Distal mediator-enriched, placental transcriptome-wide analyses illustrate the Developmental Origins of Health and Disease

Arjun Bhattacharya, PhD[1]*, Anastasia N. Freedman, BA[2], Vennela Avula[2], Rebeca Harris, BS[3], Weifang Liu, BA[4], Robert M. Joseph, PhD[5], Lisa Smeester, PhD[2,6,7], Hadley J. Hartwell, MS[2], Karl C.K. Kuban, MD[8], Carmen J. Marsit, PhD[9], Yun Li, PhD[4,10,11], T. Michael O'Shea, MD, MPH[12], Rebecca C. Fry, PhD[2,6,7]*†, Hudson P. Santos Jr., PhD, RN[3,6]*†

1. Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, USA
2. Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA
3. Biobehavioral Laboratory, School of Nursing, University of North Carolina, Chapel Hill, NC, USA
4. Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA
5. Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, MA, USA
6. Institute for Environmental Health Solutions, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA
7. Curriculum in Toxicology and Environmental Medicine, University of North Carolina, Chapel Hill, NC, USA
8. Department of Pediatrics, Division of Pediatric Neurology, Boston University Medical Center, Boston, MA, USA
9. Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA
10. Department of Genetics, University of North Carolina, Chapel Hill, NC, USA
11. Department of Computer Science, University of North Carolina, Chapel Hill, NC, USA
12. Department of Pediatrics, School of Medicine, University of North Carolina, Chapel Hill, NC, USA

†Equal contribution

*Please direct correspondence to AB (abtbhatt@ucla.edu), RCF (rfry@unc.edu), HPS (hsantosj@email.unc.edu).

ABSTRACT

As the master regulator of the intrauterine environment, the placenta is core to the Developmental Origins of Health and Disease (DOHaD) but is understudied in relation to tissue-specific gene and trait regulation. We performed distal mediator-enriched transcriptome-wide association studies (TWAS) for 40 health traits across 5 physiological categories, using gene expression models trained with multi-omic data from the Extremely Low Gestational Age Newborn Study ($N = 272$). At $P < 2.5 \times 10^{-6}$, we detected 248 gene-trait associations (GTAs) across 176 genes, mostly for metabolic and neonatal traits and enriched for cell growth and immunological pathways. Of these GTAs, 89 showed significant mediation through genetic variants distal to the gene, identifying potential targets for functional validation. Functional validation of a mediator gene (*EPS15*) in human placenta-derived JEG-3 trophoblasts resulted in increased expression of its predicted targets, *SPATA13* and *FAM214A*, both associated with the trait of waist-hip ratio in TWAS. These results illustrate the profound health impacts of placental genetic and genomic regulation in developmental programming across the life course.

KEYWORDS

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

MAIN

The placenta serves as the master regulator of the intrauterine environment via nutrient transfer, metabolism, gas exchange, neuroendocrine signaling, growth hormone production, and immunologic control[1–5]. Due to strong influences on postnatal health, the placenta is central to the Developmental Origins of Health and Disease (DOHaD) hypothesis, which purports that the *in utero* experience has lifelong impacts on child health by altering developmental programming and influencing risk of common, noncommunicable health conditions[6]. For example, placental biology has 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**)[7–10]. Despite its long-lasting influences on health, the placenta has not been well-studied in large consortia studies of multi-tissue gene regulation[11,12]. Studying regulatory mechanisms in the placenta underlying biological processes in developmental programming will 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[13,14]. Furthermore, there is growing evidence that the placental epigenome influences gene regulation, often distally (more than 1-3 Megabases away in the genome)[15], and that placental DNA methylation and microRNA (miRNA) expression are associated with health traits in children[16–18]. Dysfunction of transcription factor regulation in the placenta has also shown profound effects on childhood traits[19–22]. Although combining genetics, transcriptomics, and epigenomics lends insight into the influence of placental genomics on complex traits[23], 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[24–26]. However, traditional methods for TWAS largely overlook genetic variants

distal to genes of interest, ostensibly mediated through regulatory biomarkers (e.g., transcription factors, miRNAs, and DNA methylation sites)[27,28]. Not only may these distal biomarkers explain a significant portion of both gene expression heritability and trait heritability on the tissue-specific expression level[29–32], 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 investigate three broad questions: (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 gene expression, CpG methylation, and miRNA expression data from fetal-side placenta tissue from the Extremely Low Gestational Age Newborn (ELGAN) Cohort Study[33]. We trained predictive models of gene expression, enriched for distal SNPs using MOSTWAS, a recent TWAS extension that integrates multi-omic data[34]. Re-analyzing 40 GWAS of European-ancestry subjects from large consortia[35–39], 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 placental multi-omics. Results from our analysis can be explored at our Shiny R app, the ELGAN DOHaD Atlas: https://elgan-twas.shinyapps.io/dohad/.

RESULTS

Genetic heritability and correlations across traits

From large consortia[35–39], we curated GWAS summary statistics from subjects of European ancestry for 40 complex, non-communicable traits and disorders across five health categories to systematically identify potential links to genetically-regulated placental expression (**Table 1**, **Supplemental Table 1**; **Methods**). These 40 traits comprise of 3 autoimmune/autoreactive disorders, 8 body size/metabolic traits, 4 cardiovascular disorders, 14 neonatal/early childhood traits, and 11 neuropsychiatric traits/disorders

(**Supplemental Table 1**). These five categories of traits have been linked previously to placental biology and morphology[7–10].

To assess the percent variance explained by genetics in each trait and the genetic associations shared between traits, 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 S2**). At Benjamini-Hochberg FDR-adjusted $P < 0.05$, we also observed several 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 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.

Gene expression prediction models

To train predictive models of placental expression, the first step of our TWAS (**Figure 2A**), we leveraged MOSTWAS[34], a recent extension that includes distal variants in transcriptomic prediction. As large proportions of total heritable gene expression are explained by distal-eQTLs local to regulatory hotspots[27,28,30], MOSTWAS uses data-driven approaches to either 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**)[34]. In this analysis, these regulatory biomarkers include transcription-factor encoding genes, miRNAs, and CpG methylation sites from the ELGAN Study (**Methods**).

Using genotypes (from umbilical cord blood)[40], mRNA expression, CpG methylation, and miRNA expression data (from fetal-side placenta)[23] from the ELGAN Study[33] 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 S3**). Out of a total of 12,020 genes expressed across all samples in ELGAN, we successfully built significant models for 2,994 genes, such that SNP-based expression heritability is significantly positive (nominal $P < 0.05$) and five-fold cross-validation (CV) adjusted $R^2 \geq 0.01$ (**Figure 3A [Step 1]**); 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-of-sample validation, we imputed expression into individual-level genotypes from the Rhode Island Child Health Study (RICHS; $N = 149$)[41,42], 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 S4**).

Placental transcriptome-wide association studies

Overall associations and permutation tests

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[25,43], we detected 932 GTAs (spanning 686 unique genes) at $P < 2.5 \times 10^{-6}$ (corresponding to $|Z| > 4.56$), a transcriptome-wide significance threshold consistent with previous TWAS[25,31] (**Figure 3A [Step 2]**, **Supplemental Data**). As many of these loci carry significant signal because of strong trait-associated GWAS architecture, 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[25]. At FDR-adjusted permutation $P < 0.05$ and spanning 176 unique genes, we detected 248 such GTAs, of which 11 were found in autoimmune/autoreactive disorders, 136 in body size/metabolic traits, 32 in cardiovascular disorders, 39 in neonatal/childhood traits, and 30 in neuropsychiatric traits (**Figure 3C [Step 3]**, **Table 1**, **Supplemental Table S5;** Miami plots of TWAS Z-scores in **Supplemental Figures S4-S9**).

For example, the 39 GTAs detected with BMI included *LARS2* (Leucyl-tRNA Synthetase 2, OMIM: 604544) ($Z = 11.4$) and *CAST* (Calpastatin, OMIM: 114090) ($Z = -4.61$). These two GTAs have been detected using *cis*-only TWAS in different tissues[25,31]. In addition, one of the 30 genes identified in association with waist-hip ratio was prioritized in other tissues by TWAS: *NDUFS1* (NADH:Ubiquinone Oxidoreductase Core Subunit S1, OMIM: 157655) ($Z = -5.38$)[31]. 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[10]. Of the 19 birthweight-associated genes they identified, we could only train significant expression models for two in ELGAN: *PLEKHA1* (Pleckstrin Homology Domain Containing A1, OMIM: 607772) and *PSG8* (Pregnancy Specific Beta-1-Glycoprotein 8, OMIM: 176397). 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 from Peng *et al*[10]. We hypothesize that minimal overlap with susceptibility genes identified by Peng *et al* is due to differing eQTL architectures in the datasets and different inclusion criteria for significant gene expression models[10,34,44,45].

We conducted over-representation analysis for biological process, molecular function, and PANTHER gene pathway ontologies for TWAS-detected susceptibility genes (**Supplemental Figure S10**, **Supplemental Table S6**)[46]. 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 (**Supplemental Table S6**). 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 2**).

Genetic heritability and correlations of traits by predicted expression

To assess how, on the whole, 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 RHOGE, an linkage disequilibrium (LD) score regression approach[25,47]. Overall, we found 3/14 neonatal traits (childhood BMI, 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)[31]; none of the 26 traits outside the neonatal category were appreciably explained by placental GReX. **Figure 3D** shows that mean $\hat{h}^2_{GE}$ is higher in neonatal traits than other groups. A comparison of the number of GWAS-significant SNPs and TWAS-significant genes also shows that neonatal/childhood traits are enriched for placental TWAS associations, even though significant genome-wide GWAS architecture cannot be inferred for these traits (**Supplemental Figure S11**). Not only do these results highlight the power advantage of properly aligned tissue-specific TWAS compared to GWAS, they 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.

Similarly, using RHOGE[31], we assessed genetic correlations (r_{GE}) between traits at the level of placental GReX (**Supplemental Figure S12**). We found several known correlations, such as between cholesterol and triglycerides ($\hat{r}_{GE} = 0.99, P = 1.44 \times 10^{-118}$) 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: IQ and diastolic blood pressure ($\hat{\rho}_{GE} = -0.55, P = 2.44 \times 10^{-5}$) and age of asthma diagnosis and 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 gene regulatory contributions[48]. Overall, these correlations may 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 transcription factors, miRNAs, or CpG methylation sites.

Pleiotropy across overlapped genes and multiple traits

GTAs across the 40 traits showed several overlaps in signal. First, as these genes indicate trait association and do not reflect causality, we used FOCUS[49], 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 S8**). For example, we identified 3 genes associated with triglycerides at the 12q24.13 chromosomal region (*ERP29*, *RPL6*, *BRAP*), with *ERP29* (Endoplasmic Reticulum Protein 29, OMIM: 602287) defining the region's 90% credible set with approximately 95% PIP. Similarly, we detected 3 genes associated with BMI at 10q22.2 (*AP3M1*, *SAMD8*, *MRPS16*), with *AP3M1* (Adaptor Related Protein Complex 3 Subunit Mu 1, OMIM: 610366) defining the region's 90% credible set with approximately 99% PIP.

We also noticed that *ERP29* and *RPL6* (Ribosomal Protein L6, OMIM: 603703) 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 3**). Nine genes showed more than 3 GTAs across different categories. For example, *IDI1* (Isopentenyl-Diphosphate Delta Isomerase 1, OMIM: 604055), a gene involved in cholesterol biosynthesis[50], 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 ($Z = -5.56$). A link between cholesterol-related genes and schizophrenia has been detected previously, potentially due to coregulation of myelin-related genes[51,52]. 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[53]; the inclusion of this locus may have contributed to the large TWAS associations. Similarly, *SAMD4A* (Sterile Alpha Motif Domain Containing 4A, OMIM: 610747) also shows associations with 4 body size/metabolic - body fat percentage ($Z = 6.70$), cholesterol ($Z = -6.76$), HDL ($Z = -6.78$), triglycerides ($Z = -5.30$) - and 1 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* (Chemokine-Like Factor Superfamily Member 4, OMIM: 607887), an angiogenesis regulator[54], 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[55], 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 4A**). Using UK Biobank[35] GWAS summary statistics, we conducted TWAS for a variety of traits across 8 groups, defined generally around ICD code blocks (**Figure 4A, Supplemental Figure S13**); here, we grouped metabolic and cardiovascular traits into one category for ease of analysis. At FDR-adjusted $P < 0.05$, *ATPAF2* (ATP Synthase Mitochondrial F1 Complex Assembly Factor 2, OMIM: 608918), *RPL6*, and *SEC11A* (SEC11 Homolog A, Signal Peptidase Complex Subunit, OMIM: 618258) showed GTA enrichments for immune-related traits, *ATAPF2* for neonatal traits, *IDI1* for mental disorders, and *RPS25* (Ribosomal Protein S25, OMIM: 180465) for musculoskeletal traits. Across these 8 trait groups, *RPL6* showed multiple strong associations with circulatory, respiratory, immune-related, and neonatal traits (**Figure 4A**). 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 ($N = 729$) (**Online Methods**). Controlling for race, sex, gestational duration, inflammation of the chorion, and maternal age, as described before and in **Online Methods**[23], we tested for associations for 6 representative traits measured at birth or at 24

months: 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[56]. Shown in **Figure 4B** and **Supplemental Table S9**, at FDR-adjusted $P < 0.05$, we detected negative associations between *SEC11A* GReX and birthweight Z-score (effect size: -0.248, 95% adjusted[57] 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[58], and *SEC11A* has roles in cell migration and invasion[59]), understanding how the placental GReX of these genes affects neonatal traits may elucidate the potential long-lasting impacts of placental dysregulation.

Investigating mediators of distal SNP to gene relationships

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[34], we interrogated distal loci incorporated into expression models for trait associations, beyond the association at the local locus. For 89 of 248 associations, predicted expression from distal SNPs showed significant associations at FDR-adjusted $P < 0.05$ (**Figure 3A [Step 4]**, **Supplemental Table S5**). For each significant distal association, we identified a set of biomarkers that potentially affects transcription of the TWAS gene: a total of 9 transcription factor-encoding genes (TFs) and 163 CpG sites across all 89 distal associations. Particularly, we detected two TFs, *DAB2* (DAB Adaptor Protein 2, OMIM: 601236, distal mediator for *PAPPA* and diastolic blood pressure, distal $Z = -3.98$) and *EPS15* (Epidermal Growth Factor Receptor Substrate 15, OMIM: 600051), both highly expressed in placenta[60,61]. Mediated through *EPS15* (overall distal $Z = 7.11$ and 6.33, respectively), distally predicted expression of *SPATA13* (Spermatogenesis Associated 13, OMIM: 613324) and *FAM214A* (Family With Sequence Similarity 214 Member A) showed association with waist-hip ratio. Interestingly, *EPS15* itself showed a TWAS association for waist-hip ratio (**Supplemental Table S5**), and the direction of the *EPS15* GTA was opposite to those of *SPATA13* and *FAM214A*. Furthermore, *RORA* (RAR Related Orphan Receptor A, OMIM: 600825), a gene encoding a TF involved in inflammatory signaling[62], showed a negative

association with transcription of *UBA3* (Ubiquitin Like Modifier Activating Enzyme 3, OMIM: 603172), a TWAS gene for fetal birthweight. Low placental *RORA* expression was previously shown to be associated with lower birthweight[63]. Aside from functions related to transcription regulation, the 9 TFs (*CUL5, DAB2, ELL, EPS15, RORA, SLC2A4RG, SMARCC1, NFKBIA, ZC3H15*) detected by MOSTWAS were enriched for several ontologies (**Supplemental Table S10**), namely catabolic and metabolic processes, response to lipids, and multiple nucleic acid-binding processes[46].

As we observed strong correlations between expressions of TF-TWAS gene pairs in ELGAN (**Supplemental Figure S14**), we then examined the associations between TWAS-identified genes and the locus around any predicted mediating TFs in an external dataset[42]. Using RICHS, we conducted a gene-based *trans*-eQTL scan using Liu *et al*'s GBAT method[64] to computationally validate TF-TWAS gene associations. We predicted GReX of the TFs using *cis*-variants through leave-one-out cross-validation[65] and scanned for associations with the respective TWAS genes (**Figure 4C**, **Supplemental Table S11**). 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* (NF-Kappa-B Inhibitor Alpha, OMIM: 164008) and *HNRNPU* (Heterogeneous Nuclear Ribonucleoprotein U, OMIM: 602869) (effect size -0.26, FDR-adjusted $P = 1.9 \times 10^{-4}$). We also considered an Egger regression-based Mendelian randomization framework[66] in RICHS to estimate the causal effects of TFs on the associated TWAS genes (**Methods**) using, as instrumental variables, *cis*-SNPs correlated to the TF and uncorrelated with the TWAS genes. We estimated significant causal effects for two TF-TWAS gene pairs (**Figure 4C**, **Supplemental Table S12**): *EPS15* on *FAM214A* (causal effect estimate -0.58; 95% CI [0.21, 0.94]) and *RORA* on *UBA3* (0.58; [0.20, 0.96]).

We also examined 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[12], identifying 34 CpG sites (mediating 29 distinct TWAS genes) that fall in *cis*-regulatory regions (**Supplemental Table S13**). 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 transcription factor predicted to mediate genetic regulation of *FAM214A*. Furthermore, expression of *LARS2*, a TWAS gene for 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[67] and has shown BMI TWAS associations in other tissues[25,31]. Results from these external datasets add more evidence that these mediators play a role in gene regulation of these TWAS-identified genes.

In-vitro assays of transcription factor activity

Based on our computational results, we experimentally investigated whether the inverse relationship between TF *EPS15* and its two prioritized target TWAS genes, *SPATA13* and *FAM214A*, is supported *in vitro*. We employed gene silencing technology to knock down *EPS15* expression in human placenta-derived JEG-3 trophoblast cells and assessed the gene expression of the targets via qRT-PCR. We used a FANA oligonucleotide targeting *EPS15* for specific silencing and compared the knockdown variant with scramble oligo and no-addition controls. Addition of FANA-EPS15 to JEG-3 cells decreased *EPS15* gene expression, while increasing the expression of *SPATA13* and *FAM214* (**Figure 5A**). *EPS15* expression demonstrated a 50% decrease, while *SPATA13* and *FAM214A* demonstrated increases in expression of 795% and 377%, respectively. At FDR-adjusted $P < 0.10$, though not statistically significant against the control, 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 transcription factor and target mRNA were statistically significant (**Figure 5B**). These results support our computational findings that increased *EPS15* expression may decrease expression of *SPATA13* and *FAM214A*. Though we could not study the effects of these three genes on body size-related traits, this study prioritizes *EPS15* as a potential regulator for multiple genes downstream, perhaps for genes affecting cell adhesion and growth in the placenta, like *SPATA13*[68–70].

DISCUSSION

The placenta has historically been understudied in large multi-tissue consortia efforts that study tissue-specific regulatory mechanisms[11,12]. To address this gap, we systematically categorized placental gene-

trait associations relevant to the DOHaD hypothesis using distal mediator-enriched TWAS and deployed these results at the ELGAN DOHaD Atlas (https://elgan-twas.shinyapps.io/dohad/). By integrating multi-omic data from the ELGAN Study[33] with 40 GWAS, we detected 176 unique 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. 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 at-birth traits in ELGAN and found associations with neonatal body size and infant cognitive development. Furthermore, we could only estimate significantly positive placental GReX-mediated heritability for neonatal traits but not for later-in-life traits. 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 carry over later-in-life or as etiologic antecedents for complex traits.

MOSTWAS also allows for hypothesis generation for regulation of TWAS-detected genes, through distal mediating biomarkers, like transcription factors, miRNAs, or products downstream of CpG methylation islands[34]. Our computational results prioritized 89 GTAs with strong distal associations. We interrogated one such functional hypothesis: *EPS15*, a transcription factor-encoding gene in the EGFR pathway[1], regulates two TWAS genes associated with waist-hip ratio - *FAM214A*, a gene of unknown function, and *SPATA13*, a gene that regulates cell migration and adhesion[68–70]. In placenta-derived trophoblasts, knockdown of *EPS15* showed increased expression of both *FAM214A* and *SPATA13*. *FAM214A* and *SPATA13* both showed positive associations with adult waist-hip ratio. In particularly, *EPS15*, mainly involved in endocytosis, is a maternally imprinted gene and thus predicted to promote offspring health[61,71–73]; its 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, but this *in vitro* assay shows 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, although TWAS is unlikely to be subject to reverse-causality (trait cannot affect expression, independent of genetics), instances of horizontal SNP pleiotropy, where SNPs influence the trait and expression independently, were not examined here. 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[74,75]. However, significant TWAS genes did not show associations for gestational duration, suggesting minimal bias from this collider effect. An interesting future endeavor could include negative controls to account for unmeasured confounders in predictive models[76] to allow for more generalizability of predictive models. Fourth, though we did scan neonatal traits in ELGAN using individual-level genotypes, the sample size is small; larger GWAS with longitudinal traits could allow for rigorous Mendelian randomization studies[77] that investigate relationships between traits across the life course, in the context of placental regulation. 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. 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 consider the placenta as a core tissue for learning about the developmental origins of trait and disease etiology.

ONLINE METHODS

Data acquisition and quality control

Genotypes and multi-omic (mRNA, miRNA, and CpG methylation) data were collected from umbilical cord blood and the fetal side of the placenta of subjects enrolled in the ELGAN[33] Study, as described in

previous work[23,40,78] and in detail in **Supplemental Methods**. Genotype data was assayed on Illumina 1 Million Quad and HumanOmniExpress-12 v1.0 arrays[79]. We removed SNPs with call rate < 90% and MAF < 1% and samples with missing rate > 10%[80]. We imputed genotypes to the TOPMed Freeze 5 reference panel[81] using eagle for phasing and minimac4 for imputation[82,83] and considered only autosomal variants. This resulted in 6,567,190 total variants. Quality control for CpG methylation was performed at the sample level, excluding samples that failed and technical duplicates. The *ComBat* function was used from the sva package to adjust for batch effects from sample plate and cell-type heterogeneity[84]. mRNA and miRNA were aligned to the GENCODE Release 31 reference transcriptome and quantified using Salmon[85] and the HTG EdgeSeq System[86]. We upper-quartile normalized distributional differences between lanes[87] and used RUVSeq and limma to estimate and remove unwanted variation[88,89]. Overall, we considered 846,233 CpG sites, 12,020 genes, and 1,898 miRNAs. We downloaded quality-controlled genotypes and obtained normalized RNA-seq data for RICHS data for validation of gene expression models[42]. Summary statistics were downloaded from the following consortia: the UK Biobank[35], Early Growth Genetics Consortium[36], Genetic Investigation of Anthropometric Traits[37], Psychiatric Genomics Consortium[38], and the Complex Trait Genetics Lab[39] (**Supplemental Table 1**). Genomic coordinates were transformed to the hg38 reference genome using liftOver[90,91].

Estimation of SNP heritability of gene estimation

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[92]. Analysis was conducted using GCTA v1.93.1[93]. 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[34] (**Supplemental Figure S1**, **Supplemental Methods**). Briefly, 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 **Supplemental Methods**. We considered only genes with significantly positive heritability at nominal $P < 0.05$ using a likelihood ratio test and five-fold cross-validation $R^2 \geq 0.01$.

TWAS tests of association

The association between predicted expression and traits was assessed in GWAS summary statistics using the weighted burden test and 1000 Genomes Project CEU population as an LD reference[25,34,43,94] with Bonferroni-corrected significance threshold of $P < 2.5 \times 10^{-6}$. We only consider GWAS of subjects from European ancestry, as ELGAN data does not have a large enough sample size of non-Europeans to accurately map distal-eQTLs. In individual level data from ELGAN, we multiplied the genotype matrix by the SNP-gene weights to construct imputed expression in ELGAN; for samples used in model training, we used the cross-validated predicted expression. We tested the significance of expression-trait associations conditional on SNP-trait effects at a locus using the permutation test from Gusev *et al*[25]. We also tested the trait association at distal variants using the added last test from MOSTWAS[34]. Briefly, we computed

the weighted Z-score at the distal loci, conditional on the weighted Z-score at the local locus and test this using the same null distribution assumptions as in the weighted burden test from Gusev *et al*[25]. These tests are explained in detail in **Supplemental Methods**.

Genetic heritability and correlation estimation

At the genome-wide level, we estimated genetic heritability and correlation between traits using LD score regression[47,95] using GWAS summary statistics. 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 level. Briefly, we assume that the expected χ^2 statistic under a complex trait is a linear function of the LD score; the effect size of the LD score on the χ^2 is proportional to h^2_{GE}. We estimated the LD scores of each gene by predicted expression in European samples of 100 Genomes and computing the sample correlations and inferred h^2_{GE} using ordinary least squares. We employed RHOGE[31] to estimate and test for significant genetic correlations between traits at the predicted expression level (details in **Supplemental Methods**).

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[23,78,96] because of their strong correlations with the outcomes and with placental transcriptomics and methylomics.

Validation analyses in RICHS

Using genotype and RNA-seq expression data from RICHS[42], we attempted to validate TF-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*[64] to assess associations between the loci around TFs and the expression of TWAS genes in RICHS. GBAT tests the association between the predicted expression of a TF with the expression of a TWAS gene, improving power of trans-eQTL mapping[65]. We also conduct directional Egger regression-based Mendelian randomization to estimate and test the causal effects of the expression of the TF on the expression of the TWAS gene[97].

In-vitro functional assays

Cell culture and treatment

The JEG-3 immortalized trophoblast cell was 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.

Assessment of mRNA expression by quantitative Real-Time Polymerase Chain Reaction (qRT-PCR)

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[98].

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.

ACKNOWLEDGEMENTS

We thank Michael Love, Kanishka Patel, Michael Gandal, Chloe Yap, Bogdan Pasaniuc, and Jon Huang for engaging conversation during the research process. 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.

ETHICS DECLARATIONS

The study was approved by the Institutional Review Board of the University of North Carolina at Chapel Hill. All participants consented to the study as per IRB protocol.

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 Institute of Nursing Research (K23NR017898; R01NR019245), and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01HD092374; R03HD101413; P50HD103573). The funders had no role in study design or analysis.

WEB RESOURCES

UK Biobank summary statistics: http://www.nealelab.is/uk-biobank

GIANT Consortium summary statistics: https://portals.broadinstitute.org/collaboration/giant/index.php/Main_Page

PGC summary statistics: https://www.med.unc.edu/pgc/download-results/

EGG summary statistics: https://egg-consortium.org/

CTG Lab summary statistics: https://ctg.cncr.nl/software/

RICHS eQTL dataset: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001586.v1.p1

ENCODE placental epigenomic annotations: https://www.encodeproject.org/ (accession numbers in **Supplemental Table S13**)

MOSTWAS software: https://bhattacharya-a-bt.github.io/MOSTWAS/articles/MOSTWAS_vignette.html

ELGAN DOHaD Atlas: https://elgan-twas.shinyapps.io/dohad/

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. All models and full TWAS results are found at https://doi.org/10.5281/zenodo.4618036[99].

AUTHOR CONTRIBUTIONS

AB, RCF, and HPS designed this study. AB, VA, WL, and YL supervised and conducted computational analyses. ANF, HJH, RCF, and HPS supervised and conducted experiments. RMJ, RH, LS, KCKK, TMO,

CJM, RCF, and HPS collected the data. AB, TMO, RH, RCF, and HPS interpreted results. AB, RCF, and HPS wrote the paper. All authors read and edited the paper.

REFERENCES

1. McKay, R. Developmental biology: Remarkable role for the placenta. *Nature* vol. 472 298–299 (2011).
2. Hirschmugl, B. *et al.* Maternal obesity modulates intracellular lipid turnover in the human term placenta. *Int. J. Obes.* **41**, 317–323 (2017).
3. Shrestha, D., Ouidir, M., Workalemahu, T., Zeng, X. & Tekola-Ayele, F. Placental DNA methylation changes associated with maternal prepregnancy BMI and gestational weight gain. *Int. J. Obes.* 1–11 (2020) doi:10.1038/s41366-020-0546-2.
4. Baron-Cohen, S. *et al.* Foetal oestrogens and autism. *Mol. Psychiatry* 1–9 (2019) doi:10.1038/s41380-019-0454-9.
5. Thornburg, K. L., O'Tierney, P. F. & Louey, S. The Placenta is a Programming Agent for Cardiovascular Disease. *Placenta* **31**, S54 (2010).
6. Gillman, M. W. Developmental origins of health and disease. *New England Journal of Medicine* vol. 353 1848–1850 (2005).
7. Ursini, G. *et al.* Convergence of placenta biology and genetic risk for schizophrenia article. *Nat. Med.* **24**, 792–801 (2018).
8. 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).
9. Bronson, S. L. & Bale, T. L. The Placenta as a Mediator of Stress Effects on Neurodevelopmental Reprogramming. *Neuropsychopharmacology* vol. 41 207–218 (2016).
10. Peng, S. *et al.* Genetic regulation of the placental transcriptome underlies birth weight and risk of childhood obesity. *PLOS Genet.* **14**, e1007799 (2018).
11. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (80-.).* **369**, 1318–1330 (2020).
12. Abascal, F. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes.

Nature **583**, 699–710 (2020).

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

14. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *bioRxiv* 787903 (2019) doi:10.1101/787903.

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

16. Lesseur, C., Paquette, A. G. & Marsit, C. J. Epigenetic Regulation of Infant Neurobehavioral Outcomes. *Med. epigenetics* **2**, 71–79 (2014).

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

18. Maccani, M. A., Padbury, J. F., Lester, B. M., Knopik, V. S. & Marsit, C. J. Placental miRNA expression profiles are associated with measures of infant neurobehavioral outcomes. *Pediatr. Res.* **74**, 272–278 (2013).

19. Meinhardt, G. *et al.* Pivotal role of the transcriptional co-activator YAP in trophoblast stemness of the developing human placenta. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 13562–13570 (2020).

20. 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).

21. Tilley, S. K. *et al.* Placental CpG methylation of infants born extremely preterm predicts cognitive impairment later in life. *PLoS One* **13**, e0193271 (2018).

22. Iglesias-Platas, I. *et al.* Altered expression of the imprinted transcription factor PLAGL1 deregulates a network of genes in the human IUGR placenta. *Hum. Mol. Genet.* **23**, 6275–6285 (2014).

23. 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).

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

25. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies.

Nat. Genet. **48**, 245–252 (2016).

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

27. 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).

28. 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).

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

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

31. 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).

32. Sinnott-Armstrong, N., Naqvi, S., Rivas, M. & Pritchard, J. K. GWAS of three molecular traits highlights core genes and pathways alongside a highly polygenic background. *Elife* **10**, (2021).

33. 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).

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

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

36. 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).

37. 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).

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

39. 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).

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

41. 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).

42. 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).

43. Pasaniuc, B. *et al.* Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics* **30**, 2906–2914 (2014).

44. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1–20 (2018).

45. Fryett, J. J., Inshaw, J., Morris, A. P. & Cordell, H. J. Comparison of methods for transcriptome imputation through application to two common complex diseases. *Eur. J. Hum. Genet.* **26**, 1658–1667 (2018).

46. 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).

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

48. 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).

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

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

51. 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).

52. Prabakaran, S. *et al.* Mitochondrial dysfunction in schizophrenia: Evidence for compromised brain metabolism and oxidative stress. *Mol. Psychiatry* **9**, 684–697 (2004).

53. 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).

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

55. 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).

56. O'Shea, T. M. *et al.* Accuracy of the Bayley-II mental development index at 2 years as a predictor of cognitive impairment at school age among children born extremely preterm. *J. Perinatol.* **38**, 908–916 (2018).

57. Benjamini, Y. *et al.* False discovery rate-adjusted multiple confidence intervals for selected parameters. *Journal of the American Statistical Association* vol. 100 71–81 (2005).

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

59. 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).

60. 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).

61. 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).

62. 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).

63. 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).

64. Liu, X. *et al.* GBAT: a gene-based association test for robust detection of trans-gene regulation.

Genome Biol. **21**, 211 (2020).

65. 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).

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

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

68. 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).

69. 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).

70. Kawasaki, Y. *et al.* Identification and characterization of Asef2, a guanine-nucleotide exchange factor specific for Rac1 and Cdc42. *Oncogene* **26**, 7620–7627 (2007).

71. 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).

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

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

74. 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).

75. Mitchell, R. E., Paternoster, L. & Davey Smith, G. Mendelian Randomization in Case Only Studies: A Promising Approach to be Applied With Caution. *Am. J. Cardiol.* **122**, 2169–2171 (2018).

76. Sanderson, E., Macdonald-Wallis, C. & Davey Smith, G. Negative control exposure studies in the presence of measurement error: implications for attempted effect estimate calibration. *Int. J. Epidemiol.* **47**, 587–596 (2018).

77. Smith, G. D. & Ebrahim, S. Mendelian randomization: prospects, potentials, and limitations. *Int. J.*

Epidemiol. **33**, 30–42 (2004).

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

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

80. 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).

81. 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).

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

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

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

85. 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).

86. 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).

87. 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).

88. 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).

89. 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).

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

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

92. 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).

93. 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).

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

95. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

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

97. 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).

98. 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).

99. 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.

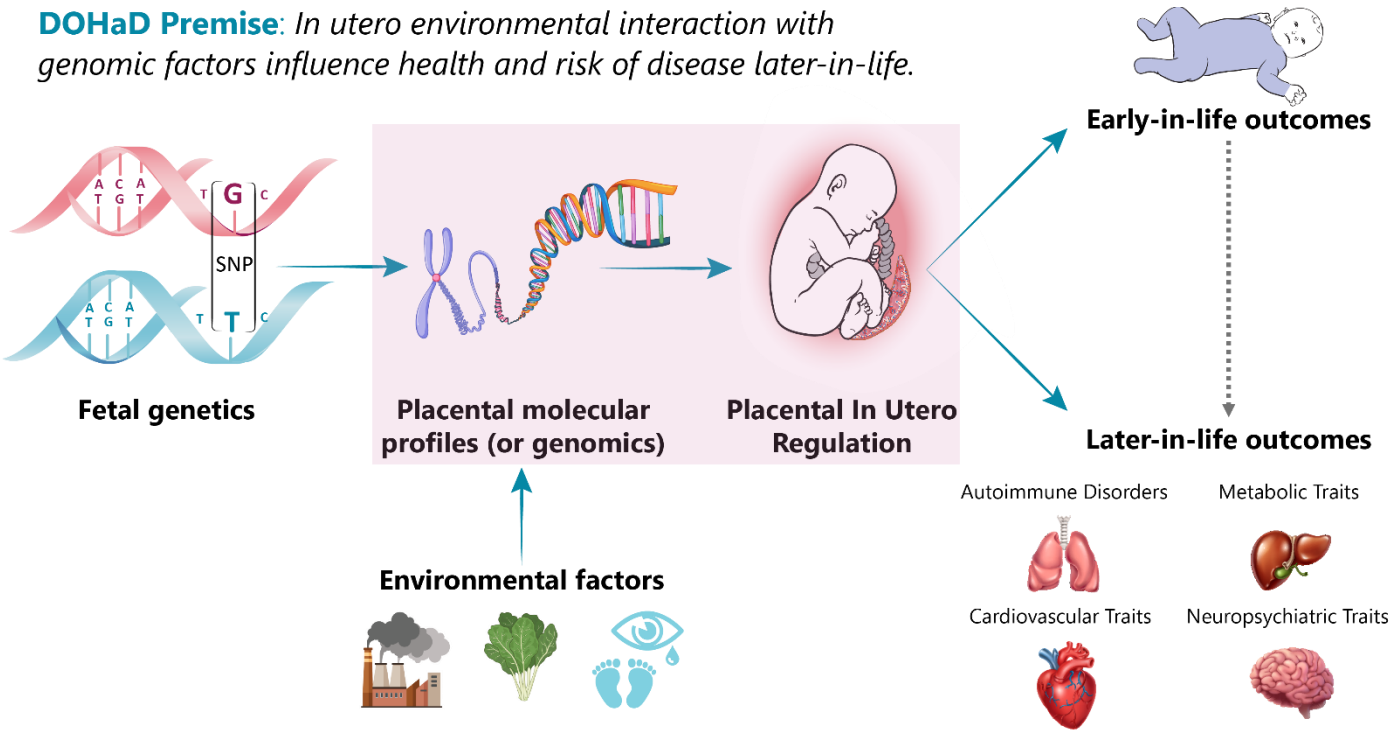

Figure 1: ***Overview of the placenta and the DOHaD Hypothesis*.** The placenta facilitates many important functions *in utero*, including nutrient transfer, metabolism, gas exchange, neuroendocrine signaling, growth hormone production, and immunologic control. As such, it is known as a master regulator of the intrauterine environment and is core to the Developmental Origins of Health and Disease (DOHaD) hypothesis. Placental genomic regulation is purported to be 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.

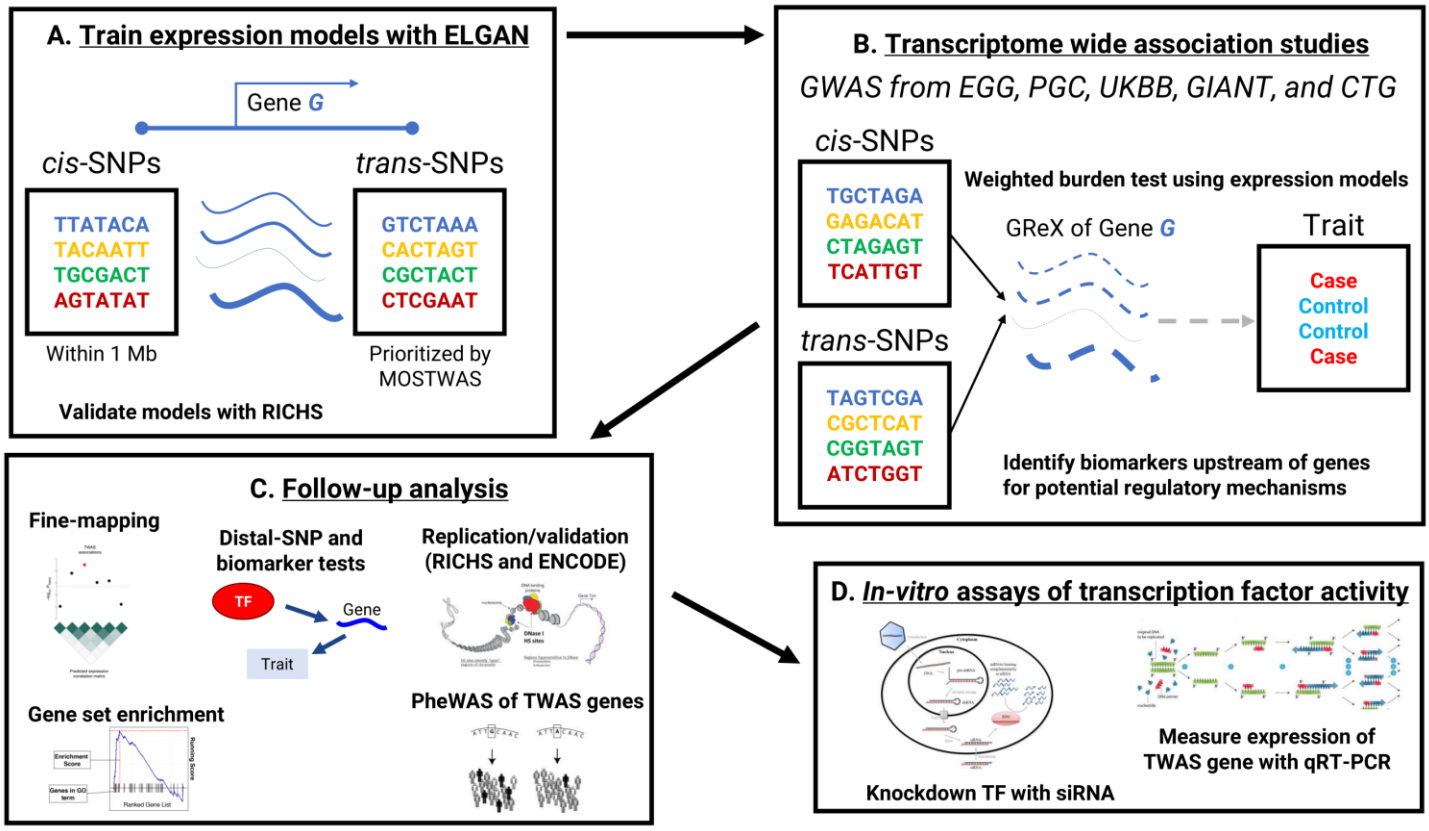

Figure 2: ***Overview of analysis scheme.*** (A) First, predictive models of placental expression are trained from germline genetics, enriched for mediating biomarkers (transcription factors, microRNAs, CpG methylation sites) using MOSTWAS[34]. Models are externally validated in RICHS[42] eQTL data. (B) Predictive models are integrated with GWAS for 40 traits to detect overall and distally-mediated gene-trait associations (GTAs) between placental genetically-regulated expression (GReX) and traits. (C) GTAs are followed up with gene set enrichment and ontology analyses, probabilistic fine-mapping, and phenome-wide scans of genes with multiple GTAs. Distal mediators are also identified for GTAs and associations between transcription factors (TFs) or CpG sites and TWAS genes are investigated further in external datasets (RICHS and ENCODE[12]). (D) *In-vitro* validation of prioritized TF-TWAS gene pairs are conducted using placenta-derived trophoblasts by knocking down the TF-coding gene and measuring expression of TWAS gene.

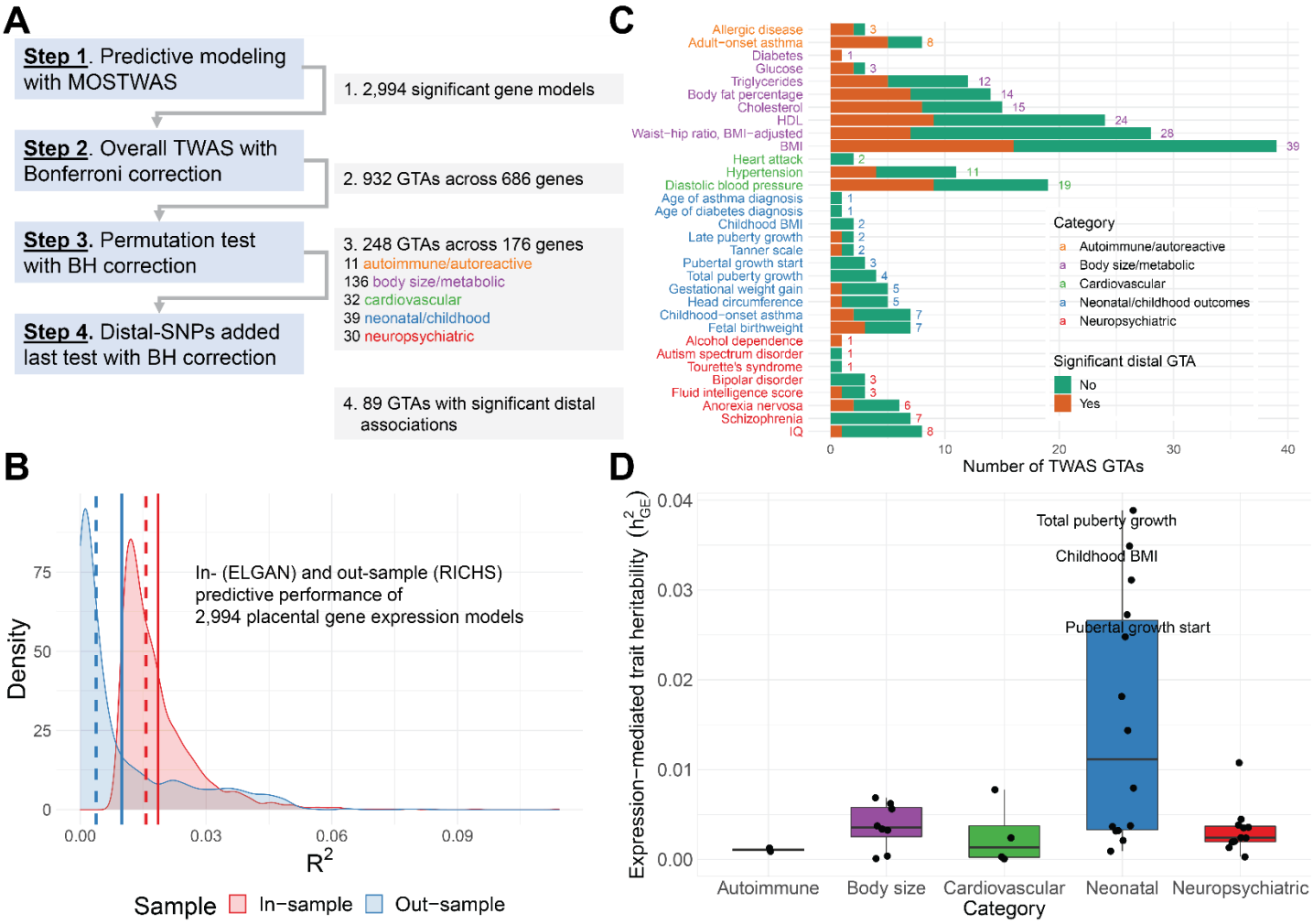

Figure 3: ***Placental MOSTWAS prediction and association test results.*** (A) Overview of TWAS association testing pipeline with number of gene-trait associations (GTAs) across unique genes over various levels of TWAS tests. (B) Kernel density plots of in- (through cross-validation in ELGAN, red) and 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) Box plot of expression-mediated trait heritability (Y-axis) by category (X-axis), with labels if $\hat{h}^2_{GE}$ is significantly greater than 0 using jack-knife test of significance.

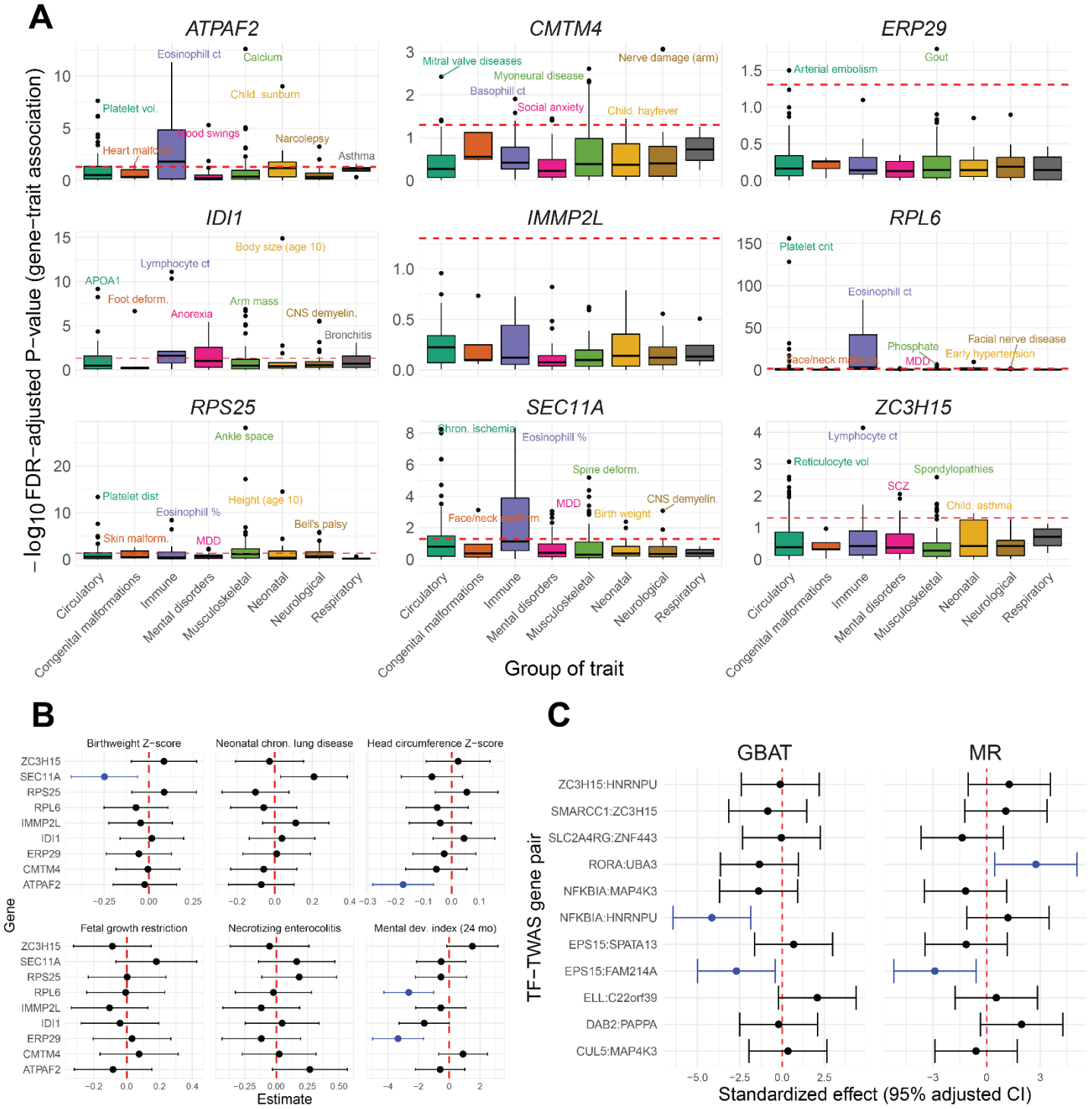

Figure 4: ***Computational follow-up analyses of TWAS-prioritized genes.*** (A) Boxplot of -log10 FDR-adjusted P-value of multi-trait scans of GTAs in UKBB, grouped by 8 ICD code blocks across 9 genes with multiple TWAS GTAs across different trait categories. The red dotted line represents FDR-adjusted 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 TF-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 TF-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.

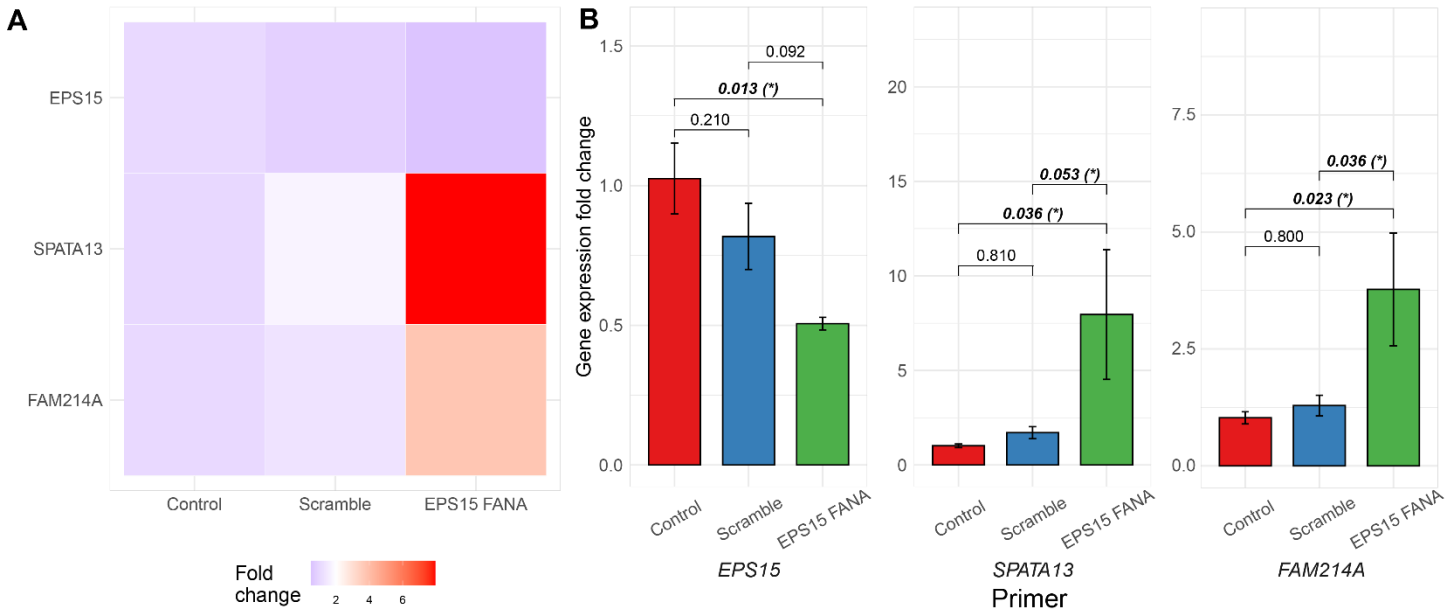

Figure 5: ***Expression of EPS15, SPATA13, and FAM214A in in vitro experiments of EPS15 knockdown*.** (A) Heatmap demonstrating the changes in gene expression from the qRT-PCR data. Color changes are associated with the fold change in gene expression. (B) Bar graph of the gene expression fold-changes from the qRT-PCR from JEG-3 RNA. Nominal P-values of pairwise t-tests are shown, with an asterisk if Benjamini-Hochberg adjusted $P < 0.10$. Note differences in Y-axis scales.

Table 1: Summary of traits and GWAS and TWAS results

Category	Trait	Sample Size	Cases	Number of significant GWAS SNPs	Number of TWAS Genes (total)	Number of TWAS Genes (not overlapping GWAS loci)
Autoimmune/ autoreactive	Adult-onset asthma	420,473	48865	27,763	8	4
	Allergic disease	242,569	180,129	8,979	3	3
	Epilepsy	409,331	6082	1	0	0
Body size/ metabolic	BMI	806,834	-	42,616	39	24
	Body fat percentage	412,960	-	61,460	14	3
	Cholesterol	400,963	-	55,265	15	6
	Diabetes	420,473	16550	10,863	1	1
	Glucose	366,759	-	12,082	3	2
	HDL	367,021	-	82,962	24	4
	Triglycerides	400,639	-	42,637	12	3
	Waist-hip ratio, BMI-adjusted	694,649	-	24,275	28	15
Cardiovascular	Angina	420,473	13408	1,191	0	0
	Diastolic blood pressure	396,667	-	83,826	19	9
	Heart attack	420,473	9743	1,024	2	2
	Hypertension	419,799	109307	32,296	11	7
Neonatal/ childhood traits	Age of asthma diagnosis	13,511	-	378	1	1
	Age of diabetes diagnosis	19,000	-	9,509	1	1
	Age of hayfever diagnosis	24,013	-	1,315	0	0
	Childhood BMI	10,555	-	510	2	2
	Childhood obesity	21,309	8613	804	0	0
	Childhood onset asthma	314,633	13,962	23,260	7	3
	Fetal birthweight	298,142	-	9,388	7	4
	Gestational duration	84,689	-	59	0	0
	Gestational weight gain	16,469	-	2	5	5
	Head circumference	10,678	-	0	5	5
	Late puberty growth	9,228	-	13	2	2
	Pubertal growth start	13,960	-	89	3	3
	Tanner scale	9,916	-	9	2	2
	Total puberty growth	10,799	-	22	4	4
Neuropsychiatric	ADHD	55,374	20183	317	0	0
	Alcohol dependence	52,848	11569	6	1	1

Anorexia nervosa	72,517	16,992	326	6	6
Autism spectrum disorder	46,350	18,381	93	1	1
Bipolar disorder	51,710	20,352	264	3	2
Fluid intelligence score	135,088	-	7,434	3	3
IQ	269,867	-	12,110	8	5
Major depressive disorder	480,359	135458	912	0	0
Obsessive compulsive disorder	9,725	2688	0	0	0
Schizophrenia	150,064	36989	14,505	7	6
Tourette's syndrome	14,307	4819	1	1	1

Table 2: 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 3: Susceptibility genes associated with multiple traits. TWAS gene, location (chromosome:start-end), 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