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: 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

SUMMARY

As the master regulator *in utero*, the placenta is core to the Developmental Origins of Health and Disease (DOHad) hypothesis but is understudied in tissue-specific regulatory consortia. To identify placental gene-trait associations across the life course, we performed distal mediator-enriched transcriptome-wide association studies (TWAS) for 40 traits, using gene expression models trained with placental multi-omics from the Extremely Low Gestational Age Newborn Study. At $P < 2.5 \times 10^{-6}$, we detected 248 gene-trait associations (GTAs), mostly for neonatal and metabolic traits, across 176 genes, enriched for cell growth and immunological pathways. 89 GTAs showed significant mediation through distal genetic variants, identifying functional hypotheses for upstream regulation. Investigation of one hypothesis in human placenta-derived trophoblasts showed that knockdown of mediator gene *EPS15* upregulated its predicted targets, *SPATA13* and *FAM214A*, both associated with waist-hip ratio in TWAS, as well as multiple genes involved in metabolic or hormone-related pathways. These results suggest profound health impacts of placental genetic and 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 control (Baron-Cohen et al., 2019; McKay, 2011; Thornburg et al., 2010). 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 (Gillman, 2005). 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 1A**) (Bronson and Bale, 2016; McKay, 2011; Peng et al., 2018; Tedner et al., 2012; Ursini et al., 2018). Despite its long-lasting influences on health, the placenta has not been well-studied in large consortia studies of multi-tissue gene regulation (Abascal et al., 2020; Aguet et al., 2020). 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 (Delahaye et al., 2018). Furthermore, there is growing evidence that the placental epigenome influences gene regulation, often distally (more than 1-3 Megabases away in the genome) (Marsit, 2016), and that placental DNA methylation and microRNA (miRNA) expression are associated with health traits in children (Paquette et al., 2016). Dysfunction of transcription factor regulation in the placenta has also shown profound effects on childhood traits (Aplin et al., 2020). Although combining genetics, transcriptomics, and epigenomics lends insight into the influence of placental genomics on complex traits (Santos Jr et al., 2020), 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 (Gamazon et al., 2015; Gusev et al., 2016). 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 (Pierce et al., 2018). Not only may these distal biomarkers explain a significant portion of both gene expression heritability and trait heritability on the tissue-specific expression level (Boyle et al., 2017; Liu et al., 2019), 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 1A**). We leveraged multi-omic data from fetal-side placenta tissue from the Extremely Low Gestational Age Newborn (ELGAN) Cohort Study (O'Shea et al., 2009) to train predictive models of gene expression enriched for distal SNPs using MOSTWAS, a recent TWAS extension that integrates multi-omic data (Bhattacharya et al., 2021). Using 40 GWAS of European-ancestry subjects from large consortia (Bycroft et al., 2018; Middeldorp et al., 2019; Savage et al., 2018; Sullivan et al., 2018; Willer et al., 2009), 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 1B**). To our knowledge, this is the first distal mediator-enriched TWAS of health traits that integrates placental multi-omics.

RESULTS

Complex traits are genetically heritable and correlated

From large consortia (Bycroft et al., 2018; Middeldorp et al., 2019; Savage et al., 2018; Sullivan et al., 2018; Willer et al., 2009), 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 (**Supplemental Table S1**, **Supplemental Table S2**). 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 S1**). These five categories of traits have been linked previously to placental biology and morphology (Bronson and Bale, 2016; Peng et al., 2018; Tedner et al., 2012; Ursini et al., 2018).

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 S3**). 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.

Multiple placental gene-trait associations detected across the life course

In the first step of our TWAS (**Figure 1B, Part I**), we leveraged MOSTWAS (Bhattacharya et al., 2021), a recent 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 (Liu et al., 2019; Pierce et al., 2018), 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**) (Bhattacharya et al., 2021). 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 (Ådén et al., 2013) and mRNA expression, CpG methylation, and miRNA expression data from fetal-side placenta (Santos Jr et al., 2020) from the ELGAN Study (O'Shea et al., 2009) 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, 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 2A [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; $N = 149$) (Deyssenroth et al., 2017; Peng et al., 2017), 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 2B; Supplemental Table S5**).

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 (Gusev et al., 2016; Pasaniuc et al., 2014), 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 (Gusev et al., 2016; Mancuso et al., 2017) (**Figure 2A [Step 2]**). 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 (Gusev et al., 2016). 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 2A [Step 3]**, **Supplemental Table S2 and S6;** Miami plots of TWAS Z-scores in **Supplemental Figures S4-S9**).

For example, the 39 GTAs detected with BMI included *LARS2* ($Z = 11.4$) and *CAST* ($Z = -4.61$). These two GTAs have been detected using *cis*-only TWAS in different tissues (Gusev et al., 2016; Mancuso et al., 2017). In addition, one of the 30 genes identified in association with waist-hip ratio was prioritized in other tissues by TWAS: *NDUFS1* ($Z = -5.38$) (Mancuso et al., 2017). 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(Peng et al., 2018). 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 (Peng et al., 2018). 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.

As these GTAs indicate trait association and do not reflect causality, we used FOCUS (Mancuso et al., 2019), 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 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 BMI 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 (**Supplemental Figure S10**, **Supplemental Table S7**) (Liao et al., 2019). 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 overlapped 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, 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 (Bulik-Sullivan et al., 2015; Gusev et al., 2016). 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) (Mancuso et al., 2017); none of the 26 traits outside the neonatal category were appreciably explained by placental GReX. **Figure 2D** 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**). 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.

Similarly, using RHOGE (Mancuso et al., 2017), 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 (Misra et al., 2012). 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.

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 (Nakamura et al., 2015), 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 (Nagarajan et al., 2002). 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 (Lee et al., 2019); the inclusion of this locus may have contributed to the large TWAS associations. Similarly, *SAMD4A* 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*, an angiogenesis regulator (Chrifi et al., 2019), 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 (Kokkinos et al., 2010), 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 3A**). Using UK Biobank (Bycroft et al., 2018) GWAS summary statistics, we conducted TWAS for a variety of traits across 8 groups, defined generally around ICD code blocks (**Figure 3A, 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 3A**). 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, head circumference Z-score, fetal growth restriction, birth weight Z-score, necrotizing enterocolitis, and Bayley II Mental Development Index (MDI) at 24 months (Santos Jr et al., 2020). Shown in **Figure 3B** 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) (Kenmochi et al., 1998; Oue et al., 2014), 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 traits 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 (Lusis et al., 2016). This panel includes 100 inbred mice strains with extensive collection of obesity-related phenotypes from over 12,000 genes. 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, 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 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 (Bhattacharya et al., 2021), 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 S6**). 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* (distal mediator for *PAPPA* and diastolic blood pressure, distal $Z = -3.98$) and *EPS15*, both highly expressed in placenta (Nelissen et al., 2011; Tao et al., 2016). 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 TF involved in inflammatory signaling (Oh et al., 2019), 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 (Everson et al., 2018). 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 S12**), namely catabolic and metabolic processes, response to lipids, and multiple nucleic acid-binding processes (Liao et al., 2019).

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. Using RICHS, we conducted a gene-based *trans*-eQTL scan using Liu *et al*'s GBAT method (Liu et al., 2020) to computationally validate TF-TWAS gene associations. We predicted GReX of the TFs using *cis*-variants through leave-one-out cross-validation and scanned for associations with the respective TWAS genes (**Figure 3C**, **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 (Burgess and Thompson, 2017) in RICHS to estimate the causal effects of TFs on the associated TWAS genes (**Methods and Materials**) 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 3C**, **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]).

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 (Abascal et al., 2020), 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 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 (Reiling et al., 2010) and has shown BMI TWAS associations in other tissues (Gusev et al., 2016; Mancuso et al., 2017). 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 TF *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 trophoblast cells and assessed the gene expression of the targets in no-addition controls, scramble oligo controls, and the knockdown variant via qRT-PCR. 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 transcription factor and target mRNA were statistically significant (**Figure 4A**).

To further investigate the transcriptomic consequences of *EPS15* knockdown *in vitro*, we measured transcriptome-wide gene expression via RNA-seq and conducted differential gene expression analysis across the knockdown cells and scramble oligo controls (Love et al., 2014, 2020; Patro et al., 2017). We detected 2,366 genes down-regulated and 2,212 genes up-regulated in the *EPS15* knockdown cells at FDR-adjusted P < 0.01, validating the negative correlations between *EPS15* and *SPATA13* and *FAM214A* observed in qRT-PCR (**Figure 4B**, **Supplemental Table S16-S17**). In particular, these down-regulated genes were enriched for cell cycle, cell proliferation, or replication pathways, while up-regulated genes were enriched for pathways related to hormones or metabolism, like parathyroid hormone synthesis, insulin resistance, and fructose metabolism (**Figure 4B, Supplemental Table S18-S19**). Enrichments for biological, cellular, and molecular ontologies support these pathway 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* as a potential regulator for multiple genes downstream, perhaps for genes affecting metabolic pathways or cell adhesion and growth in the placenta, like *SPATA13* (Jean et al., 2013), which then affect complex traits later-in-life.

DISCUSSION

The placenta has historically been understudied in large multi-tissue consortia efforts that study tissue-specific regulatory mechanisms (Abascal et al., 2020; Aguet et al., 2020). To address this gap, we systematically categorized placental gene-trait associations relevant to the DOHaD hypothesis using distal mediator-enriched TWAS. By integrating multi-omic data from the ELGAN Study (O'Shea et al., 2009) 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 early-life 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 persist later-in-life 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 (Bhattacharya et al., 2021). Our computational results prioritized 89 GTAs with strong distal associations. We interrogated one such functional hypothesis: *EPS15*, a predicted transcription factor-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 (Bristow et al., 2009; Jean et al., 2013; Kawasaki et al., 2007). In fact, EPS15 itself showed a negative TWAS association with waist-hip ratio. In placenta-derived trophoblasts, knockdown of *EPS15* showed increased expression of both *FAM214A* and *SPATA13*, as well as multiple genes involved in metabolic and hormone-related pathways. In particular, *EPS15*, mainly involved in endocytosis, is a maternally imprinted gene and predicted to promote offspring health (Diplas et al., 2009; Marsit et al.,

2012; Nelissen et al., 2011; Piedrahita, 2011); 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, building off our results showing cis-GReX correlations between EPS15 mouse analog and fatty acid levels. This *in vitro* assay provides valuable evidence for the functional consequences of *EPS15* 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, 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 (Paternoster et al., 2017). 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 control variables to account for unmeasured confounders in predictive models 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 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 (Bhattacharya et al., 2020). 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 health and disease.

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 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.

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).

DECLARATION OF INTERESTS

The authors declare that they have no competing interests.

FIGURE LEGENDS

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. (**B**) Predictive models of placental expression are trained from germline genetics, enriched for mediating biomarkers using MOSTWAS(Bhattacharya et al., 2021) and externally validated in RICHS(Peng et al., 2017) eQTL data (Part I). Predictive models are integrated with GWAS for 40 traits to detect placental gene-trait associations (GTAs) (Part II). GTAs are followed up with gene set enrichment, 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(Abascal et al., 2020) (Part III). In-vitro validation of prioritized transcription factor-TWAS gene pairs are conducted using placenta-derived trophoblasts by gene silencing and qRT-PCR to measure TWAS-gene expression (Part IV).

Figure 2: ***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 3: ***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 4: ***In vitro experiments in EPS15-knockdown human-derived trophoblasts*.** (***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.

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: Gene set enrichments of TWAS-identified genes. 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 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 TFs (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.

STAR 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 (Ådén et al., 2013; Yasuno et al., 2010). Prior to imputation, from the original set of 731,442 markers, we removed SNPs with call rate < 90% and MAF < 1%. 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 (Purcell et al., 2007). We first performed strand-flipping according to the TOPMed Freeze 5 reference panel and using eagle and minimac4 for phasing and imputation (Das et al., 2016; Kowalski et al., 2019; Loh et al., 2016). 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 (Sherry et al., 2001). 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 (Peng et al., 2017).

Expression data

mRNA expression was determined using the Illumina QuantSeq 3' mRNA-Seq Library Prep Kit, a method with high strand specificity (A Eaves et al., 2020). 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 (Patro et al., 2017) 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 (Qi et al., 2019).

Genes and miRNAs with less than 5 counts for each sample were filtered, resulting in 11,224 genes and 2,047 miRNAs for downstream analysis. Distributional differences between lanes were first upper-quartile normalized (Anders and Huber, 2010; Bullard et al., 2010). Unwanted technical and biological variation (e.g. tissue heterogeneity) was then estimated using RUVSeq (Risso et al., 2014), where we empirically defined transcripts not associated with outcomes of interest as negative control housekeeping probes

(Gagnon-Bartsch and Speed, 2012). One dimension of unwanted variation was removed from the variance-stabilized transformation of the gene expression data using the limma package (Gagnon-Bartsch and Speed, 2012; Love et al., 2014; Phipson et al., 2016; Risso et al., 2014). We obtained processed RNA expression data from the Rhode Island Children's Health Study, as described before (Peng et al., 2017). Overall, after QC and normalization, we considered 12,020 genes and 1,898 miRNAs.

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 (Addo et al., 2019; Bulka et al., 2019; Clark et al., 2019; Santos et al., 2019). 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) (Aryee et al., 2014; Fortin et al., 2014, 2017; Johnson et al., 2007; Santos et al., 2019). DNA methylation data was imported into R for pre-processing using the minfi package (Fortin et al., 2014, 2017). 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 (Triche et al., 2013) for background subtraction and dye normalization, followed by the typical functional normalization method with the top two principal components of the control matrix(Fortin et al., 2014, 2017). 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 (Leek and Storey, 2007). 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 (Du et al., 2010). Overall, after QC and normalization, we considered 846,233 CpG sites.

GWAS summary statistics

Summary statistics were downloaded from the following consortia: the UK Biobank (Bycroft et al., 2018), Early Growth Genetics Consortium (Middeldorp et al., 2019), Genetic Investigation of Anthropometric Traits (Willer et al., 2009), Psychiatric Genomics Consortium (Sullivan et al., 2018), and the Complex Trait Genetics Lab (Savage et al., 2018) (**Supplemental Table 1**). Genomic coordinates were transformed to the hg38 reference genome using liftOver (Lawrence et al., 2009, 2020).

QTL mapping

We conducted genome-wide eQTL mapping between all genotypes and all genes in the transcriptome using a standard linear regression in MatrixeQTL (Shabalin, 2012). 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 (Stegle et al., 2012). Mediators here are defined as RNA expression of genes that code for transcription factors, miRNAs, and 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 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 (Yang et al., 2015). Analysis was conducted using GCTA v1.93.1 (Yang et al., 2011). 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 (Bhattacharya et al., 2021) (**Supplemental Figure S1**, **Supplemental Text**). 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* (Bhattacharya et al., 2021). 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

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(Gusev et al., 2016). 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 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(Gusev et al., 2016). 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., 2021)

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 (Bulik-Sullivan et al., 2015). 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 (Gusev et al., 2016; Mancuso et al., 2017). We assume that the expected χ^2 statistic under a complex trait is a linear function of the LD score (Bulik-Sullivan et al., 2015). 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 ROHGE to estimate and test for significant genetic correlations between traits at the predicted expression level; see (Mancuso et al., 2017).

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 (Santos et al., 2018, 2019; Santos Jr et al., 2020) 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 (Peng et al., 2017), 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* (Liu et al., 2020) 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 (Mefford et al., 2020). 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 (Bowden et al., 2016).

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

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 (Livak and Schmittgen, 2001).

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(Love et al., 2020) and summarized to the gene-level. Differential expression analysis between *EPS15* knockdown samples and scramble oligo controls (biological and technical duplicate) was conducted using DESeq2 (Love et al., 2014).

Data and materials 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. The MOSTWAS software is accessible at https://bhattacharya-a-bt.github.io/MOSTWAS/articles/MOSTWAS_vignette.html. All models and full TWAS results can be accessed at https://doi.org/10.5281/zenodo.4618036 (Bhattacharya and Santos Jr, 2021).

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 ($P < 5 \times 10^{-8}$), and number of significant total and GWAS-overlapping TWAS associations ($P < 2.5 \times 10^{-6}$) 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: Self-reported demographics of ELGAN sample. Counts and proportions of self-reported race in ELGAN sample.

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 TF-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 S18: 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.

REFERENCES

A Eaves, L., Phookphan, P., E Rager, J., Bangma, J., Santos, H.P., Smeester, L., O'Shea, T.M., C Fry, R., Eaves, L.A., Phookphan, P., et al. (2020). A role for microRNAs in the epigenetic control of sexually dimorphic gene expression in the human placenta. Epigenomics *12*, 1543–1558.

Abascal, F., Acosta, R., Addleman, N.J., Adrian, J., Afzal, V., Aken, B., Akiyama, J.A., Jammal, O. Al, Amrhein, H., Anderson, S.M., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature *583*, 699–710.

Addo, K.A., Bulka, C., Dhingra, R., Santos, H.P., Jr, Smeester, L., O'Shea, T.M., and Fry, R.C. (2019). Acetaminophen use during pregnancy and DNA methylation in the placenta of the extremely low gestational age newborn (ELGAN) cohort. Environ. Epigenetics *5*.

Ådén, U., Lin, A., Carlo, W., Leviton, A., Murray, J.C., Hallman, M., Lifton, R.P., Zhang, H., and Ment, L.R. (2013). Candidate gene analysis: Severe intraventricular hemorrhage in inborn preterm neonates. J. Pediatr. *163*.

Aguet, F., Barbeira, A.N., Bonazzola, R., Brown, A., Castel, S.E., Jo, B., Kasela, S., Kim-Hellmuth, S., Liang, Y., Oliva, M., et al. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science (80-.). *369*, 1318–1330.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome

Biol. *11*, R106.
Aplin, J.D., Myers, J.E., Timms, K., and Westwood, M. (2020). Tracking placental development in health and disease. Nat. Rev. Endocrinol. *16*, 479–494.
Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363–1369.
Baron-Cohen, S., Tsompanidis, A., Auyeung, B., Nørgaard-Pedersen, B., Hougaard, D.M., Abdallah, M., Cohen, A., and Pohl, A. (2019). Foetal oestrogens and autism. Mol. Psychiatry 1–9.
Bhattacharya, A., and Santos Jr, H. (2021). Distal-mediator enriched placental gene expression models and TWAS results for DOHaD-related traits.
Bhattacharya, A., García-Closas, M., Olshan, A.F., Perou, C.M., Troester, M.A., and Love, M.I. (2020). A framework for transcriptome-wide association studies in breast cancer in diverse study populations. Genome Biol. *21*, 42.
Bhattacharya, A., Li, Y., and Love, M.I. (2021). MOSTWAS: Multi-Omic Strategies for Transcriptome-Wide Association Studies. PLOS Genet. *17*, e1009398.
Bowden, J., Davey Smith, G., Haycock, P.C., and Burgess, S. (2016). Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet. Epidemiol. *40*, 304–314.
Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell *169*, 1177–1186.
Bristow, J.M., Sellers, M.H., Majumdar, D., Anderson, B., Hu, L., and Webb, D.J. (2009). The Rho-family GEF Asef2 activates Rac to modulate adhesion and actin dynamics and thereby regulate cell migration. J. Cell Sci. *122*, 4535–4546.
Bronson, S.L., and Bale, T.L. (2016). The Placenta as a Mediator of Stress Effects on Neurodevelopmental Reprogramming. Neuropsychopharmacology *41*, 207–218.
Bulik-Sullivan, B., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M., Corvin, A., et al. (2015). LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.
Bulka, C.M., Dammann, O., Santos, H.P., VanderVeen, D.K., Smeester, L., Fichorova, R., O'Shea, T.M., and Fry, R.C. (2019). Placental CpG methylation of inflammation, angiogenic, and neurotrophic genes and retinopathy of prematurity. Investig. Ophthalmol. Vis. Sci. *60*, 2888–2894.
Bullard, J.H., Purdom, E., Hansen, K.D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics *11*, 94.
Burgess, S., and Thompson, S.G. (2017). Interpreting findings from Mendelian randomization using the MR-Egger method. Eur. J. Epidemiol. *32*, 377–389.
Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209.
Chrifi, I., Louzao-Martinez, L., Brandt, M.M., van Dijk, C.G.M., Bürgisser, P.E., Zhu, C., Kros, J.M., Verhaar, M.C., Duncker, D.J., and Cheng, C. (2019). CMTM4 regulates angiogenesis by promoting cell surface recycling of VE-cadherin to endothelial adherens junctions. Angiogenesis *22*, 75–93.
Clark, J., Martin, E., Bulka, C.M., Smeester, L., Santos, H.P., O'Shea, T.M., and Fry, R.C. (2019). 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.
Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat. Genet. *48*, 1284–1287.
Delahaye, F., Do, C., Kong, Y., Ashkar, R., Salas, M., Tycko, B., Wapner, R., and Hughes, F. (2018). Genetic variants influence on the placenta regulatory landscape. PLoS Genet. *14*, e1007785.
Deyssenroth, M.A., Peng, S., Hao, K., Lambertini, L., Marsit, C.J., and Chen, J. (2017). Whole-transcriptome analysis delineates the human placenta gene network and its associations with fetal growth. BMC Genomics *18*, 520.
Diplas, A.I., Lambertini, L., Lee, M.J., Sperling, R., Lee, Y.L., Wetmur, J., and Chen, J. (2009). Differential expression of imprinted genes in normal and IUGR human placentas. Epigenetics *4*, 235–240.
Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W.A., Hou, L., and Lin, S.M. (2010). Comparison of

Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics *11*, 587.
Everson, T.M., Punshon, T., Jackson, B.P., Hao, K., Lambertini, L., Chen, J., Karagas, M.R., and Marsit, C.J. (2018). Cadmium-associated differential methylation throughout the placental genome: Epigenome-wide association study of two U.S. birth cohorts. Environ. Health Perspect. *126*.
Fortin, J.-P., Labbe, A., Lemire, M., Zanke, B.W., Hudson, T.J., Fertig, E.J., Greenwood, C.M., and Hansen, K.D. (2014). Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. *15*, 503.
Fortin, J.P., Triche, T.J., and Hansen, K.D. (2017). Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics *33*, 558–560.
Gagnon-Bartsch, J.A., and Speed, T.P. (2012). Using control genes to correct for unwanted variation in microarray data. Biostatistics *13*, 539–552.
Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S. V, Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Nicolae, D.L., Cox, N.J., et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet. *47*, 1091–1098.
Gillman, M.W. (2005). Developmental origins of health and disease. N. Engl. J. Med. *353*, 1848–1850.
Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252.
Jean, L., Majumdar, D., Shi, M., Hinkle, L.E., Diggins, N.L., Ao, M., Broussard, J.A., Evans, J.C., Choma, D.P., and Webb, D.J. (2013). Activation of Rac by Asef2 promotes myosin IIdependent contractility to inhibit cell migration on type I collagen. J. Cell Sci. *126*, 5585–5597.
Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127.
Kawasaki, Y., Sagara, M., Shibata, Y., Shirouzu, M., Yokoyama, S., and Akiyama, T. (2007). Identification and characterization of Asef2, a guanine-nucleotide exchange factor specific for Rac1 and Cdc42. Oncogene *26*, 7620–7627.
Kenmochi, N., Kawaguchi, T., Rozen, S., Davis, E., Goodman, N., Hudson, T.J., Tanaka, T., and Page, D.C. (1998). A map of 75 human ribosomal protein genes. Genome Res. *8*, 509–523.
Kokkinos, M.I., Murthi, P., Wafai, R., Thompson, E.W., and Newgreen, D.F. (2010). Cadherins in the human placenta - Epithelial-mesenchymal transition (EMT) and placental development. Placenta *31*, 747–755.
Kowalski, M.H., Qian, H., Hou, Z., Rosen, J.D., Tapia, A.L., Shan, Y., Jain, D., Argos, M., Arnett, D.K., Avery, C., et al. (2019). 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.
Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. Bioinformatics *25*, 1841–1842.
Lawrence, M., Carey, V., and Gentleman, R. (2020). liftOver.
Lee, C.J., Park, M.S., Kim, M., Ann, S. jin, Lee, J., Park, S., Kang, S.M., Jang, Y., Lee, J.H., and Lee, S.H. (2019). CETP, LIPC, and SCARB1 variants in individuals with extremely high high-density lipoprotein-cholesterol levels. Sci. Rep. *9*, 1–7.
Leek, J.T., and Storey, J.D. (2007). Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. PLoS Genet. *3*, e161.
Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. *47*, 199–205.
Liu, X., Li, Y.I., and Pritchard, J.K. (2019). Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. Cell *177*, 1022-1034.e6.
Liu, X., Mefford, J.A., Dahl, A., He, Y., Subramaniam, M., Battle, A., Price, A.L., and Zaitlen, N. (2020). GBAT: a gene-based association test for robust detection of trans-gene regulation. Genome Biol. *21*, 211.
Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods *25*, 402–408.
Loh, P.R., Danecek, P., Palamara, P.F., Fuchsberger, C., Reshef, Y.A., Finucane, H.K., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. Nat. Genet. *48*, 1443–1448.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.
Love, M.I., Soneson, C., Hickey, P.F., Johnson, L.K., Pierce, N.T., Shepherd, L., Morgan, M., and Patro, R. (2020). Tximeta: Reference sequence checksums for provenance identification in RNA-seq. PLOS Comput. Biol. *16*, e1007664.
Lusis, A.J., Seldin, M.M., Allayee, H., Bennett, B.J., Civelek, M., Davis, R.C., Eskin, E., Farber, C.R., Hui, S., Mehrabian, M., et al. (2016). The hybrid mouse diversity panel: A resource for systems genetics analyses of metabolic and cardiovascular traits. J. Lipid Res. *57*, 925–942.
Mancuso, N., Shi, H., Goddard, P., Kichaev, G., Gusev, A., and Pasaniuc, B. (2017). Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. Am. J. Hum. Genet. *100*, 473–487.
Mancuso, N., Freund, M.K., Johnson, R., Shi, H., Kichaev, G., Gusev, A., and Pasaniuc, B. (2019). Probabilistic fine-mapping of transcriptome-wide association studies. Nat. Genet. *51*, 675–682.
Marsit, C.J. (2016). Placental Epigenetics in Children's Environmental Health. Semin. Reprod. Med. *34*, 36–41.
Marsit, C.J., Lambertini, L., MacCani, M.A., Koestler, D.C., Houseman, E.A., Padbury, J.F., Lester, B.M., and Chen, J. (2012). Placenta-imprinted gene expression association of infant neurobehavior. J. Pediatr. *160*, 854.
McKay, R. (2011). Developmental biology: Remarkable role for the placenta. Nature *472*, 298–299.
Mefford, J., Park, D., Zheng, Z., Ko, A., Ala-Korpela, M., Laakso, M., Pajukanta, P., Yang, J., Witte, J., and Zaitlen, N. (2020). Efficient Estimation and Applications of Cross-Validated Genetic Predictions to Polygenic Risk Scores and Linear Mixed Models. J. Comput. Biol. *27*, 599–612.
Middeldorp, C.M., Felix, J.F., Mahajan, A., Ahluwalia, T.S., Auvinen, J., Bartels, M., Bilbao, J.R., Bisgaard, H., Bønnelykke, K., Boomsma, D.I., et al. (2019). The Early Growth Genetics (EGG) and EArly Genetics and Lifecourse Epidemiology (EAGLE) consortia: design, results and future prospects. Eur. J. Epidemiol. *34*, 9.
Misra, D.P., Salafia, C.M., Charles, A.K., and Miller, R.K. (2012). Placental measurements associated with intelligence quotient at age 7 years. J. Dev. Orig. Health Dis. *3*, 190–197.
Nagarajan, R., Le, N., Mahoney, H., Araki, T., and Milbrandt, J. (2002). Deciphering peripheral nerve myelination by using Schwann cell expression profiling. Proc. Natl. Acad. Sci. U. S. A. *99*, 8998–9003.
Nakamura, K., Mori, F., Tanji, K., Miki, Y., Yamada, M., Kakita, A., Takahashi, H., Utsumi, J., Sasaki, H., and Wakabayashi, K. (2015). Isopentenyl diphosphate isomerase, a cholesterol synthesizing enzyme, is localized in Lewy bodies. Neuropathology *35*, 432–440.
Nelissen, E.C.M., van Montfoort, A.P.A., Dumoulin, J.C.M., and Evers, J.L.H. (2011). Epigenetics and the placenta. Hum. Reprod. Update *17*, 397–417.
O'Shea, T.M., Allred, E.N., Dammann, O., Hirtz, D., Kuban, K.C.K., Paneth, N., and Leviton, A. (2009). The ELGAN study of the brain and related disorders in extremely low gestational age newborns. Early Hum. Dev. *85*, 719–725.
Oh, S.K., Kim, D., Kim, K., Boo, K., Yu, Y.S., Kim, I.S., Jeon, Y., Im, S.K., Lee, S.H., Lee, J.M., et al. (2019). RORα is crucial for attenuated inflammatory response to maintain intestinal homeostasis. Proc. Natl. Acad. Sci. U. S. A. *116*, 21140–21149.
Oue, N., Naito, Y., Hayashi, T., Takigahira, M., Kawano-Nagatsuma, A., Sentani, K., Sakamoto, N., Zarni Oo, H., Uraoka, N., Yanagihara, K., et al. (2014). Signal peptidase complex 18, encoded by SEC11A, contributes to progression via TGF-α secretion in gastric cancer. Oncogene *33*, 3918–3926.
Paquette, A.G., Houseman, E.A., Green, B.B., Lesseur, C., Armstrong, D.A., Lester, B., and Marsit, C.J. (2016). Regions of variable DNA methylation in human placenta associated with newborn neurobehavior. Epigenetics *11*, 603–613.
Pasaniuc, B., Zaitlen, N., Shi, H., Bhatia, G., Gusev, A., Pickrell, J., Hirschhorn, J., Strachan, D.P., Patterson, N., Price, A.L., et al. (2014). Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. Bioinformatics *30*, 2906–2914.
Paternoster, L., Tilling, K., and Davey Smith, G. (2017). Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges. PLOS Genet. *13*, e1006944.
Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.

Peng, S., Deyssenroth, M.A., Di Narzo, A.F., Lambertini, L., Marsit, C.J., Chen, J., and Hao, K. (2017). Expression quantitative trait loci (eQTLs) in human placentas suggest developmental origins of complex diseases. Hum. Mol. Genet. *26*, 3432–3441.

Peng, S., Deyssenroth, M.A., Di Narzo, A.F., Cheng, H., Zhang, Z., Lambertini, L., Ruusalepp, A., Kovacic, J.C., Bjorkegren, J.L.M., Marsit, C.J., et al. (2018). Genetic regulation of the placental transcriptome underlies birth weight and risk of childhood obesity. PLOS Genet. *14*, e1007799.

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

Piedrahita, J.A. (2011). The role of imprinted genes in fetal growth abnormalities. Birth Defects Res. Part A - Clin. Mol. Teratol. *91*, 682–692.

Pierce, B.L., Tong, L., Argos, M., Demanelis, K., Jasmine, F., Rakibuz-Zaman, M., Sarwar, G., Islam, M.T., Shahriar, H., Islam, T., et al. (2018). Co-occurring expression and methylation QTLs allow detection of common causal variants and shared biological mechanisms. Nat. Commun. *9*, 1–12.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet *81*, 559–575.

Qi, Z., Wang, L., Desai, K., Cogswell, J., Stern, M., Lawson, B., Kerkar, S.P., and Vitazka, P. (2019). 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.

Reiling, E., Jafar-Mohammadi, B., Van ’t Riet, E., Weedon, M.N., Van Vliet-Ostaptchouk, J. V, Hansen, T., Saxena, R., Van Haeften, T.W., Arp, P.A., Das, S., et al. (2010). Genetic association analysis of LARS2 with type 2 diabetes. Diabetologia *53*, 103–110.

Risso, D., Ngai, J., Speed, T.P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. Nat. Biotechnol. *32*, 896–902.

Santos, H.P., Nephew, B.C., Bhattacharya, A., Tan, X., Smith, L., Alyamani, R.A.S., Martin, E.M., Perreira, K., Fry, R.C., and Murgatroyd, C. (2018). Discrimination exposure and DNA methylation of stress-related genes in Latina mothers. Psychoneuroendocrinology *98*, 131–138.

Santos, H.P., Bhattacharya, A., Martin, E.M., Addo, K., Psioda, M., Smeester, L., Joseph, R.M., Hooper, S.R., Frazier, J.A., Kuban, K.C., et al. (2019). Epigenome-wide DNA methylation in placentas from preterm infants: association with maternal socioeconomic status. Epigenetics *14*, 751–765.

Santos Jr, H.P., Bhattacharya, A., Joseph, R.M., Smeester, L., Kuban, K.C.K., Marsit, C.J., O’Shea, T.M., and Fry, R.C. (2020). Evidence for the placenta-brain axis: multi-omic kernel aggregation predicts intellectual and social impairment in children born extremely preterm. Mol. Autism *11*, 97.

Savage, J.E., Jansen, P.R., Stringer, S., Watanabe, K., Bryois, J., De Leeuw, C.A., Nagel, M., Awasthi, S., Barr, P.B., Coleman, J.R.I., et al. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat. Genet. *50*, 912–919.

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

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). DbSNP: The NCBI database of genetic variation. Nucleic Acids Res. *29*, 308–311.

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

Sullivan, P.F., Agrawal, A., Bulik, C.M., Andreassen, O.A., Børglum, A.D., Breen, G., Cichon, S., Edenberg, H.J., Faraone, S. V, Gelernter, J., et al. (2018). Psychiatric Genomics: An Update and an Agenda. Am. J. Psychiatry *175*, 15.

Tao, W., Moore, R., Smith, E.R., and Xu, X.X. (2016). Endocytosis and physiology: Insights from disabled-2 deficient mice. Front. Cell Dev. Biol. *4*, 129.

Tedner, S.G., Örtqvist, A.K., and Almqvist, C. (2012). Fetal growth and risk of childhood asthma and allergic disease. Clin. Exp. Allergy *42*, 1430–1447.

Thornburg, K.L., O’Tierney, P.F., and Louey, S. (2010). The Placenta is a Programming Agent for Cardiovascular Disease. Placenta *31*, S54.

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

Ursini, G., Punzi, G., Chen, Q., Marenco, S., Robinson, J.F., Porcelli, A., Hamilton, E.G., Mitjans, M., Maddalena, G., Begemann, M., et al. (2018). Convergence of placenta biology and genetic risk for schizophrenia article. Nat. Med. *24*, 792–801.

Willer, C.J., Speliotes, E.K., Loos, R.J.F., Li, S., Lindgren, C.M., Heid, I.M., Berndt, S.I., Elliott, A.L., Jackson, A.U., Lamina, C., et al. (2009). Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat. Genet. *41*, 25–34.

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

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A.E., Lee, S.H., Robinson, M.R., Perry, J.R.B., Nolte, I.M., van Vliet-Ostaptchouk, J. V, et al. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat. Genet. *47*, 1114–1120.

Yasuno, K., Bilguvar, K., Bijlenga, P., Low, S.K., Krischek, B., Auburger, G., Simon, M., Krex, D., Arlier, Z., Nayak, N., et al. (2010). Genome-wide association study of intracranial aneurysm identifies three new risk loci. Nat. Genet. *42*, 420–425.

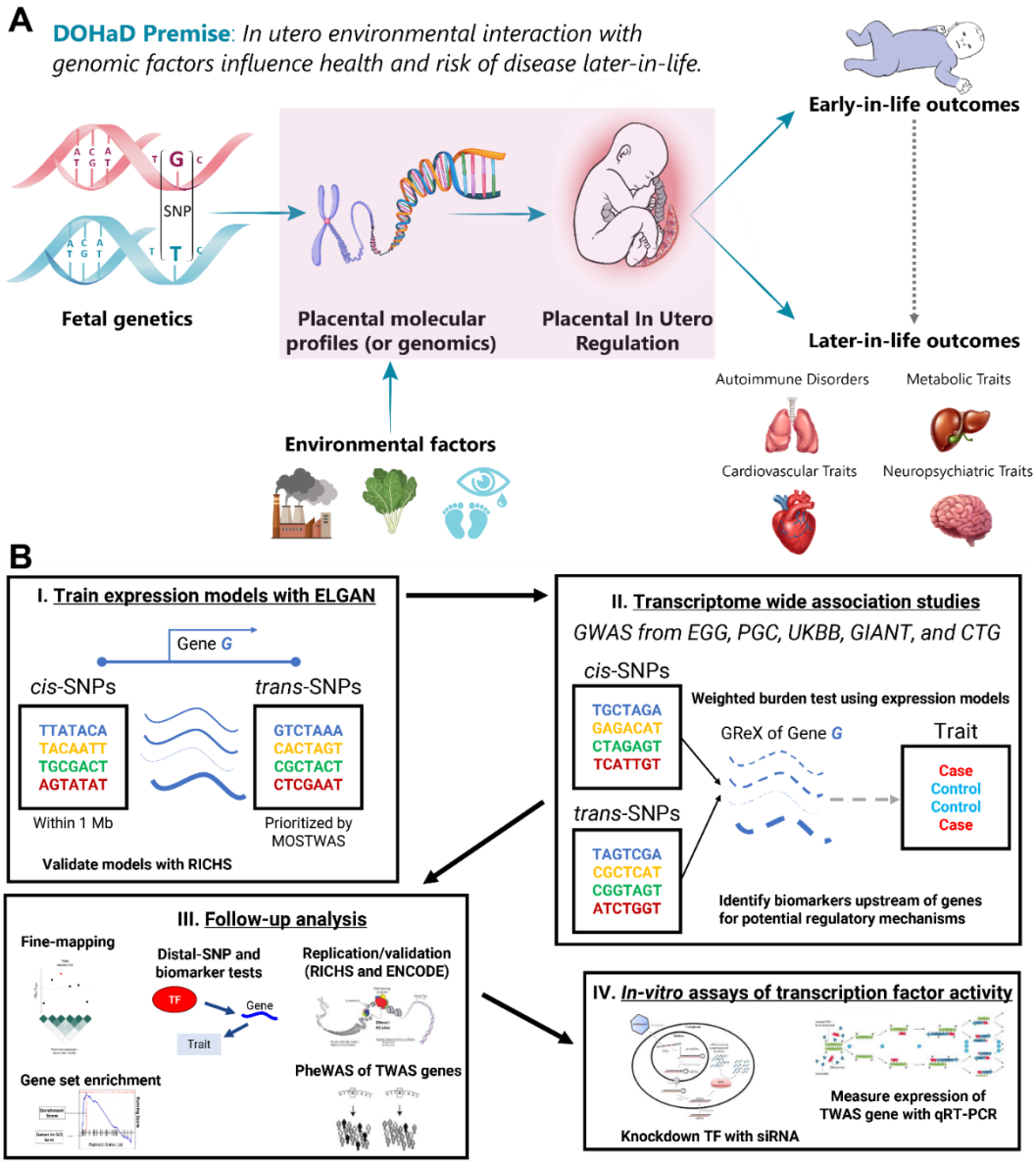

Figure 1: **Overview of the DOHaD Hypothesis and study scheme.** (**A**) The placenta facilitates important functions *in utero*, including nutrient transfer, metabolism, gas exchange, neuroendocrine signaling, growth hormone production, and immunologic control. Accordingly, it is a master regulator 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. (**B**) Predictive models of placental expression are trained from germline genetics, enriched for mediating biomarkers using MOSTWAS(Bhattacharya et al., 2021) and externally validated in RICHS(Peng et al., 2017) eQTL data (Part I). Predictive models are integrated with GWAS for 40 traits to detect placental gene-trait associations (GTAs) (Part II). GTAs are followed up with gene set enrichment, 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 (Abascal et al., 2020) (Part III). In-vitro validation of prioritized transcription factor-TWAS gene pairs are conducted using placenta-derived trophoblasts by gene silencing and qRT-PCR to measure TWAS-gene expression (Part IV).

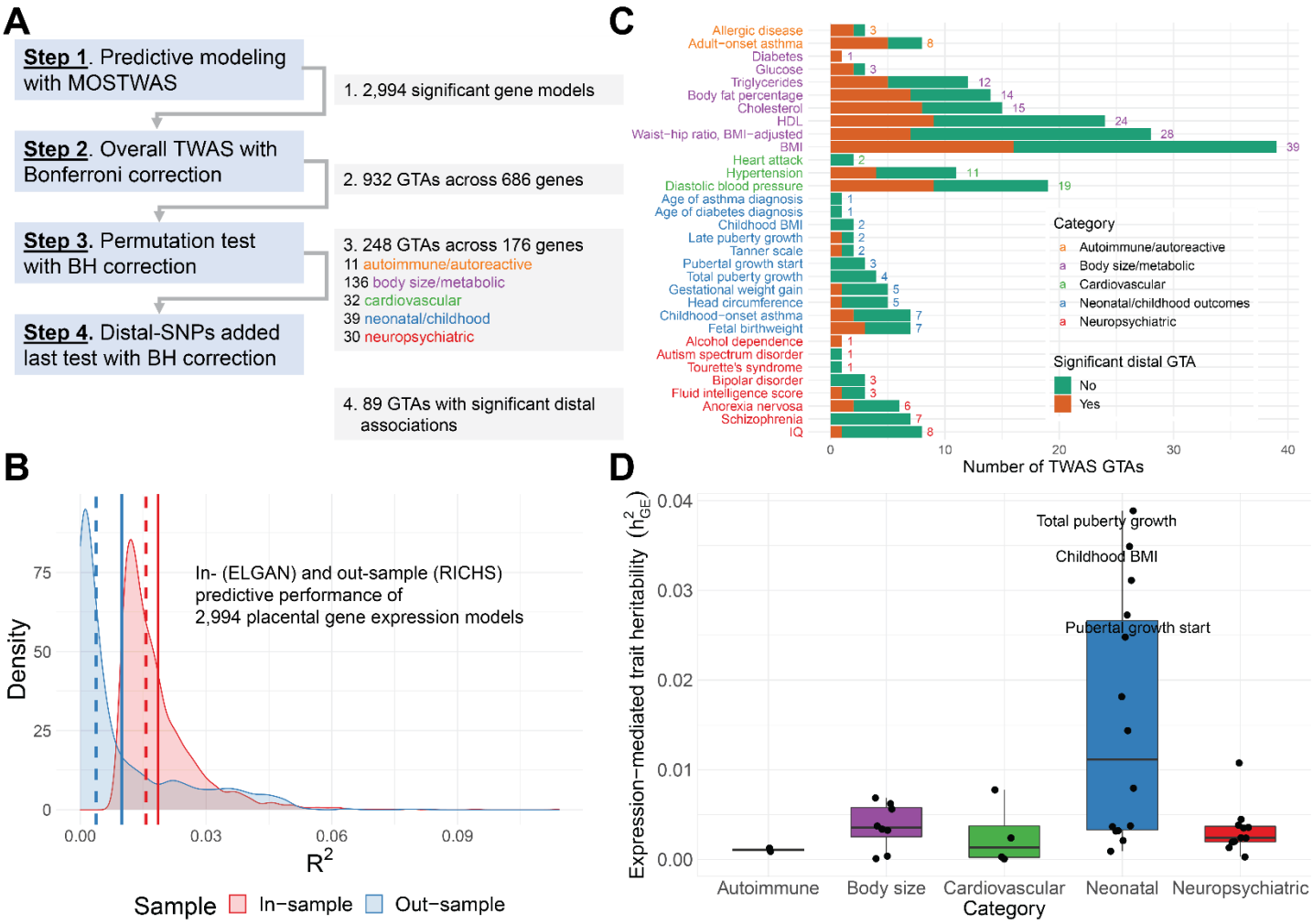

Figure 2: Placental MOSTWAS prediction and association test results. (**A**) Overview of TWAS association testing pipeline with number of gene-trait associations (GTAs) across unique genes over various levels of TWAS tests. (**B**) Kernel density plots of in- (through cross-validation in ELGAN, red) 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.

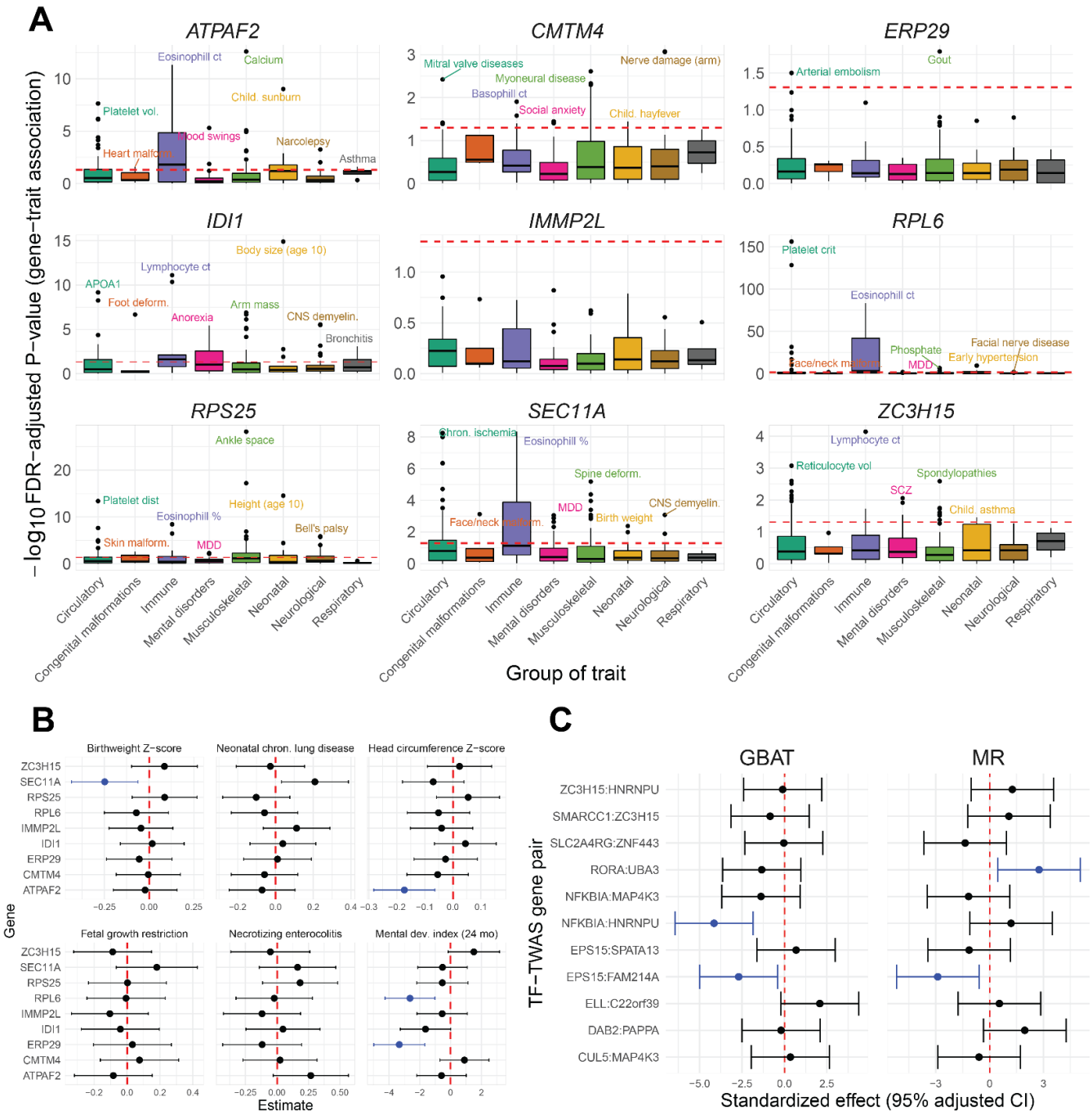

Figure 3: ***Computational follow-up analyses of TWAS-prioritized genes*.** (**A**) Boxplot of -log10 FDR-adjusted P-value of multi-trait scans of GTAs in UKBB, grouped by 8 ICD code blocks across 9 genes with multiple TWAS GTAs across different trait categories. The red dotted line represents FDR-adjusted 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.

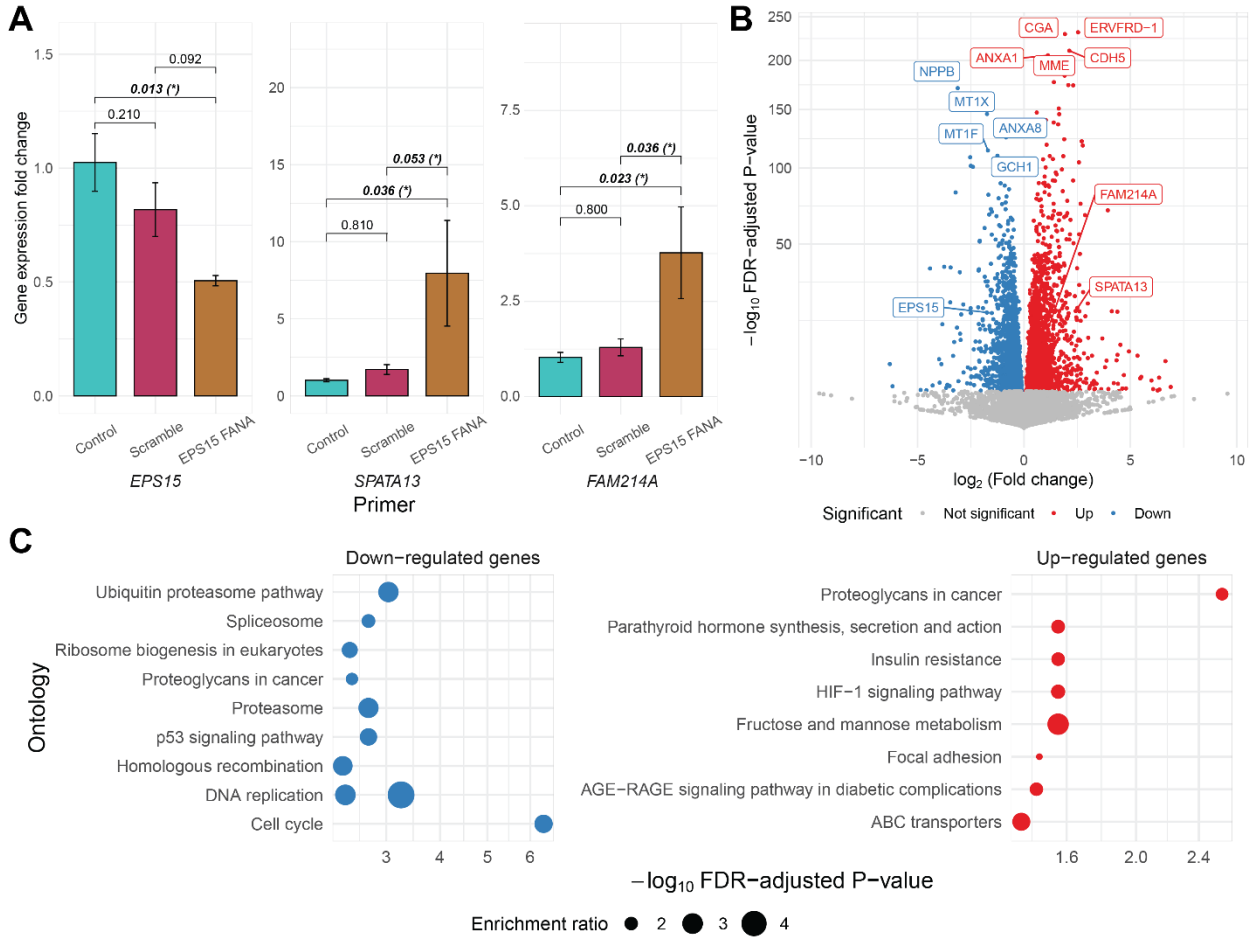

Figure 4: **In vitro experiments in EPS15-knockdown human-derived trophoblasts.** (***A***) Bar graph of the gene expression fold-changes from the qRT-PCR from JEG-3 RNA. Nominal P-values of pairwise t-tests are shown, with an asterisk if Benjamini-Hochberg *FDR*-adjusted $P < 0.10$. Note differences 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.

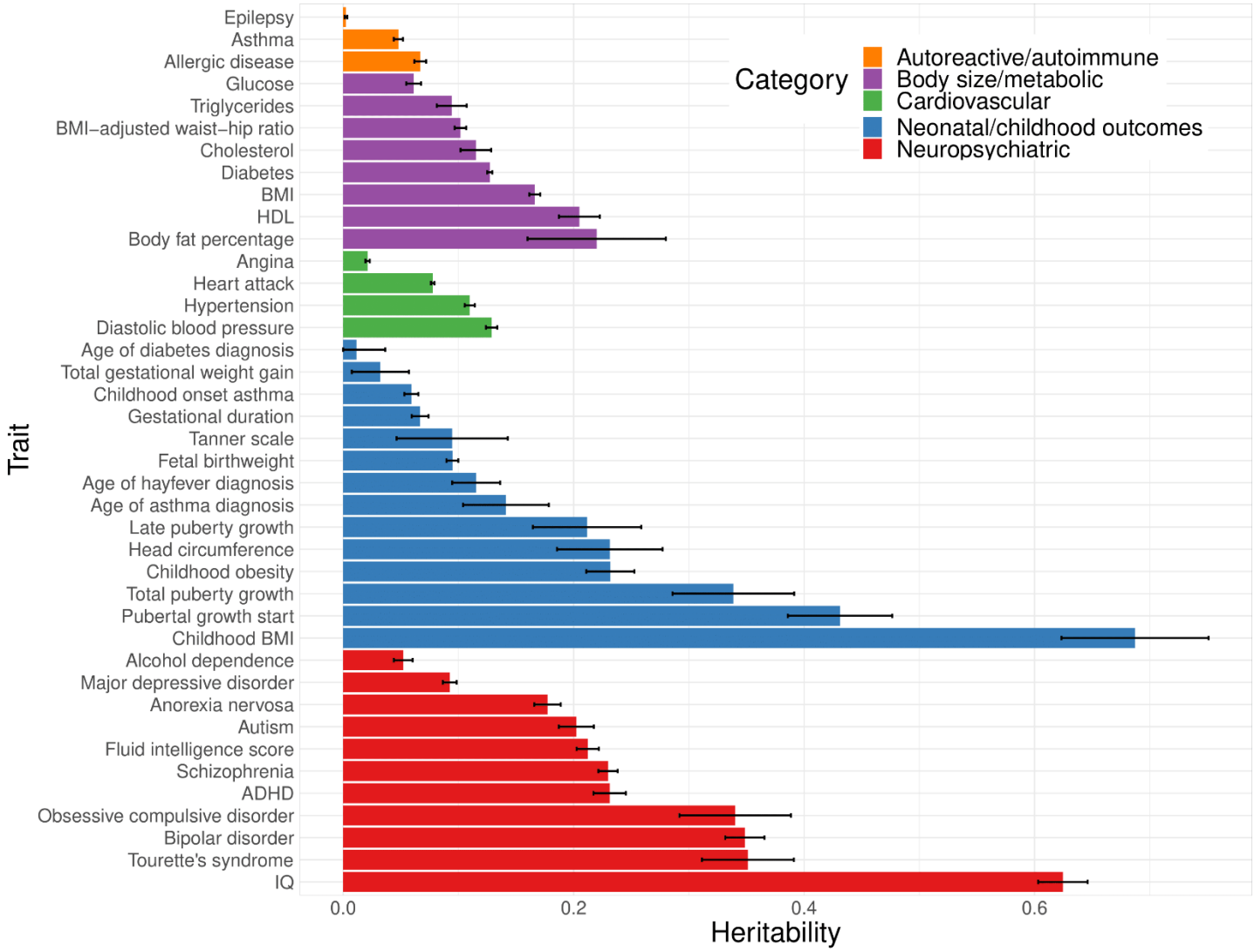

Figure S1: SNP heritability of 40 traits. Estimates of SNP heritability with 95% confidence interval (X-axis), grouped and colored by trait category (Y-axis).

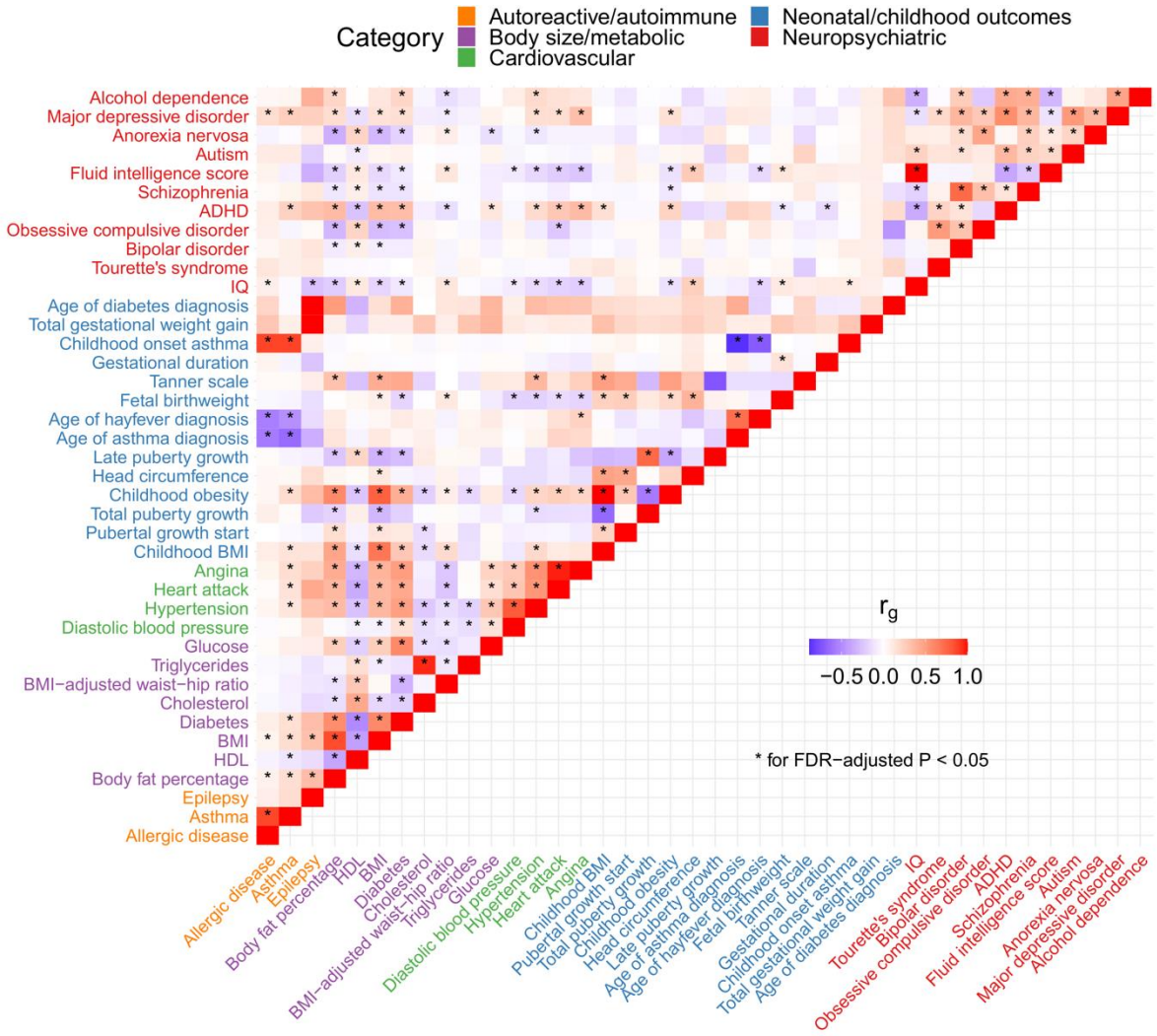

Figure S2: SNP-based genetic correlation between 40 traits. Heatmap of estimates of SNP-based genetic correlated between traits, grouped and colored by trait category. Correlations are marked with an asterisk are significantly non-zero with FDR-adjusted $P < 0.05$.

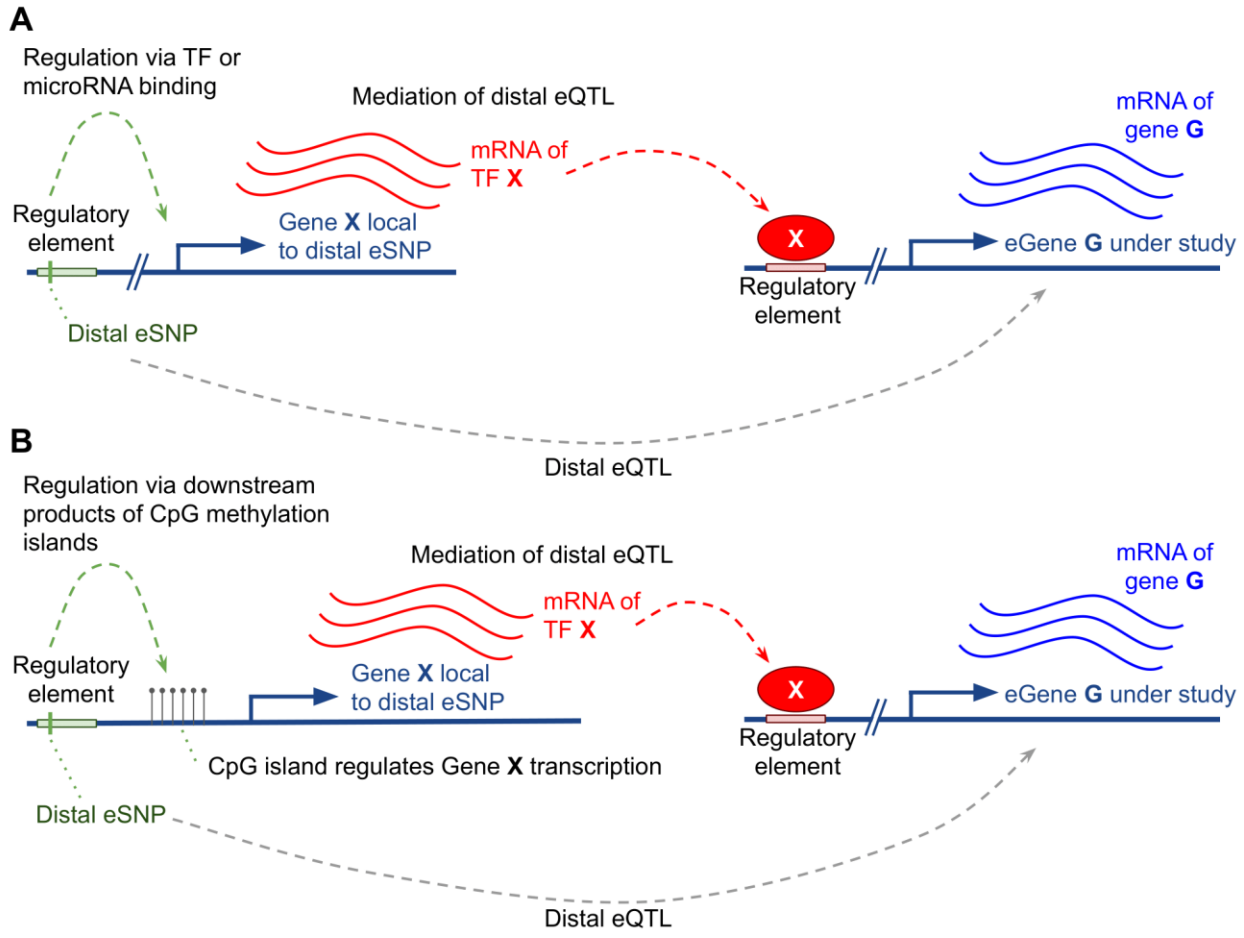

Figure S3: **Example of a biological mechanism MOSTWAS leverages in its predictive models**. Here, assume a SNP (in green) within a regulatory element affects the transcription of gene X (A) or the hyper- or hypomethylation of a CpG island upstream of gene X (B) that codes for a transcription factor 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.

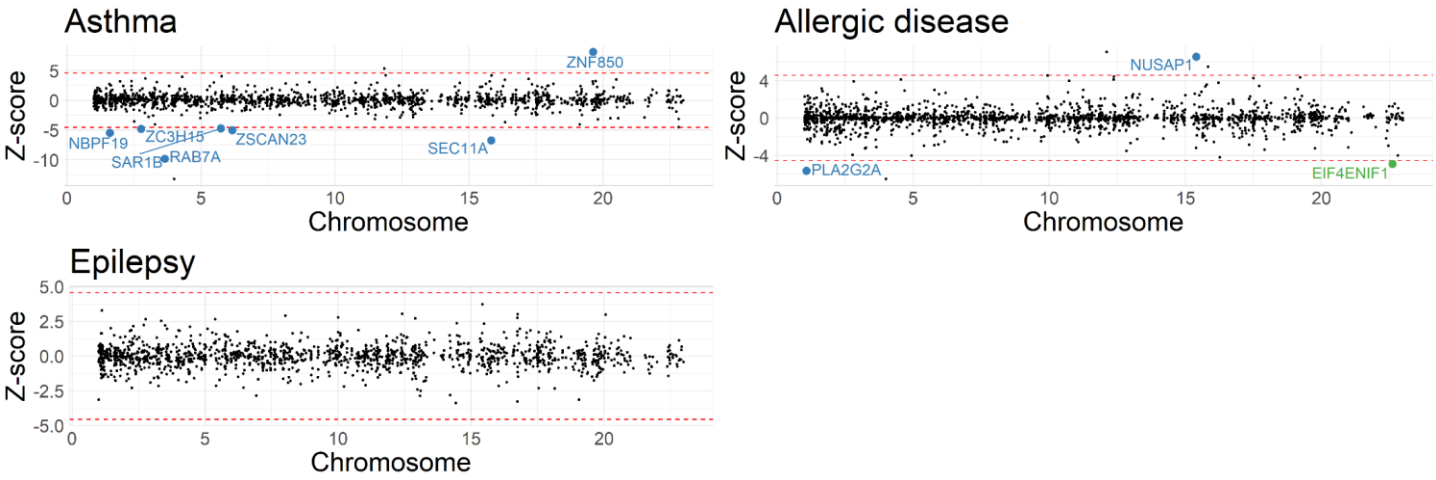

Figure S4: TWAS Miami plots for autoimmune/autoreactive disorders. Weighted Z-scores for TWAS associations (Y-axis) over genomic location of genes (X-axis). Red lines show Z-scores corresponding to $P < 2.5 \times 10^{-6}$. Genes labelled have $P < 2.5 \times 10^{-6}$, nominal permutation P <0.05, and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

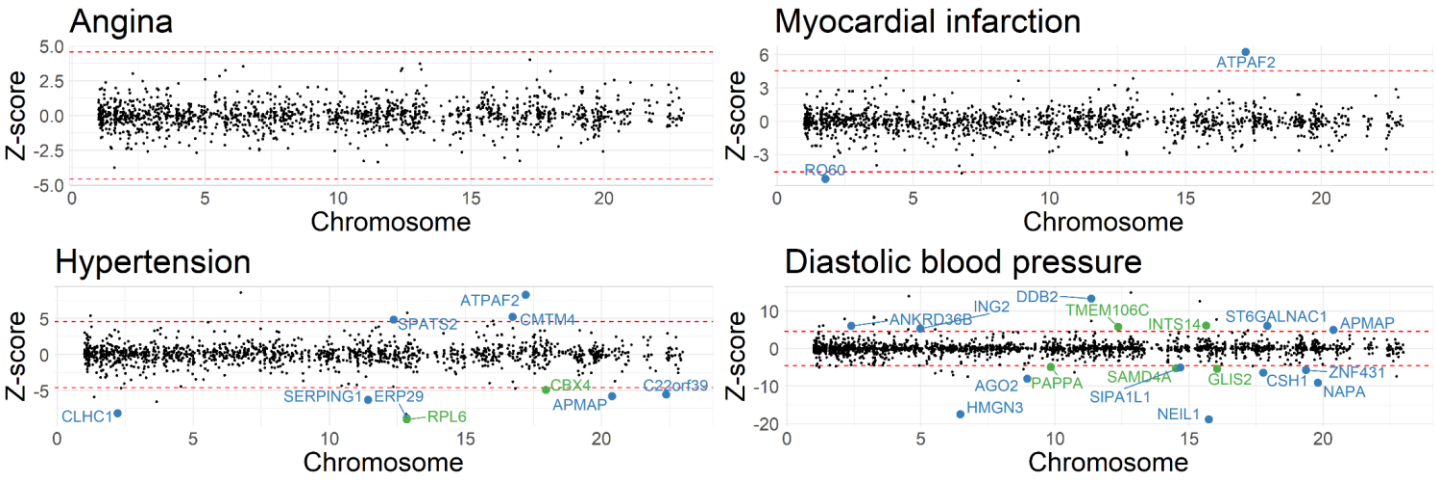

Figure S5: TWAS Miami plots for cardiovascular disorders. Weighted Z-scores for TWAS associations (Y-axis) over genomic location of genes (X-axis). Red lines show Z-scores corresponding to $P < 2.5 \times 10^{-6}$. Genes labelled have $P < 2.5 \times 10^{-6}$, nominal permutation P <0.05, and genes in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

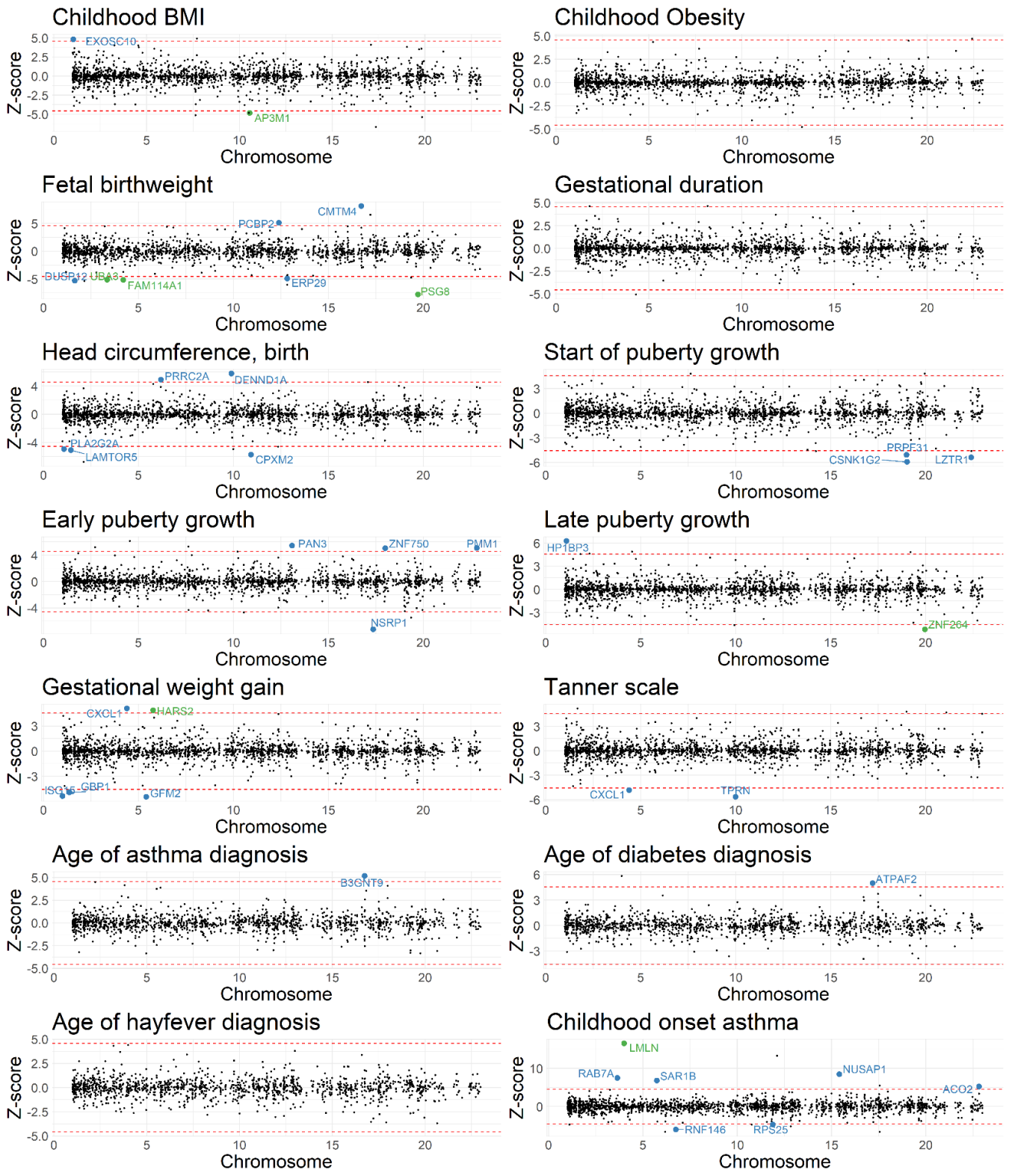

Figure S6: TWAS Miami plots for neonatal/childhood outcomes. Weighted Z-scores for TWAS associations (Y-axis) over genomic location of genes (X-axis). Red lines show Z-scores corresponding to $P < 2.5 \times 10^{-6}$. Genes labelled have $P < 2.5 \times 10^{-6}$, nominal permutation P <0.05, and genes in green showed Benjamini-Hochberg FDR-adjusted P < 0.05 for the distal-SNPs added-last test.

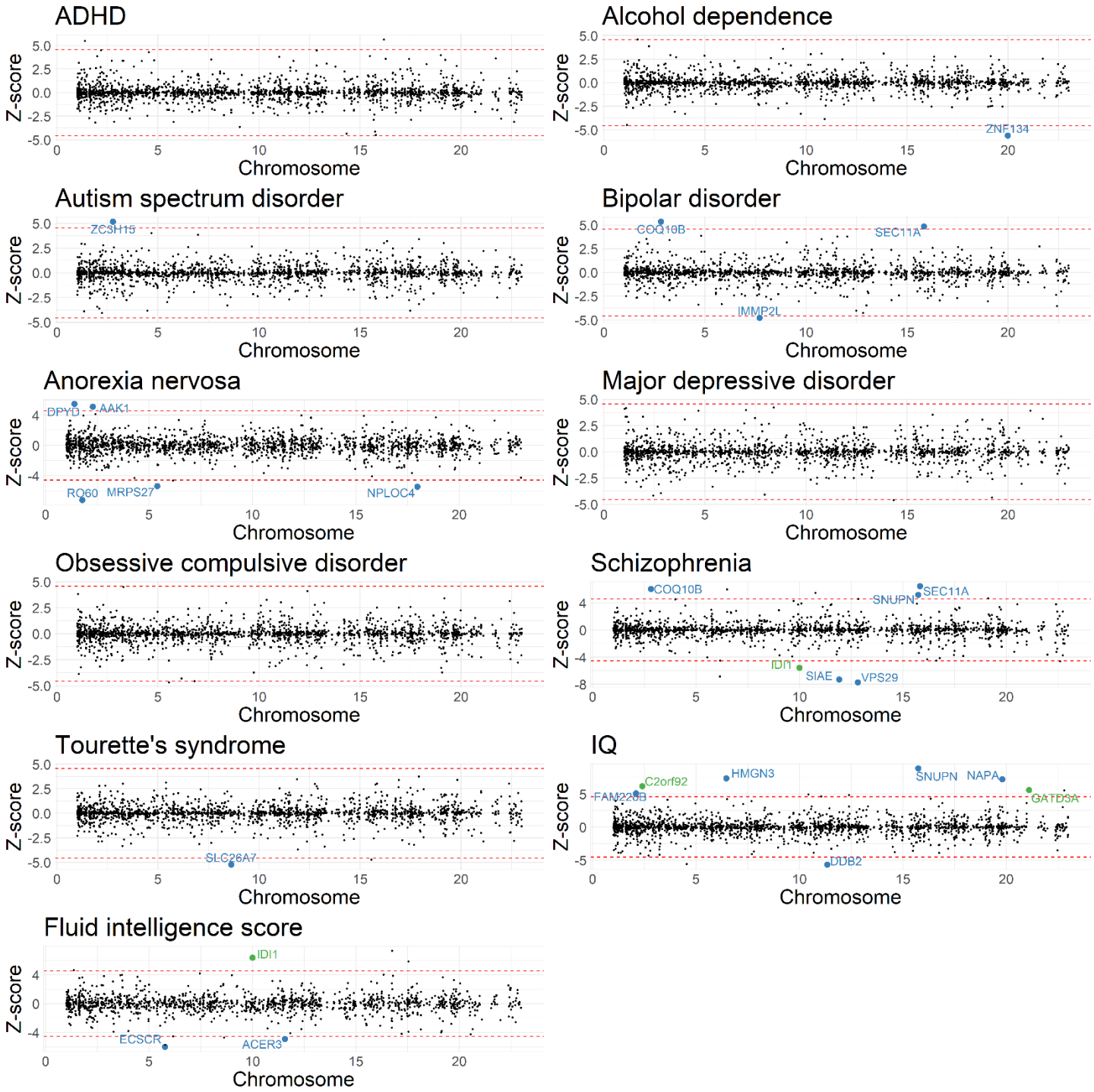

Figure S7: TWAS Miami plots for neuropsychiatric outcomes. Weighted Z-scores for TWAS associations (Y-axis) over genomic location of genes (X-axis). Red lines show Z-scores corresponding to $P < 2.5 \times 10^{-6}$. Genes labelled have $P < 2.5 \times 10^{-6}$, nominal permutation P <0.05, and genes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

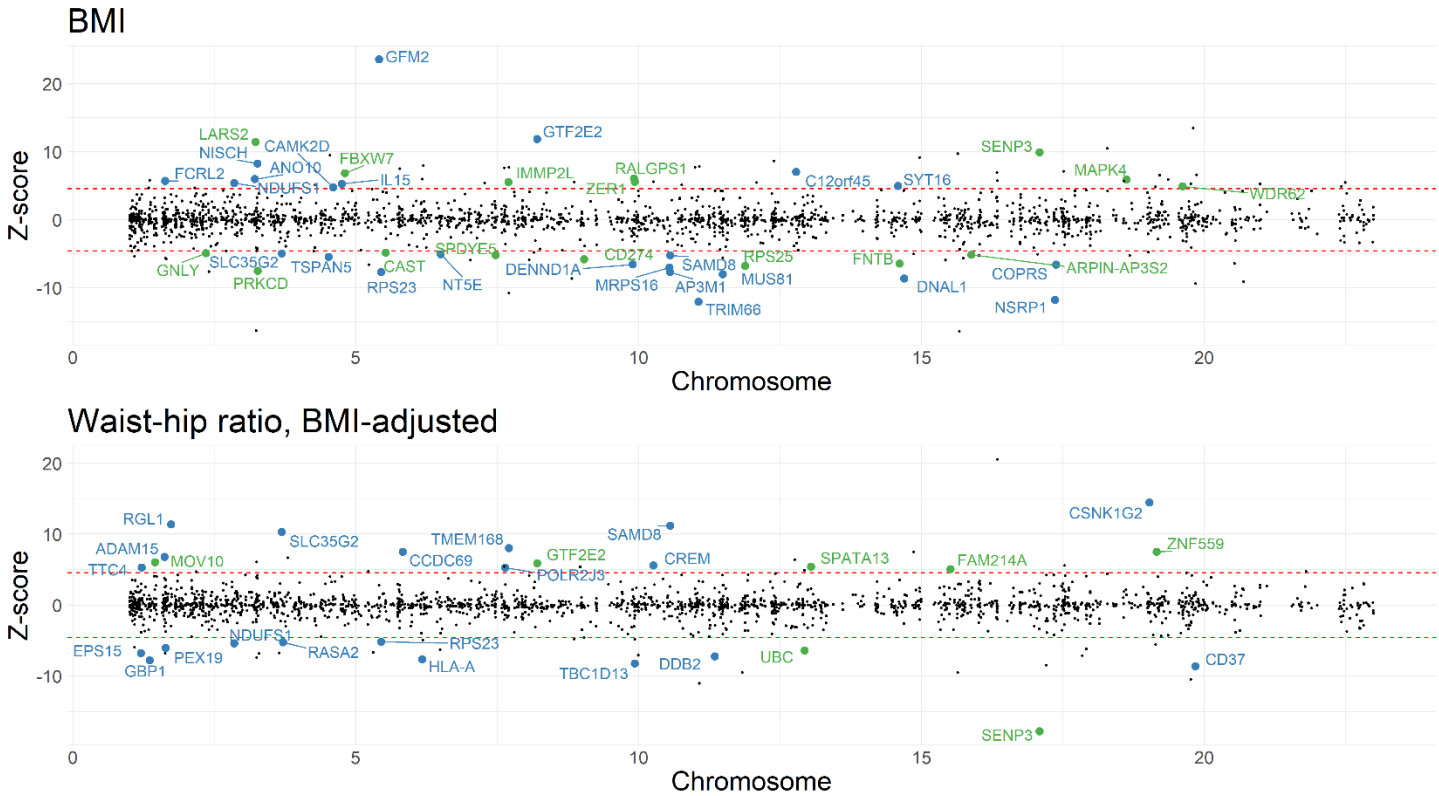

Figure S8: TWAS Miami plots for BMI and BMI-adjusted waist-hip ratio. Weighted Z-scores for TWAS associations (Y-axis) over genomic location of genes (X-axis). Red lines show Z-scores corresponding to P < 2.5×10^{-6}. Genes labelled have P < 2.5×10^{-6}, nominal permutation P <0.05, and genes in green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

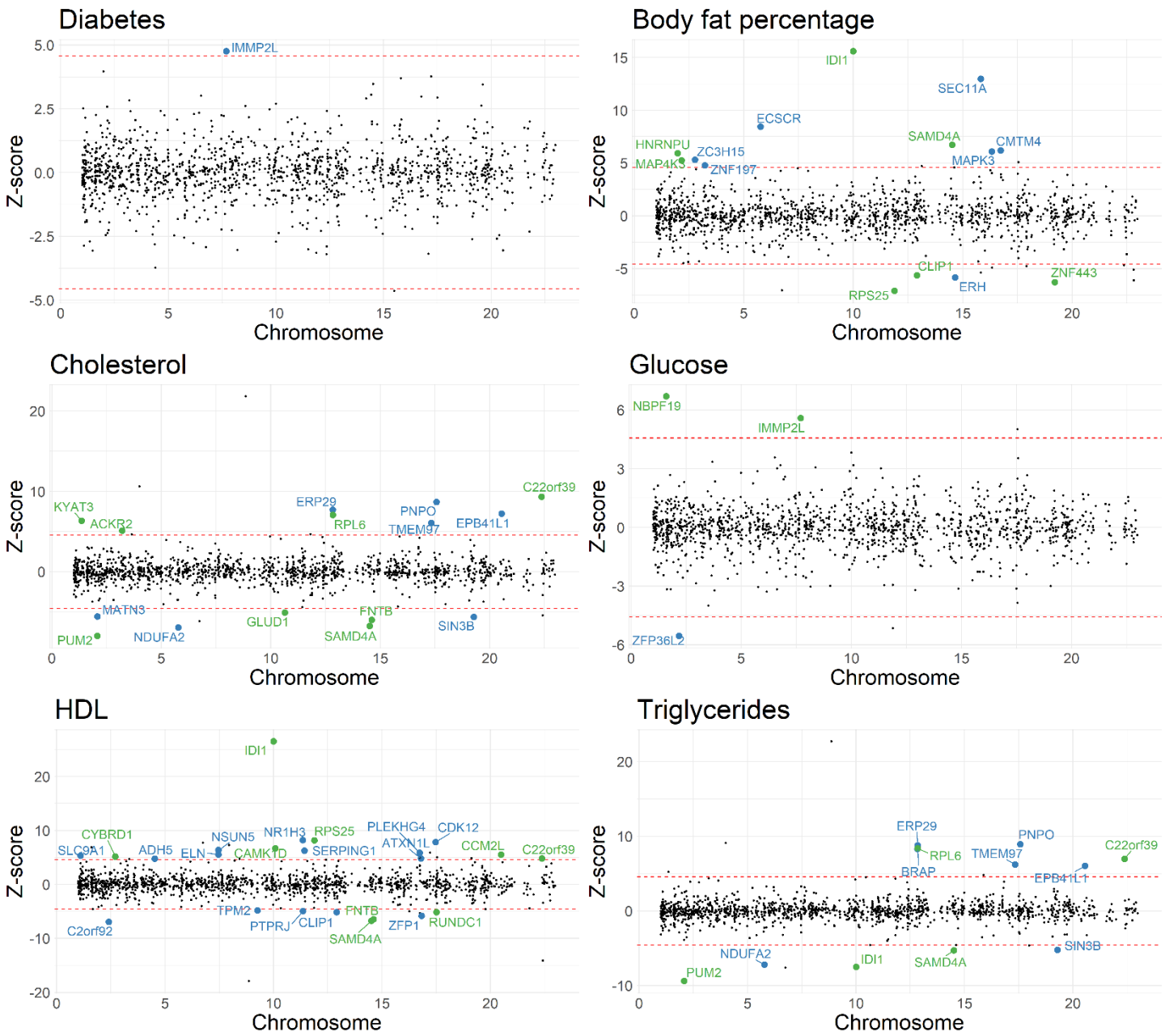

Figure S9: TWAS Miami plots for body size/metabolic traits, excluding BMI and BMI-adjusted waist-hip ratio. Weighted Z-scores for TWAS associations (Y-axis) over genomic location of genes (X-axis). Red lines show Z-scores corresponding to $P < 2.5 \times 10^{-6}$. Genes labelled have $P < 2.5 \times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

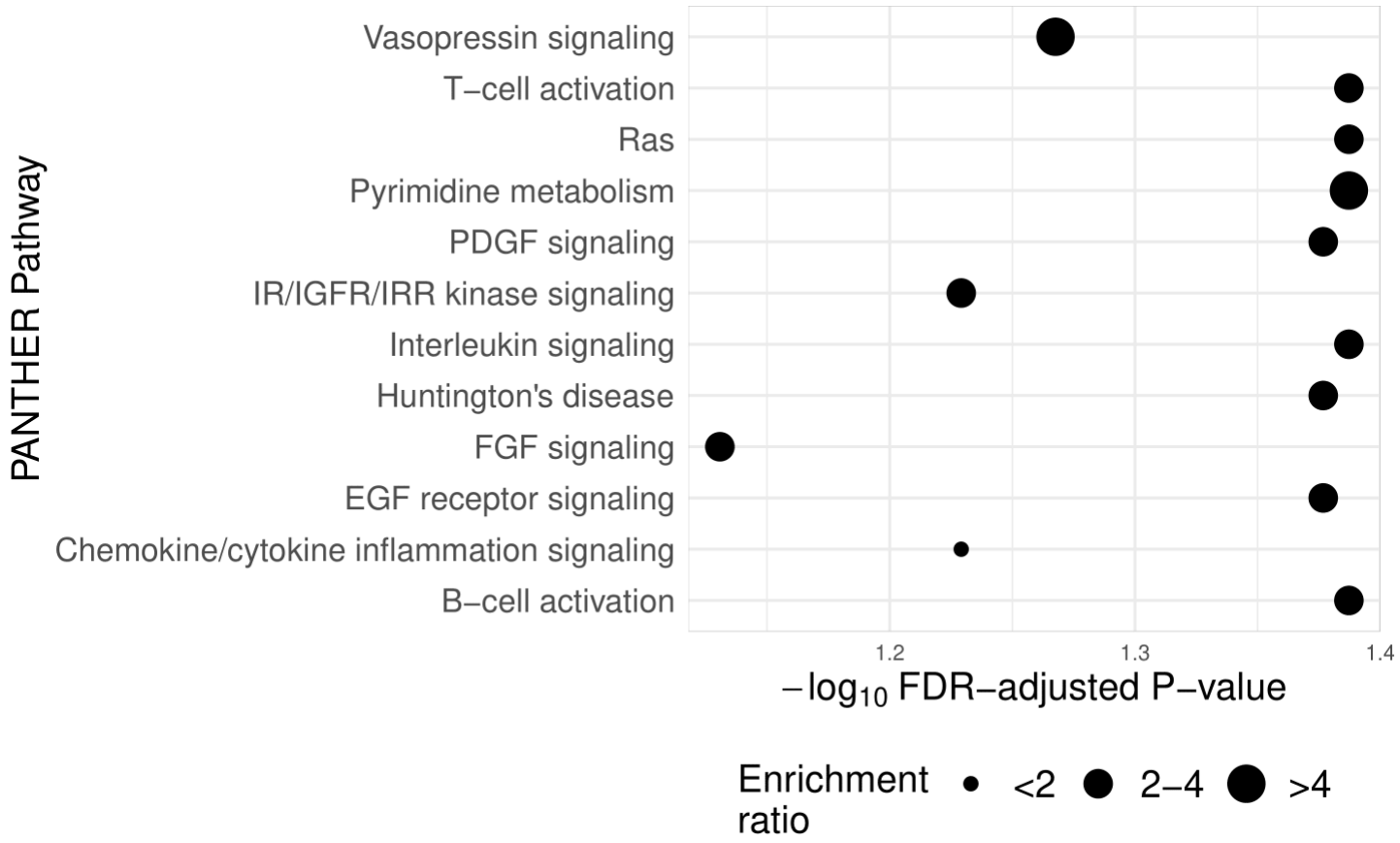

Figure S10: Gene set enrichments of TWAS-identified genes. Enrichment plot of over-representation in 176 TWAS genes of PANTHER pathways (Y-axis) with -$\log_{10}$ FDR-adjusted P-value (X-axis). The size of the point gives the relative enrichment ratio for the given pathway.

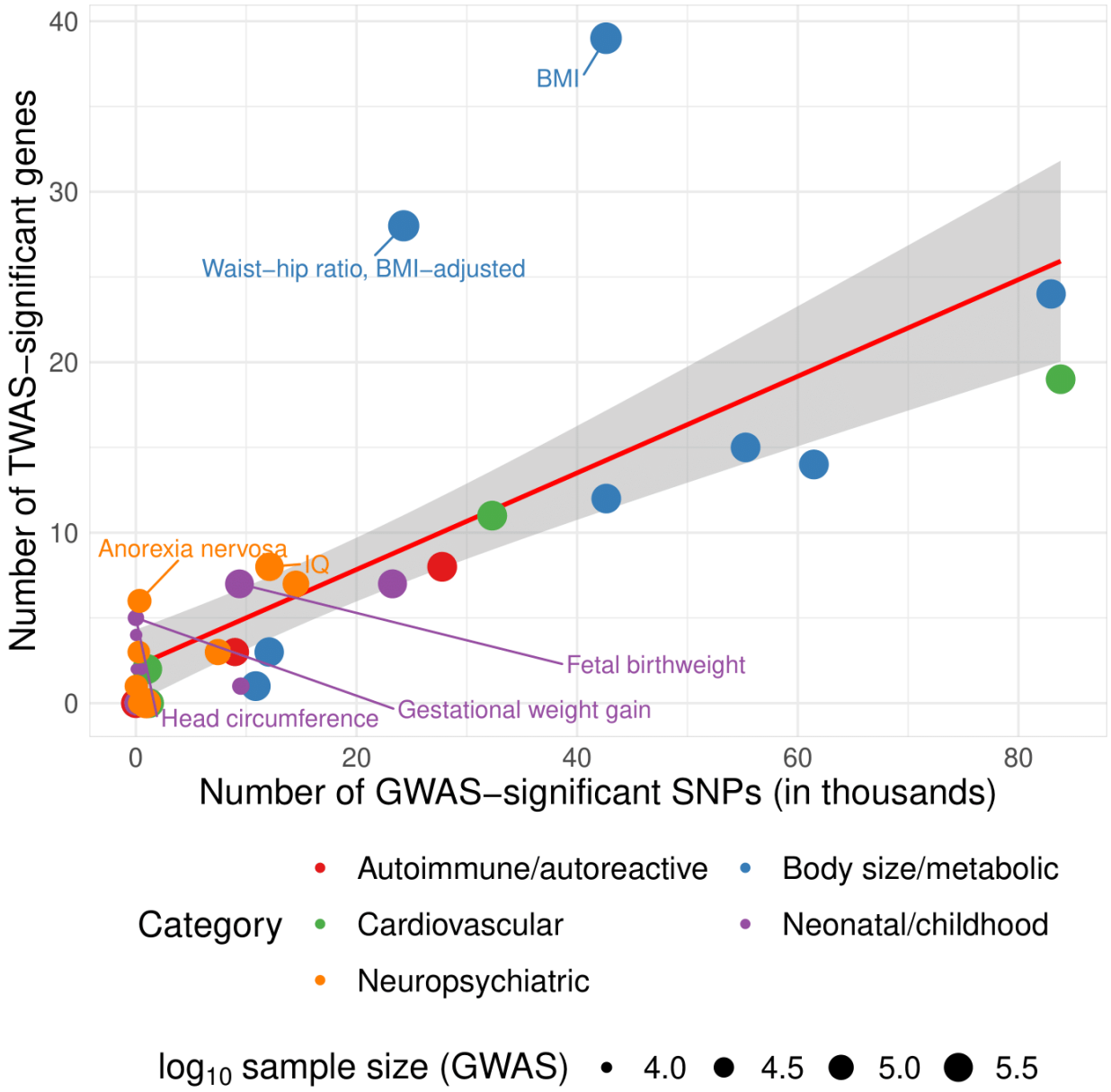

Figure S11: Comparison of GWAS and TWAS results across all 40 traits. Scatterplot of number of TWAS-significant genes (Y-axis) and number of GWAS-significant SNPs (X-axis) across all 40 traits, colored by category of the trait. The size of the point shows the $\log_{10}$ sample size of the GWAS. The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

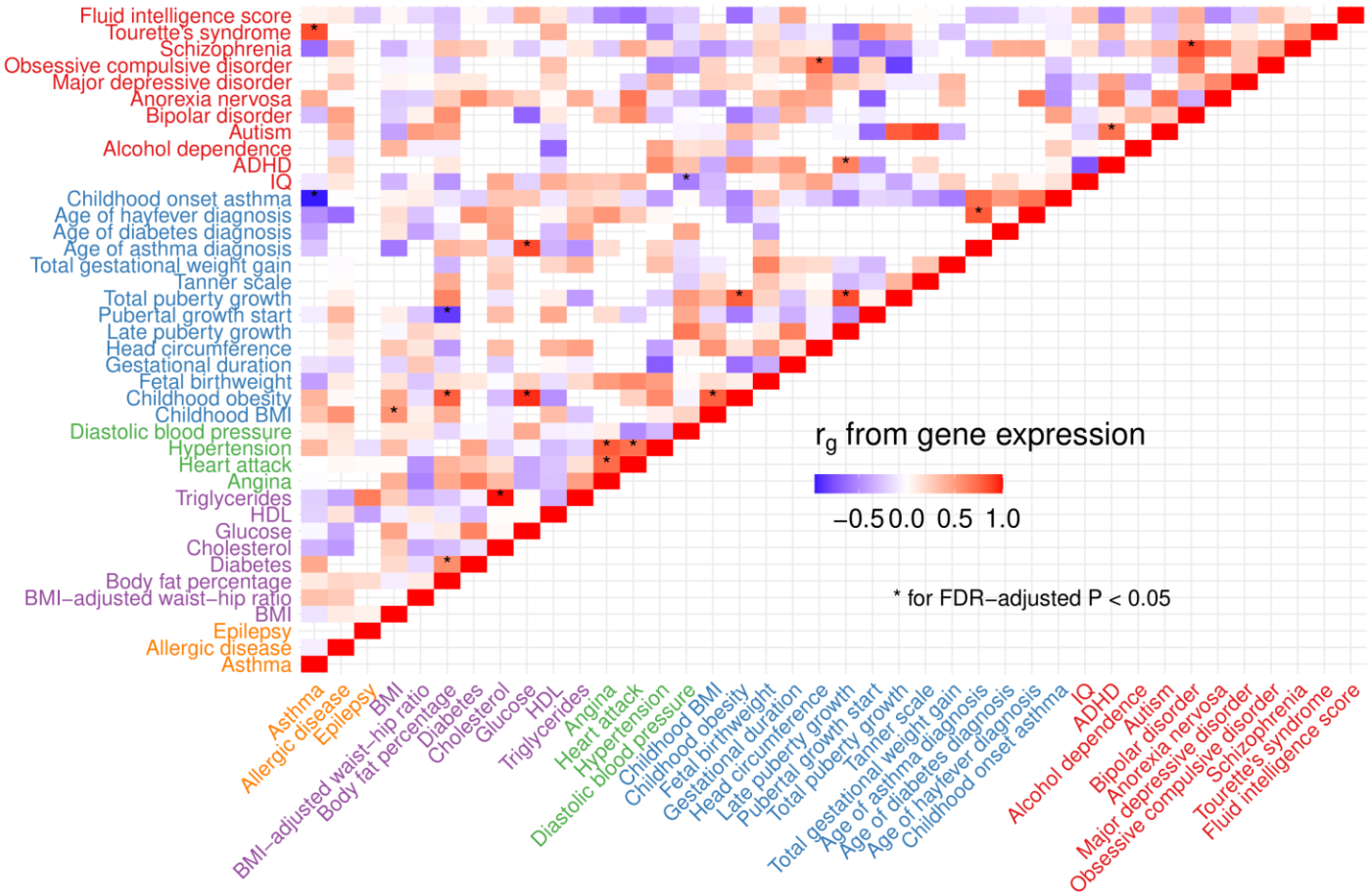

Figure S12: Heatmap of genetic correlations on the heritable gene expression level between 40 traits considered in TWAS analysis. Genetic correlations between traits at the level of the predicted expression of heritable genes. Correlations at FDR-adjusted P <0.05 are marked with an asterisk. Autoimmune/autoreactive 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.

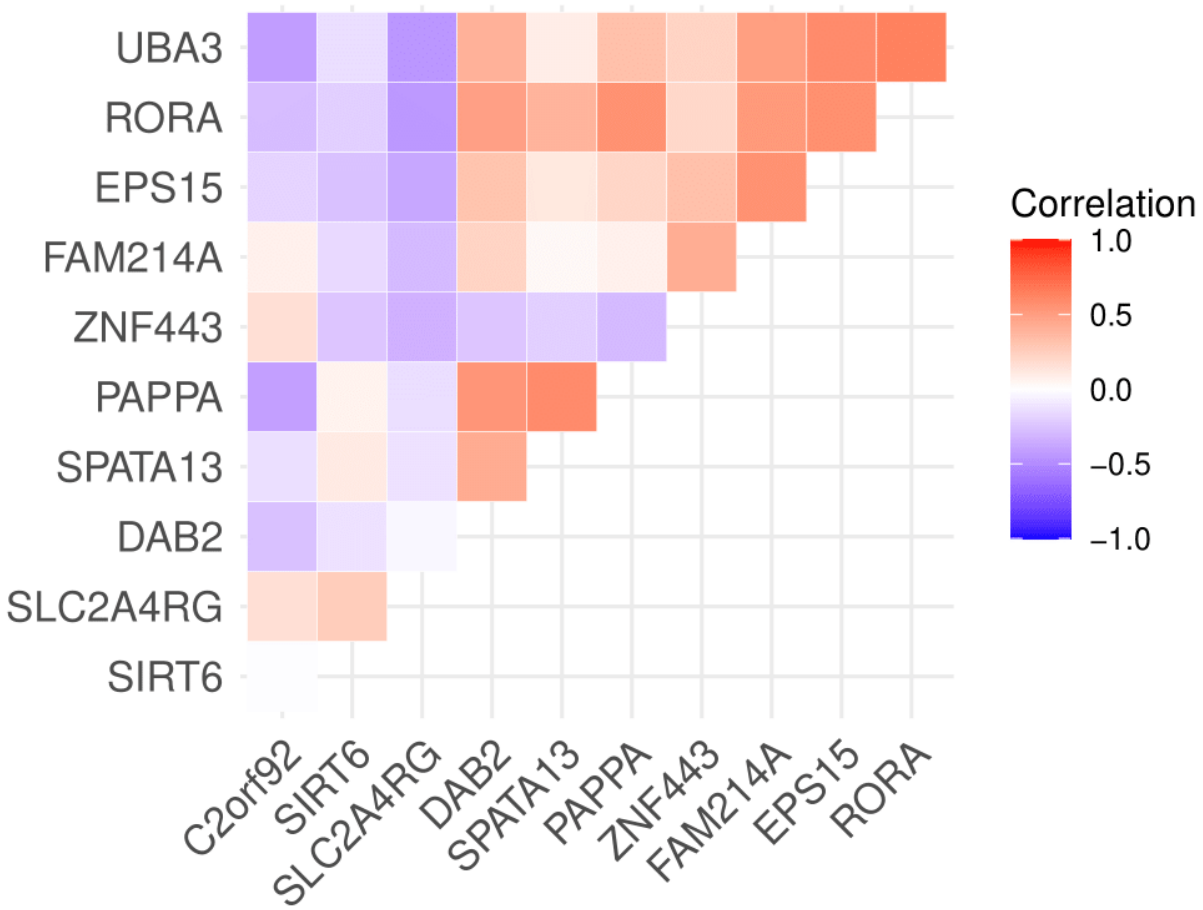

Figure S14: Heatmap of correlations between select transcription factor and TWAS-identified genes in RICHS. Correlations between the RICHS expression of TFs (Y-axis) and associated TWAS genes identified by MOSTWAS in ELGAN (X-axis).

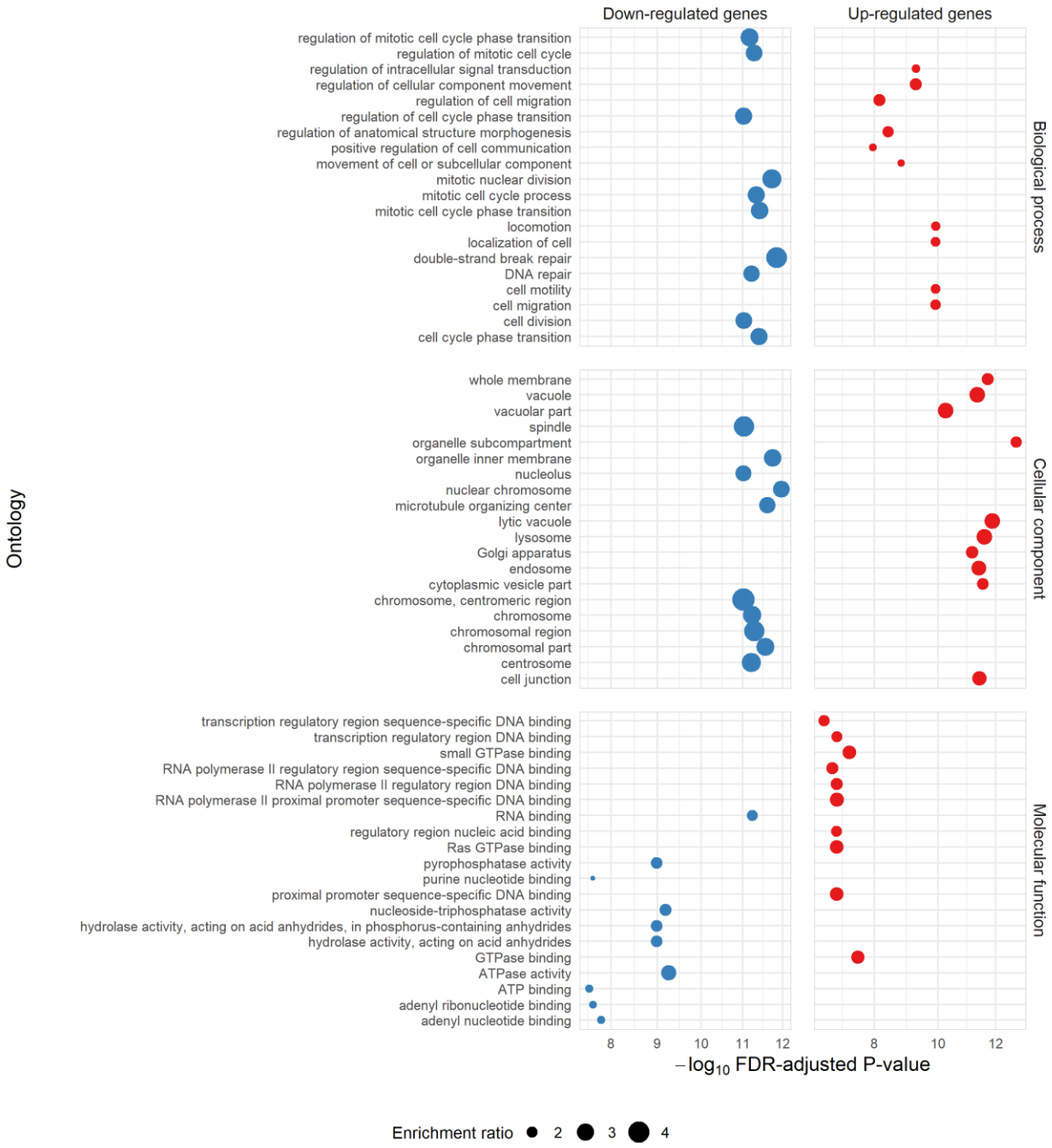

Figure S15: Over-representation enrichments of differentially expressed genes in *EPS15* knockdown. Enrichment plot of over-representation of biological process, cellular component, and molecular function ontologies (Y-axis) with -$\log_{10}$ FDR-adjusted P-value (X-axis). The size of the point gives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