-knife test of significance. (**C**) Forest plot of significant placenta expression-mediated genetic correlations and 95% FDR-adjusted confidence intervals between traits from different categories.

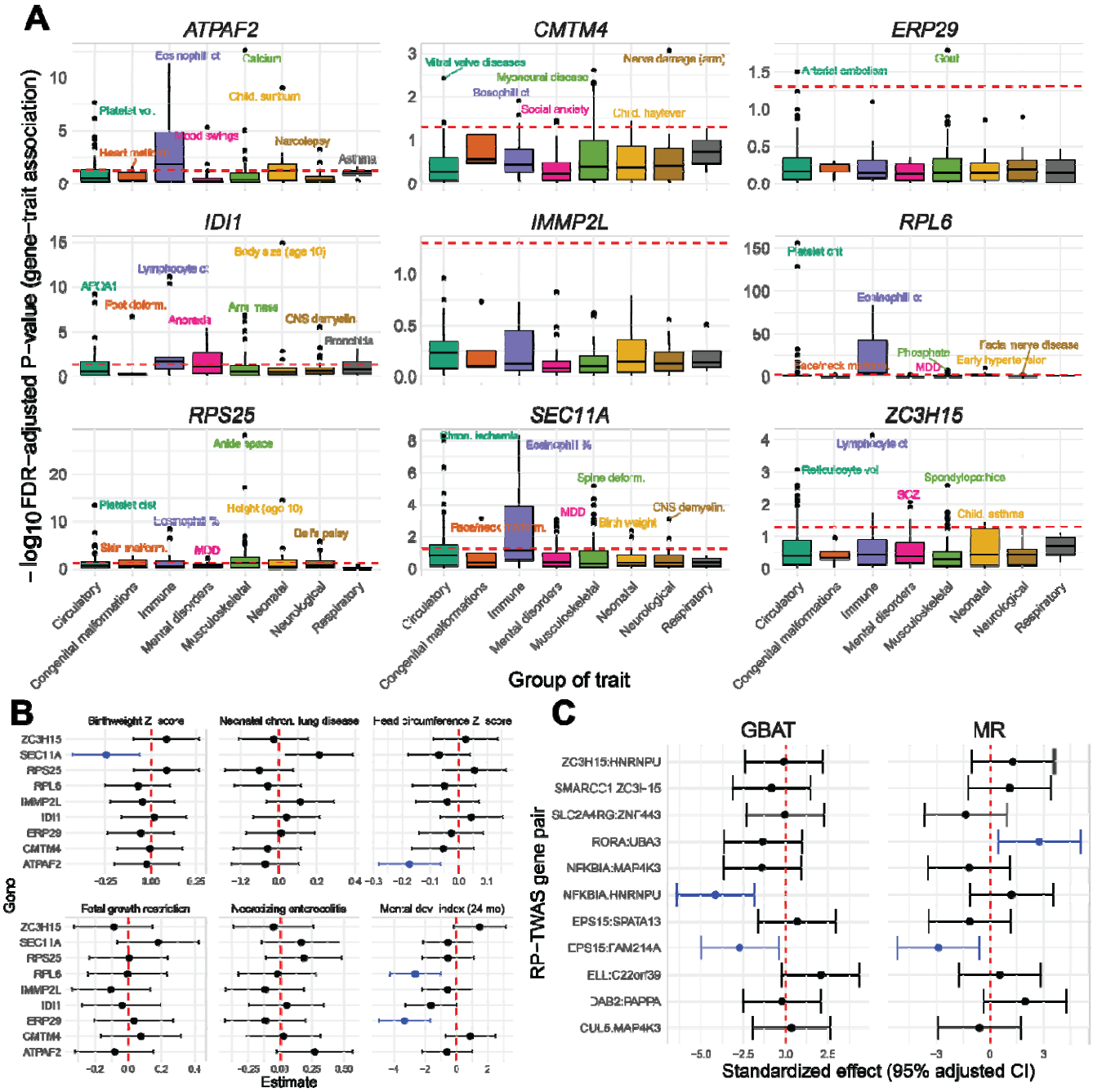

Figure 5: ***Computational follow-up analyses of TWAS-prioritized genes.*** (**A**) Boxplot of -log10 FDR-adjusted P-value of multi-trait scans of GTAs in UKBB, grouped by 8 ICD code blocks across 9 genes with multiple TWAS GTAs across different trait categories. The red dotted line represents FDR-adjusted P = 0.05. (**B**) Forest plot of GTA association estimates and 95% FDR-adjusted confidence intervals for 6 neonatal traits in ELGAN for 9 genes with multiple TWAS GTAs across categories. The red line shows a null effect size of 0, and associations are colored blue for associations at FDR-adjusted P < 0.05. (**C**) Follow-up GBAT and Mendelian randomization (MR) analysis results using RICHS data. On the left, effect size and 95% adjusted confidence intervals from GBAT (X-axis) between GReX of RP-encoding genes and TWAS gene associations (pairs given on Y-axis). On the right, MR effect size and 95% adjusted confidence interval (X-axis) of RP-gene on TWAS gene (pairs on Y-axis). The red line shows a null effect size of 0, and associations are colored blue for associations at FDR-adjusted P < 0.05.

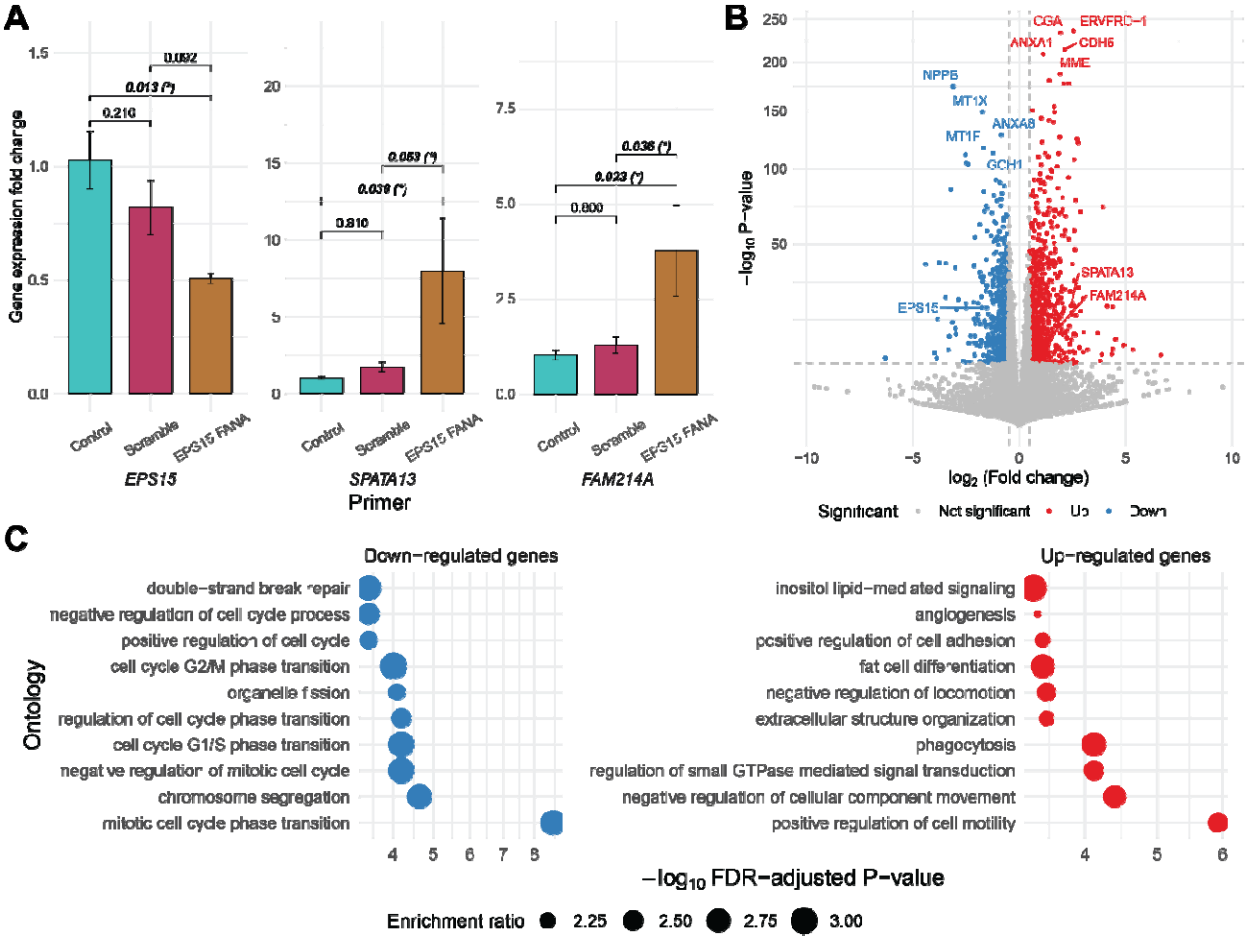

Figure 6: **In vitro experiments in EPS15-knockdown human placenta-derived choriocarcinoma epithelial cells** (***A***) Bar graph of the gene expression fold-changes from the qRT-PCR from JEG-3 RNA. Nominal P-values of pairwise t-tests are shown, with an asterisk if Benjamini-Hochberg *FDR*-adjusted . Note differences in Y-axis scales. **(B)** Volcano plot of $\log_2$ fold change (X-axis) of differential expression across EP15 knockdown cells and scramble oligo nucleotide against $-\log_{10}$ FDR-adjusted P-value (Y-axis). Up-regulated genes are in red and down-regulated genes in blue, with absolute $\log_2$-fold change and . Top up- and down-regulated genes by P-value are labeled, as well as EPS15, SPATA13, and FAM214A. **(C)** Enrichment plot of over-representation of down- (blue) and up-regulated (red) genes of biological process ontologies (Y-axis) with $-\log_{10}$ FDR-adjusted P-value (X-axis). The size of the point gives the enrichment ratio.

MAIN TABLES

Table 1: TWAS-significant genes that do not overlap significant GWAS loci (within 500 kB)

Category	Trait	Gene
Autoimmune/autoreactive	Adult-onset asthma	*NBPF19, ZC3H15, SAR1B, PPP3CB-AS1*
	Allergic disease	*PLA2G2A, NUSAP1, EIF4ENIF1*
Body size/metabolic	BMI	*FCRL2, NDUFS1, ANO10, NISCH, SLC35G2, TSPAN5, IL15, RPS23, NT5E, CD274, DENND1A, ZER1, MRPS16, SAMD8, MUS81, C12orf45, OGFOD2, SYT16, DNAL1, ARPIN-AP3S2, SENP3, NSRP1, COPRS, WDR62*
	Body fat percentage	*ZC3H15, CMTM4, ZNF443*
	Cholesterol	*KYAT3, NDUFA2, SAMD4A, FNTB, TMEM97, SIN3B*
	Diabetes	*IMMP2L*
	Glucose	*NBPF19, IMMP2L*
	HDL	*CAMK1D, FNTB, PLEKHG4, C22orf39*
	Triglycerides	*NDUFA2, SAMD4A, SIN3B*
	Waist-hip ratio, BMI-adjusted	*TTC4, GBP1, MOV10, ADAM15, NDUFS1, SLC35G2, RPS23, CCDC69, POLR2J3, TMEM168, GTF2E2, CREM, DDB2, ITGB3, CD37*
Cardiovascular	Diastolic blood pressure	*ANKRD36B, ING2, SHARPIN, PAPPA, TMEM106C, SAMD4A, SIPA1L1, ZNF431, APMAP*
	Heart attack	*RO60, ATPAF2*
	Hypertension	*SPATS2, CMTM4, ATPAF2, ITGB3, CBX4, APMAP, C22orf39*
Neonatal/childhood traits	Age of asthma diagnosis	*B3GNT9*
	Age of diabetes diagnosis	*ATPAF2*
	Childhood BMI	*EXOSC10, AP3M1*
	Childhood-onset asthma	*SAR1B, RNF146, RPS25*
	Fetal birthweight	*DUSP12, UBA3, FAM114A1, CMTM4*
	Gestational weight gain	*ISG15, GBP1, CXCL1, GFM2, HARS2*
	Head circumference	*PLA2G2A, LAMTOR5, PRRC2A, DENND1A, CPXM2*
	Late puberty growth	*HP1BP3, ZNF264*
	Pubertal growth start	*PRPF31, CSNK1G2, LZTR1*
	Tanner scale	*CXCL1, TPRN*
	Total puberty growth	*PAN3, NSRP1, ZNF750, PMM1*
Neuropsychiatric	Alcohol dependence	*ZNF134*
	Anorexia nervosa	*DPYD, RO60, AAK1, MRPS27, PPP3CB-AS1, NPLOC4*
	Autism spectrum disorder	*ZC3H15*
	Bipolar disorder	*COQ10B, SEC11A*

Fluid intelligence score	*ECSCR, IDI1, ACER3*
IQ	*FAM228B, HMGN3, DDB2, NAPA, GATD3A*
Schizophrenia	*RNU6-469P, IDI1, SIAE, VPS29, SNUPN, SEC11A*
Tourette's syndrome	*SLC26A7*

Table 2: Susceptibility genes associated with multiple traits. TWAS gene, chromosomal location, and associated trait are provided with genetic correlations between traits at SNP level are provided if significant FDR-adjusted $P < 0.05$.

Gene (Location)	Traits	Genetic Correlation
AP3M1 (10:74120256-74151085)	BMI, Childhood BMI	0.68
APMAP (20:24962943-24992789)	Diastolic blood pressure, Hypertension	0.80
ATPAF2 (17:18018019-18039166)	Heart attack, Hypertension, Age of diabetes diagnosis	HA/HT: 0.53
C22orf39 (22:19443149-19448232)	Cholesterol, HDL, Triglycerides, Hypertension	All pairs significantly correlated
C2orf92 (2:97669742-97703066)	HDL, IQ	0.11
CLIP1 (12:122271433-122422569)	Body fat percentage, HDL	-0.38
CMTM4 (16:66614749-66696707)	Body fat percentage, Hypertension, Fetal birthweight	All pairs significantly correlated, except BFP/FB
COQ10B (2:197453422-197475309)	Bipolar disorder, Schizophrenia	0.73
CSNK1G2 (19:1941161-1981337)	Waist-hip ratio, Pubertal growth start	No significant correlation
CXCL1 (4:73869391-73871302)	Gestational weight gain, Tanner Scale	No significant correlation
DDB2 (11:47214941-47239218)	Waist-hip ratio, Diastolic blood pressure, IQ	WHR/DBP: -0.10; WHR/IQ: 0.13
DENND1A (9:123401590-123930138)	BMI, Head circumference	0.12
ECSCR (5:139448553-139462743)	Body fat percentage, Fluid intelligence score	-0.20
EPB41L1 (20:36092709-36232799)	Cholesterol, Triglycerides	0.95
ERP29 (12:112013347-112023220)	Cholesterol, Triglycerides, Hypertension, Fetal birthweight	All pairs significantly correlated, expect those with FB
FNTB (14:64986788-65062652)	BMI, Cholesterol, HDL	All pairs significantly correlated
GBP1 (1:89052303-89065360)	Waist-hip ratio, Gestational weight gain	No significant correlation
GFM2 (5:74736343-74767371)	BMI, Gestational weight gain	No significant correlation
GTF2E2 (8:30578317-30658241)	BMI, Waist-hip ratio	0.04
HMGN3 (6:79201244-79234738)	Diastolic blood pressure, IQ	-0.08
IDI1 (10:1039419-1056716)	Body fat percentage, HDL, Triglycerides, Fluid intelligence score, Schizophrenia	All pairs significantly correlated
IMMP2L (7:111480816-111562517)	BMI, Diabetes, Glucose, Bipolar disorder	All pairs significantly correlated, expect those with BPD
ITGB3 (17:47253827-47313743)	Waist-hip ratio, Hypertension	-0.13
NAPA (19:47487633-47515258)	Diastolic blood pressure, IQ	-0.08
NBPF19 (1:149475897-149556361)	Adult-onset asthma, Glucose	0.08
NDUFA2 (5:140645362-140647785)	Cholesterol, Triglycerides	0.95
NDUFS1 (2:206123078-206159519)	BMI, Waist-hip ratio	0.04
NSRP1 (17:30116806-30186475)	BMI, Total puberty growth	-0.30

NUSAP1 (15:41332693-41381050)	Allergic disease, Childhood-onset asthma	0.87
PLA2G2A (1:19975430-19980439)	Allergic disease, Head circumference	No significant correlation
PNPO (17:47941522-47949308)	Cholesterol, Triglycerides	0.95
PPP3CB-AS1 (10:73495525-73520070)	Adult-onset asthma, Anorexia nervosa	0.22
PUM2 (2:20248690-20350850)	Cholesterol, Triglycerides	0.95
RAB7A (3:128726135-128814798)	Adult-onset asthma, Childhood-onset asthma	0.88
RO60 (1:193059421-193085985)	Heart attack, Anorexia nervosa	No significant correlation
RPL6 (12:112405180-112409641)	Cholesterol, Triglycerides, Hypertension	All pairs significantly correlated
RPS23 (5:82273319-82278416)	BMI, Waist-hip ratio	0.04
RPS25 (11:119015712-119018347)	BMI, Body fat percentage, HDL, Childhood-onset asthma	All pairs significantly correlated, expect those with COA
SAMD4A (14:54567611-54793315)	Body fat percentage, Cholesterol, HDL, Triglycerides, Diastolic blood pressure	All pairs significantly correlated
SAMD8 (10:75111634-75182123)	BMI, Waist-hip ratio	0.04
SAR1B (5:134601148-134632843)	Adult-onset asthma, Childhood-onset asthma	0.88
SEC11A (15:84669536-84716460)	Adult-onset asthma, Body fat percentage, Bipolar disorder, Schizophrenia	AOA/BFP: 0.16; AOA/BPD: 0.15; BFP/BPD: -0.08; BPD/SCZ: 0.73
SENP3 (17:7561991-7571969)	BMI, Waist-hip ratio	0.04
SERPING1 (11:57597553-57614853)	HDL, Hypertension	-0.26
SHMT2 (12:57229573-57234935)	HDL, IQ	0.11
SIN3B (19:16829386-16880353)	Cholesterol, Triglycerides	0.95
SLC35G2 (3:136819018-136855892)	BMI, Waist-hip ratio	0.04
SNUPN (15:75598082-75626105)	IQ, Schizophrenia	-0.20
TMEM97 (17:28319094-28328685)	Cholesterol, Triglycerides	0.95
ZC3H15 (2:186486157-186509360)	Adult-onset asthma, Body fat percentage, Autism spectrum disorder	AOA/BFP: 0.16

SUPPLEMENTAL FIGURE LEGENDS

Figure S1: SNP heritability of 40 traits. Estimates of SNP heritability with 95% confidence interval (X-axis), grouped and colored by trait category (Y-axis).

Figure S2: SNP-based genetic correlation between 40 traits. Heatmap of estimates of SNP-based genetic correlated between traits, grouped and colored by trait category. Correlations are marked with an asterisk are significantly non-zero with FDR-adjusted $P < 0.05$.

Figure S3: **Example of a biological mechanism MOSTWAS leverages in its predictive models**. Here, assume a SNP (in green) within a regulatory element affects the transcription of gene X (A) or the hyper- or hypomethylation of a CpG island upstream of gene X (B) that codes for a transcription factor or a microRNA hairpin. Transcription factor or microRNA X then binds to a distal regulatory region and affects the transcription of gene G. The association between the expression of gene X and gene G is leveraged in the first step of MeTWAS. A distal-eQTL association is also conferred between this distal-SNP and the eGene G, which is leveraged in the DePMA training process.

Figure S4: TWAS Miami plots for autoimmune/autoreactive disorders. Weighted Z-scores for TWAS associations (Y-axis) over genomic location of genes (X-axis). Red lines show Z-scores corresponding to P < 2.5×10^{-6}. Genes labelled have P < 2.5×10^{-6}, nominal permutation P <0.05, and genes in green showed Benjamini-Hochberg FDR-adjusted P < 0.05 for the distal-SNPs added-last test.

Figure S5: TWAS Miami plots for cardiovascular disorders. Weighted Z-scores for TWAS associations (Y-axis) over genomic location of genes (X-axis). Red lines show Z-scores corresponding to P < 2.5×10^{-6}. Genes labelled have P < 2.5×10^{-6}, nominal permutation P <0.05, and genes in green showed Benjamini-Hochberg FDR-adjusted P < 0.05 for the distal-SNPs added-last test.

Figure S6: TWAS Miami plots for neonatal/childhood outcomes. Weighted Z-scores for TWAS associations (Y-axis) over genomic location of genes (X-axis). Red lines show Z-scores corresponding to

P < 2.5×10^{-6}. Genes labelled have P < 2.5×10^{-6}, nominal permutation P <0.05, and genes in green showed Benjamini-Hochberg FDR-adjusted P < 0.05 for the distal-SNPs added-last test.

Figure S7: TWAS Miami plots for neuropsychiatric outcomes. Weighted Z-scores for TWAS associations (Y-axis) over genomic location of genes (X-axis). Red lines show Z-scores corresponding to P < 2.5×10^{-6}. Genes labelled have P < 2.5×10^{-6}, nominal permutation P <0.05, and genes in green showed Benjamini-Hochberg FDR-adjusted P < 0.05 for the distal-SNPs added-last test.

Figure S8: TWAS Miami plots for BMI and BMI-adjusted waist-hip ratio. Weighted Z-scores for TWAS associations (Y-axis) over genomic location of genes (X-axis). Red lines show Z-scores corresponding to P < 2.5×10^{-6}. Genes labelled have P < 2.5×10^{-6}, nominal permutation P <0.05, and genes in green showed Benjamini-Hochberg FDR-adjusted P < 0.05 for the distal-SNPs added-last test.

Figure S9: TWAS Miami plots for body size/metabolic traits, excluding BMI and BMI-adjusted waist-hip ratio. Weighted Z-scores for TWAS associations (Y-axis) over genomic location of genes (X-axis). Red lines show Z-scores corresponding to P < 2.5×10^{-6}. Genes labelled have P < 2.5×10^{-6}, nominal permutation P <0.05, and genes in green showed Benjamini-Hochberg FDR-adjusted P < 0.05 for the distal-SNPs added-last test.

Figure S10: Placental expression-mediated genetic heritability of traits. Caterpillar plot of placental expression-mediated genetic heritability of traits, colored by trait category. Wald-type 95% confidence intervals are provided for reference. Trait is labelled if the confidence interval does not intersect the null of $\mathrm{h}^2_{\mathrm{GE}} = 0$.

Figure S11: Comparison of GWAS and TWAS results across all 40 traits. Scatterplot of number of TWAS-significant genes (Y-axis) and number of GWAS-significant SNPs (X-axis) across all 40 traits, colored by category of the trait. The size of the point shows the $\log_{10}$ sample size of the GWAS. The red

line and gray band provide a regression line and 95% confidence band for the fitted values. Points are labelled if the point falls outside the confidence band.

Figure S12: Heatmap of genetic correlations on the heritable gene expression level between 40 traits considered in TWAS analysis. Genetic correlations between traits at the level of the predicted expression of heritable genes. Correlations at FDR-adjusted $P < 0.05$ are marked with an asterisk. Autoimmune/autoreactive traits are colored in yellow, body size/metabolic in purple, cardiovascular in green, neonatal/childhood outcomes in blue, and neuropsychiatric in red.

Figure S13: Miami plot of representative phenome-wide scans of GTAs in UKBB. Weighted burden Z-score (Y-axis) of GTA across all traits (X-axis), grouped and colored by ICD code block.

Figure S14: Heatmap of correlations between select transcription factor and TWAS-identified genes in RICHS. Correlations between the RICHS expression of RPs (Y-axis) and associated TWAS genes identified by MOSTWAS in ELGAN (X-axis).

Figure S15: Over-representation enrichments of differentially expressed genes in *EPS15* knockdown. Enrichment plot of over-representation of biological process, cellular component, and molecular function ontologies (Y-axis) with $-\log_{10}$ FDR-adjusted P-value (X-axis). The size of the point gives the relative enrichment ratio for the given pathway.

SUPPLEMENTAL TABLE LEGENDS

Table S1: **Overview of 40 traits and GWAS consider in analysis**. The consortium, trait category, trait, URL for summary statistics, sample size, number of cases (if binary trait), SNP heritability estimate and standard error, Lambda GC, mean χ2 statistic, reference DOI for GWAS, and expression mediated heritability and standard errors (using all and all TWAS-significant genes) are provided in order.

Table S2: Comparison of GWAS and TWAS associations. The category, trait, GWAS sample size, number of cases, number of significant GWAS SNPs (P < 5×10^{-8}), and number of significant total and GWAS-overlapping TWAS associations (P < 2.5×10^{-6}) are provided in order.

Table S3: Genetic correlations between traits at SNP- and placenta-expression mediated levels. Genetic correlations, standard errors, Z-test statistic, P-value, FDR-adjusted P-value, and genetic covariance and standard errors are provided for all pairs of traits.

Table S4: Demographic and clinical covariates summary statistics of ELGAN and RICHS samples.

Table S5: Summary of in- and out-sample predictive performance of MOSTWAS placental expression models. Mean, standard deviation, 25% quantile, median, and 75% quantile of gene expression heritability, in-sample cross-validation R^2 in ELGAN, and out-sample R^2 in RICHS.

Table S6: Summary of 248 significant TWAS gene-trait associations. For each gene and trait, the trait category, chromosomal position of the gene, expression heritability and associated likelihood ratio test P-value, cross-validation predictive performance for gene model, TWAS Z-score and P-value, permutation P-value, top SNP and P-value in GWAS among SNPs used in the gene model, distal Z-score and P-value, and identified mediators are provided, in order.

Table S7: Over-representation analysis of TWAS genes. Biological process, molecular function, and PANTHER pathway ontologies enriched for TWAS-identified genes associated with each trait at FDR-adjusted P < 0.05.

Table S8: Genetic correlations between traits at placental expression-mediated level. For each pair of traits, the genetic correlation, standard error, t-statistic and associated degrees of freedom and P-value is provided.

Table S9: Results of fine-mapping of overlapped TWAS genes using FOCUS. Overlapping genes are provided, with the associated trait, chromosomal positions, TWAS Z-scores, P-values, top GWAS SNP information, posterior inclusion probability, and whether they are included in the credible set for the region. The distal Z-score is also provided.

Table S10: Results of ELGAN phenome-wide scan of neonatal outcomes. For each gene and ELGAN phenotype, the effect size, standard error, adjusted 95% confidence interval, Z-score, P-value, and FDR-adjusted P-value are provided.

Table S11: Cis-GReX correlations of TWAS-identified genes with metabolic traits in the Hybrid Mouse Diversity Panel. For each correlation at FDR-adjusted P < 0.10, the dataset, gene (mouse analog), trait, correlation, and P-value are provided.

Table S12: Over-representation analysis of transcription factors identified as mediators. For the transcription-factor encoding genes identified as mediators, functional categories, ontologies, FDR-adjusted P-value of enrichment, number of overlapping genes in the ontology, and the total number of genes in the ontology is given.

Table S13: Trans-eQTL scan using GBAT in RICHS between genetic loci local to MOSTWAS-identified transcription factors and the expression of the target TWAS gene. The effect size, P-value, and FDR-adjusted P-value are provided.

Table S14: Results from MR-Egger to assess causal effects of transcription factors on targeted TWAS genes. For each RP-TWAS pair, the causal estimate, confidence interval, P-value, residual standard error, heterogeneity statistic, and heterogeneity P-value are provided.

Table S15: MOSTWAS-identified CpG site mediators found within ENCODE-identified placenta cis-regulatory sites. For each CpG site mediator that overlaps with a placental cis-regulatory stie, the chromosomal location of the regulatory site, the classification of the regulatory site, tissue, gestational time, sex, and accession number are provided.

Table S16: Summary statistics of down-regulated differentially expressed genes in EPS15 knockdown cells. For each gene with FDR-adjusted P < 0.01, we provide the gene name, $\log_2$ fold change, standard error, and P-values.

Table S17: Summary statistics of up-regulated differentially expressed genes in *EPS15* knockdown cells. For each gene with FDR-adjusted P < 0.01, we provide the gene name, $\log_2$ fold change, standard error, and P-values.

Table S18: Over-representation analysis of down-regulated genes. Biological process, molecular function, and PANTHER and KEGG pathway ontologies enriched for down-regulated genes in *EPS15* knockdown cells associated with each trait at FDR-adjusted P < 0.05.

Table S19: Over-representation analysis of up-regulated genes. Biological process, molecular function, and PANTHER and KEGG pathway ontologies enriched for up-regulated genes in *EPS15* knockdown cells associated with each trait at FDR-adjusted P < 0.05.