Genetic Associations with Dynamic Placental Proteins Identify Causal Biomarkers for Hypertension in Pregnancy

Qi Yan[1,*], Nathan R. Blue[2], Buu Truong[3], Yu Zhang[4], Rafael F. Guerrero[5], Nianjun Liu[4], Michael C. Honigberg[3,6], Samuel Parry[7], Rebecca B. McNeil[8], Brian M. Mercer[9], William A. Grobman[10], Robert Silver[2], Uma M. Reddy[1], Ronald J. Wapner[1], David M. Haas[11]

[1]Department of Obstetrics and Gynecology, Columbia University Irving Medical Center, New York, NY 10032,

[2]Department of Obstetrics and Gynecology, University of Utah, Salt Lake City, UT 84132,

[3]Cardiovascular Research Center, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114,

[4]Department of Epidemiology and Biostatistics, Indiana University Bloomington, Bloomington, IN 47405,

[5]Department of Biological Sciences, North Carolina State University, Raleigh, NC 27695,

[6]Cardiology Division, Department of Medicine, Massachusetts General Hospital, Boston, MA 02114,

[7]Department of Obstetrics and Gynecology, University of Pennsylvania, Philadelphia, PA 19104,

[8]RTI International, Research Triangle Park, NC 27709,

[9]Department of Obstetrics and Gynecology, Case Western Reserve University, Cleveland, OH 44109,

[10]Department of Obstetrics and Gynecology, College of Medicine, The Ohio State University, Columbus, OH 43210,

[11]Department of Obstetrics and Gynecology, Indiana University, Indianapolis, IN 46202, USA

[*]Corresponding authors:
Qi Yan, Ph.D.
Assistant Professor
Department of Obstetrics and Gynecology
Columbia University Irving Medical Center
212-305-8738
qy2253@cumc.columbia.edu

ABSTRACT

Preeclampsia (PE), a gestational hypertensive disorder, ranks as the second leading cause of maternal mortality worldwide. While PE is considered a multifactorial disease, placental insufficiency is believed to drive its progression. To noninvasively study placental physiology related to adverse pregnancy outcomes (APOs) and predict these outcomes before symptom onset, we measured nine placental protein levels in first- and second-trimester serum samples from 2,352 nulliparous pregnant women in the Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be (nuMoM2b) study. The proteins analyzed include VEGF, PlGF, ENG, sFlt-1, ADAM-12, PAPP-A, fβHCG, INHA, and AFP. Currently, little is known about the genetic variants contributing to the heritability of these proteins during pregnancy, and no studies have explored the causal relationships between early pregnancy proteins and gestational hypertensive disorders. Our study has three objectives. First, we conducted genome-wide association study (GWAS) of nine placental proteins in maternal serum during the first and second trimesters and the difference between time points to understand how genetics may influence placental proteins in early pregnancy. Second, we examined whether early pregnancy placental proteins are causal factors for PE and gestational hypertension (gHTN). Lastly, we investigated the causal relationship between PE/gHTN and long-term HTN. In conclusion, our study discovered significant genetic associations with placental proteins ADAM-12, VEGF, and sFlt-1, offering insights into their regulation during pregnancy. Mendelian randomization (MR) analyses demonstrated evidence of causal relationships between placental proteins, particularly ADAM-12, and gHTN, potentially informing prevention and treatment strategies. Our findings suggest that placental proteins like ADAM-12 could serve as biomarkers for postpartum HTN risk.

INTRODUCTION

Preeclampsia (PE), as a gestational hypertensive disorder, is the second-largest cause of maternal mortality worldwide[1]. Although PE is likely a multifactorial disease, it is believed that poor placental implantation is responsible for the disease[1]. The villous structure of the placenta facilitates gas and nutrient exchange between the mother and fetus, and placental products are secreted into the maternal circulation from the syncytiotrophoblast villous surface. However, assessing placental health before the onset of PE is difficult due to the inaccessibility of placental tissue during pregnancy. To non-invasively study placental physiology concerning adverse pregnancy outcomes (APOs) and to further predict these outcomes before the onset of symptoms, the levels of nine placental proteins were measured in first- and second-trimester serum from 2,352 nulliparous pregnant women using the large Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be (nuMoM2b) study. The proteins measured comprise vascular endothelial growth factor (VEGF), placental growth factor (PlGF), endoglin (ENG), soluble FMS-like tyrosine kinase-1 (sFlt-1), a disintegrin and metalloproteinase domain-containing protein 12 (ADAM-12), pregnancy-associated plasma protein A (PAPP-A), free beta-human chorionic gonadotropin (fβHCG), inhibin A (INHA), and alpha fetal protein (AFP). They fall into three categories that reflect different aspects of placental function: (1) angiogenesis (VEGF, PlGF, ENG, and sFlt-1)[2-9]; (2) placental implantation and development (ADAM-12 and PAPP-A)[10,11]; and (3) biomarkers for fetal chromosomal abnormalities (fβHCG, INHA, and AFP)[12,13]. Our previous study has demonstrated the association of the levels of these proteins with PE[14]. Gaining a better understanding of the genetic predisposition to the levels of these proteins could facilitate the development of new approaches for predicting and preventing APOs, such as PE. However, little is currently known about the genetic variants contributing to the heritability of these proteins during pregnancy. Previous studies have examined the effect of certain candidate single nucleotide polymorphisms (SNPs) on PAPP-A level[15,16], and a genome-wide association study (GWAS) of PlGF levels was performed but excluded pregnant women[17]. Similarly, *VEGFA* and *VLDLR* have been repeatedly identified to be associated with VEGF levels in non-pregnant populations, but genetic effects on VEGF levels during pregnancy have never been studied[18-21]. In this study, we performed a comprehensive genetic analysis of nine placental proteins measured in maternal serum during the first (visit 1) and second (visit 2) trimesters, as well as the difference between the two time points (visit 2-1), among nuMoM2b pregnant women.

Mendelian randomization (MR) provides a means to estimate the causal effect of an exposure on an outcome by using genetic variants as instrumental variables (IVs), while minimizing confounding effects and reversed causation in observational studies[22]. Since genetic variants are randomly assigned from parents to offspring, they are typically unrelated to the presence of confounders. In many cases, the exposure and outcome are available in different datasets, which is referred to as two-sample MR. The major advantage of using two-sample MR is that it only requires GWAS summary statistics, rather than individual-level data. Moreover, two-sample MR is generally considered more conservative and unbiased than one-sample MR[23]. To date, there have been no MR studies investigating the causal relationships between early pregnancy proteins and gestational hypertensive disorders. In this study, we performed two-sample MR analyses to examine the causal effects of placental proteins on PE and gestational hypertension (gHTN). We extracted IV SNPs from GWAS of placental proteins using the nuMoM2b cohort, and obtained genetic associations of PE and gHTN from the most recent large GWAS[24].

Literature also indicates that PE and gHTN are associated with HTN, cardiovascular disease, and renal disease in women later in life[25-28]. Furthermore, few studies included biologic markers measured during pregnancy. A deeper understanding of the underlying mechanisms and the identification of risk markers will allow researchers to investigate the development of cardiovascular abnormalities following pregnancy. In this study, we conducted a two-sample MR analysis to examine the causal relationship between PE/gHTN and long-term HTN. We leveraged the large GWAS of PE/gHTN and integrated it with genetic results of HTN measured several years after first delivery in the nuMoM2b Heart Health Study (nuMoM2b-HHS), in which 4,484 of the 10,038 women enrolled in nuMoM2b were followed 2-7 years (mean 3.2 years) after delivery and had laboratory data[29].

This study has three objectives. First, to fill the knowledge gap on how genetics might affect placental proteins in early pregnancy, we performed GWAS of nine placental proteins in maternal serum during visit 1 and visit 2, as well as the change in these proteins between those visits. Second, we investigated whether placental proteins in early pregnancy are causal factors for PE and gHTN. Third, we examined the causal relationship between PE/gHTN and long-term HTN.

RESULTS

Data Characteristics

Tables 1 and 2 shows the characteristics of the 2,352 nulliparous women included in the GWAS of placental proteins, with 2,263 for visit 1, 2,134 for visit 2, and 2,045 for visit 2-1. ADAM-12, AFP, PAPP-A, PlGF, and VEGF showed a significant increase in serum levels from visit 1 to visit 2, while ENG, fβHCG, and INHA showed a significant decrease. There was no significant difference in sFlt-1 levels between the two time points. The mean age at visit 1 was 26.8±5.8, and most of the participants self-reported as White (57.7%), with 16.5% Black, 16.8% Hispanic, 3.4% Asian, and 5.7% from other groups (Table 1). This subset for placental protein analyses consisted of a random sample of 11.8% of all nuMoM2b women and all women with any APOs (Figure 1A). Overall, 62.2% of this subsample had at least one APO (23.5% had PE or eclampsia (PE), 33.0% had any delivery prior to 37 weeks' gestation (PTB), 19.2% had spontaneous PTB (sPTB), 2.0% had delivery of a stillborn baby (SB), and 16.8% had birth weight for gestational age <5th percentile (SGA)).

Table 1. Summary of women characteristics included in the GWAS of placental proteins

		Total (n=2,352)
Age at visit 1 (mean±SD)		26.77±5.84
Self-reported race		
	White	1,356 (57.65%)
	Black	389 (16.54%)
	Hispanic	394 (16.75%)
	Asian	80 (3.4%)
	Other	133 (5.65%)
APO status		1,463 (62.2%)
	PE	552 (23.47%)
	PTB	776 (32.99%)
	sPTB	451 (19.18%)
	SB	48 (2.04%)
	SGA	396 (16.84%)

Table 2. Summary of the nine placental proteins (mean±SD) included in the GWAS

	visit 1 (n=2,263)	**visit 2 (n=2,134)**	**visit 2-1 (n=2,045)**	***P*-value*** **(visit 2-1)**
ADAM-12 log(ng/mL)	1.49±0.43	2.25±0.34	0.77±0.36	<0.0001
AFP log(IU/mL)	2.53±0.71	3.81±0.48	1.29±0.69	<0.0001
ENG log(ng/mL)	1.83±0.22	1.7±0.23	-0.14±0.17	<0.0001
fβHCG log(ng/mL)	2.99±0.62	1.39±0.75	-1.6±0.59	<0.0001
INHA log(pg/mL)	5.75±0.49	5.35±0.44	-0.39±0.44	<0.0001
PAPP-A log(mU/mL)	6.74±1.05	9.05±0.8	2.33±0.9	<0.0001
PlGF log(pg/mL)	3.68±0.56	5.18±0.66	1.5±0.61	<0.0001
sFlt-1 log(pg/mL)	6.77±0.43	6.77±0.48	0±0.33	0.57
VEGF log(pg/mL)	0.18±0.96	0.27±0.88	0.17±0.77†	<0.0001

* *P*-value represents the difference between visit 2 and 1 on the logarithm scale (i.e., H_0: visit 2-1 = 0)

† There is a discrepancy between the mean of VEGF at visit 2-1 and the mean difference between visit 2 and visit 1. This discrepancy is attributed to the fact that individuals who had VEGF levels measured at both visits 1 and 2 showed a much higher mean VEGF level at visit 2 (mean = 0.37) compared to all individuals who had their VEGF levels measured at visit 2 (mean = 0.27).

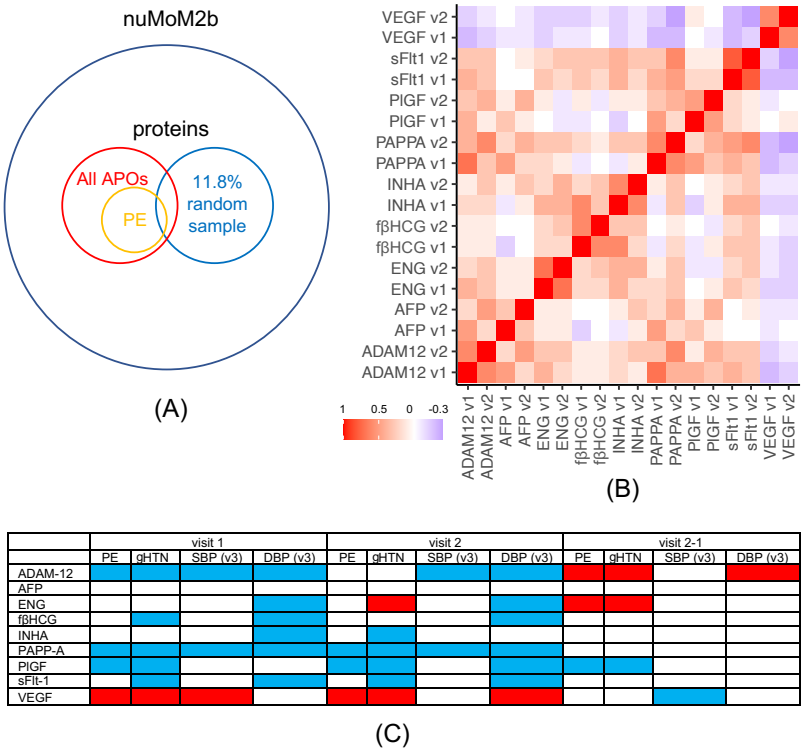

Figure 1. Overview of nine placental proteins in nuMoM2b. (A) A subset for placental protein analyses consisted of a random sample of 11.8% of all nuMoM2b women and all women with any APOs. (B) The correlation among placental proteins measured during visits 1 and 2. (C) The observational association between placental proteins (measured at visit 1, visit 2, and visit 2-1) and PE, gHTN, visit 3 systolic blood pressure (SBP), and visit 3 diastolic blood pressure (DBP), respectively, using the nuMoM2b cohort. A red color indicates a positive association and a blue color indicates an inverse association.

GWAS of placental proteins

We conducted GWAS on nine placental proteins measured at visits 1 and 2, as well as their difference (Figure S1). The QQ plot for all GWAS showed no inflation of the test statistics (Figure S2). A significant association was observed between ADAM-12 levels at visit 2 and a locus on chromosome 12p13.31 ($P < 5.6\times10^{-9}$; Figure 2C). The SNP rs6487735, located near *PZP* (123 kb from the gene promoter), showed the strongest association (test allele: T, minor allele frequency (MAF) = 0.47, effect size = -0.1 log(ng/mL), $P = 3.03\times10^{-22}$). A likely causal missense variant, rs2277413 in *PZP*, was independent of rs6487735 (linkage disequilibrium (LD) $r^2 = 0.003$, Figure 2C). This missense SNP was also significantly associated with ADAM-12 levels at visit 2 (test allele: G, MAF = 0.3, effect size = 0.07 log(ng/mL), $P = 3.32\times10^{-11}$) and showed the strongest association after additionally adjusting for rs6487735 ($P = 2.37\times10^{-10}$). The GWAS of ADAM-12 at visit 1 also identified the same *PZP* locus, which was marginally significant (Figure 2B). The top SNP, rs4316551 (test allele: T, MAF = 0.35, effect size = -0.07 log(ng/mL), $P = 6.76\times10^{-9}$), was in moderate LD ($r^2 = 0.32$) with rs6487735, which was also significant at the visit 2 analysis (effect size = -0.08 log(ng/mL), $P = 1.53\times10^{-13}$). Both rs6487735 (effect size = -0.06 log(ng/mL), $P = 2.62\times10^{-7}$) and rs2277413 (effect size = 0.06 log(ng/mL), $P = 1.06\times10^{-5}$) showed the same effect directions at visit 1 as at visit 2, although they did not reach the genome-wide significance level. In general, the GWAS results suggested that *PZP* was associated with ADAM-12 levels in maternal serum, and this association increased as pregnancy progressed. In sensitivity analyses, we repeated the GWAS of ADAM-12 using only those of self-reported as Whites to compare our mixed-ancestry findings with results from a more genetically homogeneous population, and obtained similar but less significant results compared to using all available individuals (Figure S3 and Table S1). In this context, Whites were treated as a proxy of European ancestry. The principal component analysis (PCA) using genotypic data demonstrated that the vast majority of Whites share a common ancestry (Figure S4). Separate analyses for other populations were not feasible due to insufficient sample sizes. Furthermore, the transcriptome-wide association analysis (TWAS) analysis leveraging eQTLs from GTEx v8 indicated that gene expression on chromosome 22q11.22, encompassing *ENSG00000279278* (pancreas: false discovery rate (*FDR*) $P = 3.81\times10^{-3}$; whole blood: *FDR* $P = 0.013$), *ENSG00000274422* (pancreas: *FDR* $P = 3.81\times10^{-3}$), *IGLV1-41* (vagina: *FDR* $P = 0.028$; whole blood: *FDR* $P = 0.012$), and *IGLV9-49* (whole blood: *FDR* $P = 0.027$), was associated with ADAM-12 levels at visit 2. These genes, located within the same

region, exhibited consistent effects in the pancreas, vagina, and whole blood, suggesting a common genetic influence on ADAM-12 levels across various tissues.

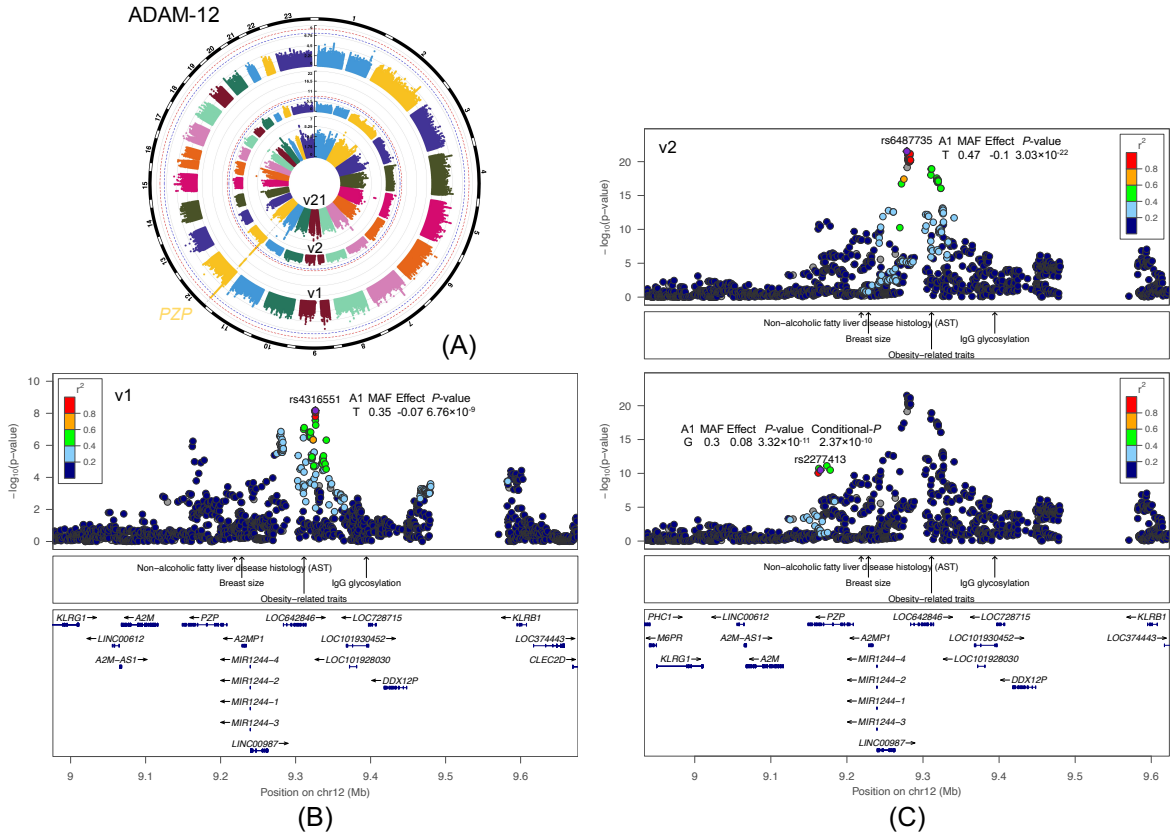

Figure 2. GWAS of ADAM-12. (A) Circular Manhattan plots. The chromosomal position of each single SNP is displayed along the circle and the negative log10 of the association *P*-value is displayed on the radius. The red line represents the genome-wide significance level ($P < 5.6\times10^{-9}$) and blue line represents the suggestive significance level ($P < 5\times10^{-8}$). Results for visit 1 are displayed on the outer circle, visit 2 on the middle circle, and visit 2-1 on the inner circle. (B) Regional plot for SNP rs4316551 from the visit 1 analysis. The lower portion of the figure displays the relative location of genes and the direction of transcription, while the middle portion shows known GWAS associations at the locus from the GWAS catalog. The x-axis displays the chromosomal position and the y-axis shows the significance of the associations. The purple diamond shows the *P*-value for the reference SNP. The circles show the *P*-values for all other SNPs and are color coded according to the level of LD with the reference SNP using the nuMoM2b cohort. (C) Regional plots for SNPs rs6487735 and rs2277413 from the visit 2 analysis.

Our GWAS analysis of VEGF identified rs6921438, located near *VEGFA* (187.7 kb away from the gene promoter) on chromosome 6q21.1, as the variant with the strongest association with VEGF levels at visit 1 (test allele: A, MAF = 0.45, effect size = -0.36 log(pg/mL), $P = 7.94\times10^{-30}$) and

visit 2 (MAF = 0.47, effect size = -0.32 log(pg/mL), $P = 2.49\times10^{-28}$) (Figure 3AD). Despite being located between *LINC01512* and *LOC101929705*, rather than immediately adjacent to *VEGFA*, this locus has consistently been replicated in multiple studies as being associated with circulating VEGF levels in non-pregnant populations[18-21]. Our study confirms the presence of *cis*-acting genetic associations with VEGF levels in nulliparous pregnant women. We subsequently investigated whether SNPs previously linked to VEGF levels in non-pregnant populations at *VEGFA* and *VLDLR* were also linked to VEGF levels in pregnant women. Our analysis showed that we were able to confirm the association with SNPs at *VEGFA*, but not *VLDLR* (see Table S2). These findings suggest that the genetic regulation of VEGF levels may differ between early pregnancy and non-pregnant periods.

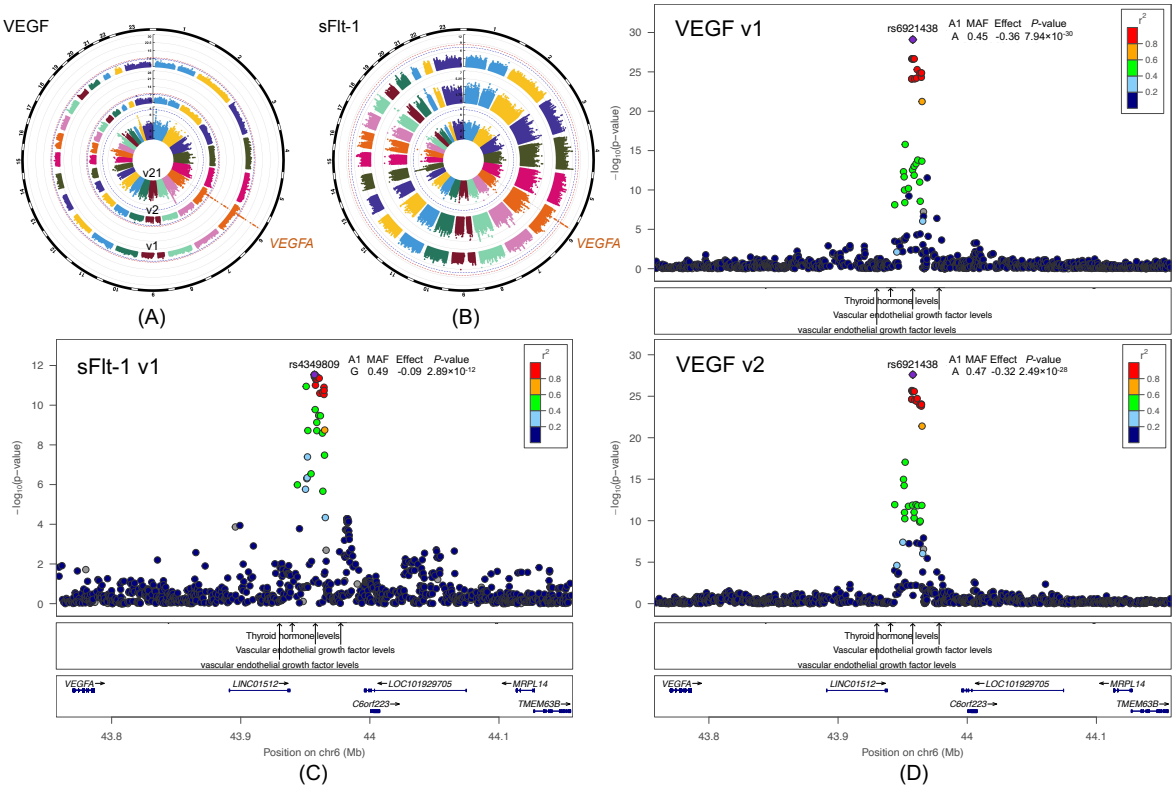

Figure 3. GWAS of sFlt-1 and VEGF. (A) and (B) Circular Manhattan plots for sFlt-1 and VEGF. The chromosomal position of each single SNP is displayed along the circle and the negative log10 of the association *P*-value is displayed on the radius. The red line represents the genome-wide significance level ($P < 5.6\times10^{-9}$) and blue line represents the suggestive significance level ($P < 5\times10^{-8}$). Results for visit 1 are displayed on the outer circle, visit 2 on the middle circle, and visit 2-1 on the inner circle. (C) Regional plot for SNP rs4349809 from the sFlt-1 visit 1 analysis. The lower portion of the figure displays the relative location of genes and the direction of transcription, while the middle portion shows known GWAS associations at the locus from the GWAS catalog.

The x-axis displays the chromosomal position and the y-axis shows the significance of the associations. The purple diamond shows the *P*-value for the reference SNP. The circles show the *P*-values for all other SNPs and are color coded according to the level of LD with the reference SNP using the nuMoM2b cohort. (D) Regional plots for SNP rs6921438 from the VEGF visit 1 and 2 analyses.

In the GWAS of sFlt-1, rs4349809 at the same *VEGFA* locus was found to be significantly associated with sFlt-1 levels at visit 1 (test allele: G, MAF = 0.49, effect size = -0.09 log(pg/mL), $P = 2.89\times10^{-12}$), but this association was not observed at visit 2 (Figure 3BC). Furthermore, rs4349809 was also significantly associated with VEGF levels in the same direction at visit 1 (effect size = -0.33 log(pg/mL), $P = 7.91\times10^{-25}$) and visit 2 (effect size = -0.3 log(pg/mL), $P = 2.34\times10^{-25}$) as with sFlt-1. VEGF promotes angiogenesis by stimulating the growth and branching of blood vessels, which is important for ensuring adequate blood flow to support the developing fetus. On the other hand, sFlt-1 acts as a natural inhibitor of VEGF by binding to it and preventing it from interacting with its receptors[30]. Despite their opposing roles in regulating angiogenesis during pregnancy, the genetic effects from the *VEGFA* locus exhibited the same association directions with their serum levels, and the locus was found to be more strongly associated with VEGF than sFlt-1. To compare our mixed-ancestry results with those from more genetically homogeneous populations with adequate sample sizes, we performed sensitivity analyses by repeating the GWAS analysis of VEGF and sFlt-1 using only self-reported Whites, who were treated as a proxy of European ancestry. The results were similar to those obtained using all individuals, but less statistically significant (Figures S5, S6 and Table S1).

The figures displaying the regional plots of other GWAS SNPs that achieved a suggestive significance level ($P < 5\times10^{-8}$) are presented in Figures S7-S10. Notably, rs981087 located near *CGA* on chromosome 6q14.3 was significantly associated with fβHCG levels at visit 1 (test allele: C, MAF = 0.48, effect size = -0.1 log(ng/mL), $P = 1.12\times10^{-8}$); rs10458657 in *VCL* on chromosome 10q22.2 was significantly associated with PAPP-A levels at visit 1 (test allele: C, MAF = 0.39, effect size = 0.19 log(mU/mL), $P = 2.89\times10^{-8}$); rs5951549 in *PTCHD1-AS* on chromosome Xp22.11 was significantly associated with VEGF levels at visit 2 (test allele: C, MAF = 0.07, effect size = 0.32 log(pg/mL), $P = 1.9\times10^{-8}$); and rs72886119 in *NCMAP* on chromosome 1p36.11 was significantly associated with the change of VEGF levels between visit 2 and visit 1 (test allele: C, MAF = 0.43,

effect size = 0.16 log(pg/mL), $P = 4.18\times10^{-8}$). A summary of all GWAS SNPs with $P < 5\times10^{-8}$ can be found in Table S3.

Observational associations between placental proteins and PE/gHTN

Our initial investigation focused on exploring the correlation among placental proteins measured during visits 1 and 2. We found that the same protein across both visits had a strong positive correlation, whereas different proteins within the same visit had a higher correlation than those across visits. Additionally, we observed that most proteins showed an inverse correlation with VEGF, while the majority of the remaining proteins exhibited a positive correlation (Figure 1B). Furthermore, we proceeded to examine the association between nine placental proteins and PE/gHTN using logistic regression, adjusting for the same factors as in the protein GWAS except for APO status, in a subset of the nuMoM2b cohort comprising the random sample of 11.8% of all women and women with PE or gHTN (Figure 1A). Additionally, we performed parallel analyses for visit 3 systolic blood pressure (SBP) (22-29 weeks) and visit 3 diastolic blood pressure (DBP), using linear regression in the 11.8% random sample. We found that VEGF measured at both visits exhibited a positive association with PE/gHTN, whereas PAPP-A and PlGF measured at both visits were inversely associated with PE/gHTN (Bonferroni-adjusted $P < 5.6\times10^{-3}$, as shown in Figure 1C and Table S4). ADAM-12 showed a significant negative association at visit 1, but not at visit 2. Furthermore, we observed a positive association between the change of ADAM-12 and ENG from visit 1 to visit 2 (visit 2-1) and PE/gHTN, as well as a negative association between PlGF (visit 2-1) and PE/gHTN. Hence, our findings provide observational evidence linking the levels of placental proteins with the risk of PE/gHTN in the nuMoM2b cohort.

Evidence for the causal impact of placental proteins on PE/gHTN

We followed the steps described in Methods to select independent IV SNPs with $P < 1\times10^{-5}$ from GWAS of placental proteins. We conducted a two-sample MR analysis to assess the causal effects of placental proteins measured in the first two visits on PE/gHTN, using the MR-robust adjusted profile scoring (MR-RAPS) approach[31]. Our analysis revealed a significant effect ($P < 5.6\times10^{-3}$) between ADAM-12 at visit 1 and gHTN (odds ratio (OR) = 0.78, 95% CI: 0.67-0.9, $P = 8.6\times10^{-4}$) (Figure 4). To verify the robustness of this potential causal relationship, we performed inverse variance weighting (IVW)[32] and MR pleiotropy residual sum and outlier (MR-PRESSO)[33] analyses

for replication studies. Both analyses confirmed the causal association between ADAM-12 and gHTN (Figures S11 and S12). No heterogeneity was observed in the IVW analyses as indicated by Cochran's Q, and the MR-PRESSO global pleiotropy test showed no sign of horizontal pleiotropy in IV SNPs for this replicated association. Furthermore, this replicated causal association was consistent with the observational association. In summary, our two-sample MR analyses provided evidence for a causal relationship between ADAM-12 at visit 1 and gHTN.

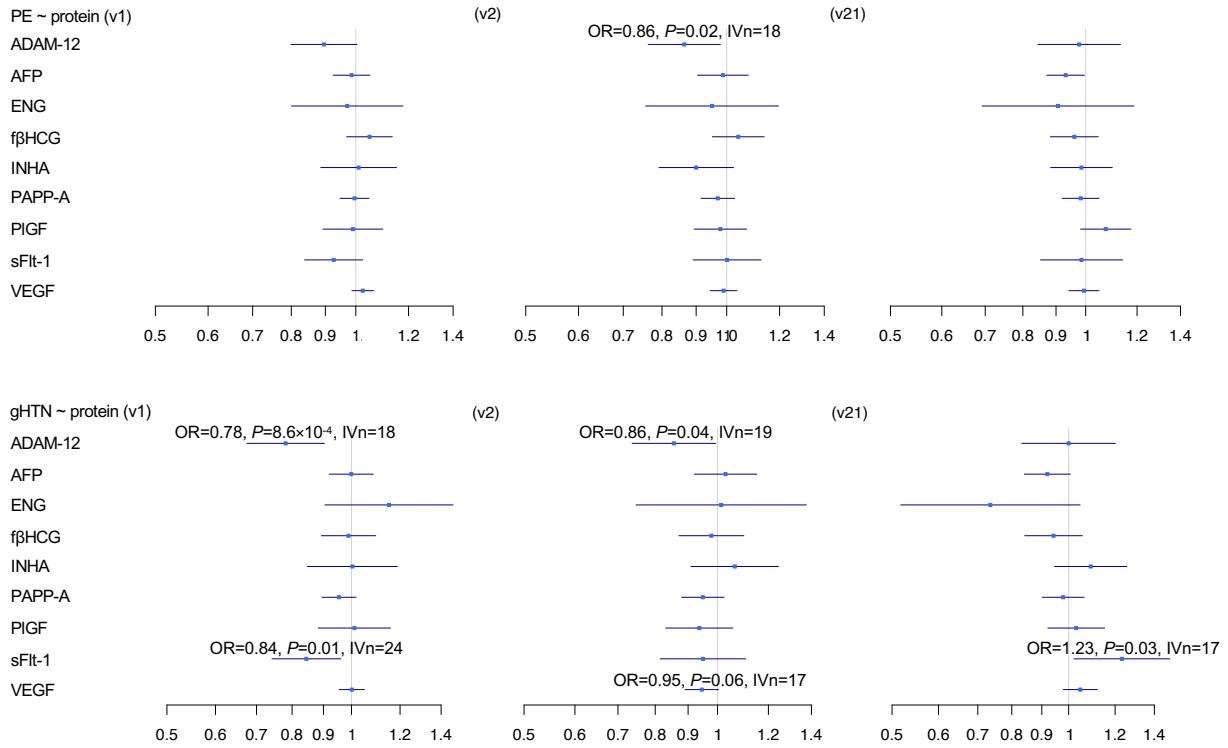

Figure 4. Causal estimates of nine placental proteins on PE and gHTN. *P*-values were determined by the two-sample MR-RAPS method. The squares represent the causal estimates on the odds ratio (OR) scale, and the whiskers show the corresponding 95% confidence intervals. The labels v1, v2 and v21 indicate visit 1, visit 2 and their difference (visit 2-1), respectively. IVn: number of IV SNPs used for the estimation of the causal effects.

The MR-RAPS analyses also suggested other associations at the nominal level ($P < 0.05$), although they did not reach significance after multiple-testing adjustment. For instance, ADAM-12 at visit 2 exhibited a potential inverse association with gHTN (OR = 0.86, 95% CI: 0.74-0.99, $P = 0.04$) and PE (OR = 0.86, 95% CI: 0.76-0.98, $P = 0.02$). Furthermore, we found that sFlt-1 measured at visit 1 was inversely associated with gHTN (OR = 0.84, 95% CI: 0.74-0.96, $P = 0.01$, Figure 4), and this inverse association was replicated by the other two MR methods (Figures S11 and S12). Moreover,

the observational association among only nuMoM2b individuals also exhibited the same effect direction. Although it has been repeatedly reported that higher levels of sFlt-1 were associated with PE when measured near the onset of PE, the association remains unclear when sFlt-1 is measured during early pregnancy[3,34,35]. Despite finding of an inverse association between visit 1 sFlt-1 and gHTN, the changes in sFlt-1 levels from visit 1 to visit 2 showed a possible positive association with gHTN (OR = 1.23, 95% CI: 1.02-1.49, $P = 0.03$). This suggests that the increase in sFlt-1 levels during pregnancy may be associated with gHTN, and not just the absolute sFlt-1 levels. The level of VEGF at visit 2 showed a marginal association with gHTN (OR = 0.95, 95% CI: 0.89-1, $P = 0.06$) and the other two MR methods yielded similar results. However, these results were in the opposite direction to the observational association (Figure 1C). It has been suggested that the lower level of VEGF is associated with PE/gHTN when measured close to the disease onset, possibly due to its inverse correlation with sFlt-1[30]. Thus, the two-sample MR analyses can potentially rectify any biased observational associations resulting from unobserved confounding effects. Additionally, none of the significant associations at $P < 0.05$ displayed any signs of heterogeneity, as evident from Cochran's Q statistics, and horizontal pleiotropy, as indicated by the MR-PRESSO global pleiotropy test.

As sensitivity analyses, we further conducted one-sample MR using the nuMoM2b individuals. Firstly, we obtained genetic effects of PE, HTN, SBP, and DBP by analyzing all women who consented to genetic data analyses in nuMoM2b to prepare for conducting MR analyses. We focused on ADAM-12 at visit 1 for the one-sample MR analysis because it displayed a suggestive causal relationship with gHTN from the two-sample MR analysis. The results of the one-sample IVW MR analysis revealed that ADAM-12 at visit 1 had an inverse association with SBP (effect size = -1.4, $P = 0.03$) and DBP (effect size = -1.5, $P = 2.2\times10^{-3}$) at visit 3 (Figure S13AB), although we could not replicate the association with gHTN. The observational results also displayed similar associations between ADAM-12 at visit 1 and SBP and DBP at visit 3 (Figure 1C). For the above two-sample MR analyses, we solely considered common variants as IVs since rare variants were typically excluded from GWAS. In cases where rare variants were included, they were generally grouped into sets (e.g., genes) before analysis to enhance the variability of data. However, it was challenging to ensure that the same SNPs were grouped into the same gene and the same test alleles were used between two cohorts, particularly when only summary statistics (e.g., burden statistics)

were accessible. As we had access to all the data in the nuMoM2b cohort, we could easily harmonize the genetic associations with the exposure and outcome, providing an opportunity to incorporate rare variants as additional IVs. Therefore, we took the genotyped rare variants into consideration for the analysis. We collapsed rare variants into annotated genes to conduct genome-wide rare variant association studies of placental proteins. Genes that had a burden $P < 1\times10\text{-}5$ were chosen as additional IVs. Next, we performed similar rare variant analyses on the four HTN related outcomes to prepare for conducting MR analyses. We conducted a one-sample IVW analysis using both common and rare variants as IVs and obtained similar results, with effect sizes of -1.5 ($P = 0.02$) for SBP and -1.4 ($P = 3.7\times10^{-3}$) for DBP at visit 3, as compared to common variants alone (Figure S13C, D). The reason for this could be due to the fact that our data were genotyped using microarray technology, which only covered a small portion of rare variants. All of the analyses displayed no indications of heterogeneity.

Evidence for causal relationships between PE/gHTN and long-term postpartum HTN

Additionally, we performed a two-sample MR analysis to investigate the causal relationship between PE/gHTN during the first pregnancy and long-term HTN occurring 2-7 years (mean 3.2 years) after the first pregnancy. We once again utilized the recent GWAS of PE/gHTN[24], and significant SNPs with $P < 5\times10^{-8}$ were used as IVs. We then combined the results with genetic associations of long-term HTN (2-7 years after the first pregnancy) using the nuMoM2b-HHS cohort, which comprised 972 cases and 3,409 controls. The MR-RAPS analyses revealed significant evidence for causal effects of PE (OR = 1.75, 95% CI: 1.16-2.65, $P = 8.3\times10^{-3}$) and gHTN (OR = 1.84, 95% CI: 1.21-2.81, $P = 4.7\times10^{-3}$) on the long-term postpartum HTN (Figure 5). The IVW and MR-PRESSO analyses exhibited consistent indication for causal outcomes (Figure 5). Furthermore, Cochran's Q showed no evidence of heterogeneity, and the MR-PRESSO global pleiotropy test revealed no horizontal pleiotropy in the selected IV SNPs. Moreover, the causal associations observed were in agreement with the observational outcomes obtained from the nuMoM2b-HHS cohort, in which both PE (OR = 3.65, $P = 4.6\times10^{-21}$) and gHTN (OR = 2.59, $P = 9.2\times10^{-28}$) were significantly associated with the risk of developing HTN 2-7 years after the first pregnancy. Thus, our two-sample MR analyses identified robust evidence for a causal relationship between PE/gHTN at the first pregnancy and HTN 2-7 years later.

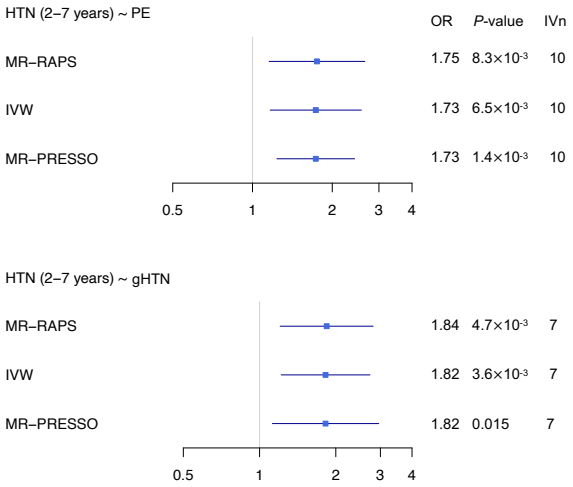

Figure 5. Causal estimates of PE and gHTN on long-term postpartum HTN. *P*-values were determined by the two-sample MR-RAPS, IVW, and MR-PRESSO methods. The squares represent the causal estimates on the OR scale, and the whiskers show the corresponding 95% confidence intervals. IVn: number of IV SNPs used for the estimation of the causal effects.

Mediation effect (ME) of ADAM-12 on long-term postpartum HTN via PE/gHTN

We conducted a mediation analysis to explore the potential indirect effect of placental proteins on long-term postpartum HTN via PE/gHTN, using the effect estimates from the two-step MR analyses. This analysis was focused on ADAM-12 at visit 1, which exhibited strong evidence of a causal association with gHTN. Our results showed a significant ME, whereby ADAM-12 at visit 1, inversely associated with gHTN, was further inversely associated with postpartum HTN 2-7 years later, as estimated by MR-RAPS (ME = -0.15, P = 0.03). The IVW analysis also replicated this ME (ME_{IVW} = -0.14, P_{IVW} = 0.04; ME_{PRESSO} = -0.14, P_{PRESSO} = 0.08).

DISCUSSION

Our observational findings provide insights into the correlation among placental proteins measured during the first two trimesters of pregnancy and their potential relationships with the risk of developing PE and gHTN. We found that each protein's serum level at visit 1 had a strong positive correlation with its level at visit 2, whereas different proteins within the same visit had a higher correlation with each other than those across visits. This suggests that longitudinal changes in protein level may be more informative than cross-sectional comparisons for identifying potential biomarkers of disease risk. Our results also revealed that most placental proteins showed an inverse correlation with VEGF, while the remaining proteins exhibited a positive correlation. These

findings are consistent with previous reports that have demonstrated a complex interplay between VEGF and other placental proteins[4].

In our GWAS, we identified significant associations between ADAM-12 levels and the *PZP* locus on chromosome 6q21.1, as well as between VEGF/sFlt-1 and the *VEGFA* locus on chromosome 6q21.1. Our findings not only contribute to the growing body of literature on the genetic determinants of placental proteins but also provide insights into the genetic regulation of these proteins during pregnancy.

Our GWAS results found a significant association between ADAM-12 levels at visit 2 and two SNPs: rs6487735, located near the *PZP* gene, and rs2277413, a missense SNP within *PZP*. *PZP* is responsible for encoding a pregnancy zone glycoprotein (PZP) that is primarily synthesized by the liver and secreted into the bloodstream[36]. The PZP protein is thought to regulate immunity during pregnancy and may also have inhibitory functions for proteases[37,38]. However, to date, no studies have reported a relationship between ADAM-12 and the *PZP* gene or its encoded protein. It is possible that PZP could potentially inhibit the protease activity of ADAM-12 or other proteases within the same family. Further research is necessary to confirm this and explore the potential implications of their interplay in various biological processes and pathological conditions, including hypertensive disorders of pregnancy. According to the GWAS catalog, this genetic locus is associated with a variety of traits, including non-alcoholic fatty liver disease histology[39], breast size[40], obesity-related traits[41], and IgG glycosylation (Figure 2)[42]. Our TWAS results showed that the expression of *IGLV* genes was significantly associated with ADAM-12 levels at visit 2. This finding suggests that immunoglobulin light chains may regulate ADAM-12 activity and have a potential role in modulating immune-related processes in which ADAM-12 is involved[43].

Regarding VEGF, our study identified rs6921438 near *VEGFA* as the most strongly associated variant with VEGF levels at both visits. This finding is in line with previous reports that have consistently replicated the association between the *VEGFA* locus and circulating VEGF levels in non-pregnant populations[18-21]. Importantly, our study extends these findings to nulliparous pregnant women, confirming the presence of *cis*-acting genetic associations with VEGF levels during pregnancy. The same locus was also found to be associated with thyroid hormone levels[44,45].

In our analysis of sFlt-1, we found that rs4349809 at the *VEGFA* locus was significantly associated with sFlt-1 levels at visit 1, but not at visit 2. This SNP was also significantly associated with VEGF levels in the same direction at both visits. Despite their antagonistic functions, the genetic effects from the *VEGFA* locus exhibited the same association directions with their serum levels. These findings are noteworthy, as VEGF and sFlt-1 play opposing roles in regulating angiogenesis during pregnancy[30]. This observation highlights the complex interplay between pro-angiogenic and anti-angiogenic factors during pregnancy and suggests that the *VEGFA* locus may have a more pronounced influence on VEGF than sFlt-1.

Our MR analysis using MR-RAPS, IVW, and MR-PRESSO suggested a causal association between ADAM-12 levels at visit 1 and gHTN. The direction of the significant effect was consistent when using IV SNPs with $P < 5\times10^{-8}$ (Figure S14), although only one SNP met the cutoff. This finding is also consistent with our observational analysis. While ADAM-12 levels did not meet the significance threshold for evidence of a causal relationship with PE, the consistent direction and similar ORs for ADAM-12 levels at both visits for PE and gHTN (Figure 4) suggests that a similar relationship to PE remains possible. This is consistent with clinical evidence that gHTN has similar clinical features and frequently progresses to PE when pregnancy is expectantly managed[46,47]. The fact that ADAM-12 levels were not causally related to PE in our study may be due to limited instrument strength. Alternatively, it is possible that progression from gHTN to PE requires a "second hit" or occurs through activation of an alternative pathway that remains unknown. Importantly, the robust causal evidence provided by the MR analysis supports the potential promise of ADAM-12 as both a predictive marker for gHTN, as well as a potential target for therapeutic intervention.

Our MR analyses also identified several potential associations between placental proteins and PE/gHTN. Although the associations did not reach significance after multiple-testing adjustment, they provide hypothesis-generating insights for further investigation. The first of these includes the aforementioned potential inverse causal association of ADAM-12 levels at visit 2 with gHTN and PE. Additionally, we observed a potential inverse causal relationship between sFlt-1 measured at visit 1 and gHTN, which was consistent with our observational association (Figure 4). This is

in contrast to previous reports linking higher sFlt-1 levels with PE when measured near the onset of the disease[3,34,35]. However, we also found that changes in sFlt-1 levels from visit 1 to visit 2 (visit 2-1) showed a possible positive causal relationship with gHTN, suggesting that the increase in sFlt-1 levels during pregnancy may be associated with gHTN, and not just the absolute sFlt-1 levels. Taken together, these findings highlight the complex relationship between sFlt-1 levels during early pregnancy and gHTN. The VEGF levels at visit 2 showed a marginal inverse causal relationship with gHTN, which was in the opposite direction to the observational association. This finding is consistent with previous literature suggesting that lower levels of VEGF are associated with PE/gHTN when measured close to the disease onset[30]. The inconsistency between the MR analysis and the observational association highlights the potential for unobserved confounding effects in observational studies. In this regard, MR analyses have the potential to mitigate these biases and provide more reliable estimates of causal effects.

Moreover, our MR analysis provided evidence for a causal effect of PE and gHTN during the first pregnancy on the development of HTN 2-7 years later. This finding is consistent with the observational outcomes obtained from the nuMoM2b-HHS cohort and other studies that have also identified a link between these pregnancy complications and long-term maternal cardiovascular risk[48]. Our findings improve this body of work by providing new clarity on the causal role of hypertensive disorders of pregnancy in long-term maternal cardiovascular health. Prior observational studies, which established the importance of pregnancy as a window to future health, were unable to determine whether pregnancy complications such as gHTN and PE contribute to or are merely indicators of underlying predisposition to cardiovascular disease. This knowledge gap complicated preconception counseling and reproductive planning for those considering whether to pursue additional pregnancies based on the potential impact of those pregnancies on their own health. Therefore, identifying the causal associations of PE and gHTN on long-term postpartum HTN serve to highlight the critical importance of investigations into prevention and treatment of hypertensive disorders of pregnancy, which may have the potential to improve long-term maternal health. The additional mediation analysis offered further insights, indicating that the impact of ADAM-12 on postpartum HTN could be explained in part by its indirect effect through gHTN. This suggests that the relationship of placental proteins on long-term postpartum HTN might be mediated to some extent by their impact on the development of PE/gHTN during pregnancy.

We recognize several study limitations. First, the sample size of the nuMoM2b cohort was limited, which may have reduced the statistical power to detect small genetic effects on placental proteins. However, this study is the first to systematically examine multiple protein levels during pregnancy using a GWAS approach. Second, only nine proteins related to placental health were examined in this study, and untargeted proteomics data using techniques such as Olink or SomaScan[49] may identify additional genetically regulated proteins during pregnancy, although these techniques have their limitations and can only capture a small proportion of the proteome[50]. Third, the lack of available data concerning the same questions prevented the replication of the GWAS of placental proteins in other cohorts. Fourth, as with any MR analysis, some assumptions are untestable, such as horizontal pleiotropy. To address the assumptions of MR and control for different types of biases, several sensitivity analyses were conducted following the latest guidelines[23]. These included the use of multiple MR approaches with different assumptions, comparison of MR analysis results with observational associations, conducting additional one-sample MR analyses, and selecting different IV SNPs by setting different *P*-value cutoffs.

In conclusion, our study identified significant genetic associations with placental proteins ADAM-12, VEGF, and sFlt-1, providing insights into their regulation during pregnancy. MR analyses revealed evidence for causal relationships between placental proteins, particularly ADAM-12, and PE/gHTN, which could lead to potential prevention and treatment strategies for these conditions. Our findings also suggest that placental proteins, such as ADAM-12, have the potential to serve as biomarkers for postpartum HTN risk. Further research is needed to explore the biological mechanisms underlying these associations and investigate potential therapeutic targets for managing postpartum HTN in women with a history of PE/gHTN.

METHODS

Study Cohort

The nuMoM2b study

The nuMoM2b enrolled 10,038 women between 2010 and 2013 from eight geographically disparate clinical sites in the United States. Prior pregnancy history is often the best predictor of future pregnancy outcomes, but it is challenging to assess risks during a woman's first pregnancy.

Therefore, nuMoM2b was designed to recruit a large, racially and geographically diverse cohort of nulliparous pregnant women (Table S5). The primary aim of the study was to determine maternal characteristics, including genetic factors, physiological responses to pregnancy, and environmental factors that can predict adverse pregnancy outcomes[51]. The study participants were longitudinally followed and underwent four study visits, from the first trimester to after birth. Throughout pregnancy, various data were obtained, including detailed interviews, questionnaires, research ultrasounds, maternal biometric measurements, and biospecimens (Table S6). Peripheral maternal blood samples were collected at three study visits during pregnancy: visit 1 (6-13 weeks), visit 2 (16-21 weeks), and visit 3 (22-29 weeks). The methods of the nuMoM2b study have been described in detail elsewhere[51], and the study was approved by the Institutional Review Boards at all participating centers[29,51,52]. Genome-wide genotyping of 9,757 women with sufficient material was performed using whole blood samples collected at the first visit, using the Infinium Multi-Ethnic Global D2 BeadChip (Illumina, Miami, USA).

Maternal serum samples were collected at two visits, 6-13 weeks and 16-21 week, to measure the levels of nine placental proteins, including VEGF (pg/mL), PlGF (pg/mL), ENG (ng/mL), sFlt-1 (pg/mL), ADAM-12 (ng/mL), PAPP-A (mU/mL), fβHCG (ng/mL), INHA (pg/mL), and AFP (IU/mL). The analysis was conducted in a subsample of the nuMoM2b cohort, which included a random sample of 11.8% of all women (n = 1,088) and women who experienced any of the following APOs (n = 1,502): PTB, PE, SGA, or SB. Notably, the 11.8% random sample included 177 women who had experienced APOs. Of the 2,413 (n = 1,088+1,502-177) women, genotypes were available for 2,352. The protein levels were log-transformed for subsequent analyses.

The nuMoM2b-HHS study

The nuMoM2b-HHS was carried out as a sequel study to gain a better understanding of the influence of pregnancy outcomes on subsequent health[29]. The study comprised a subset of 7,003 women from the nuMoM2b cohort who underwent screening at 6-month intervals, and 4,484 of them completed the laboratory assessments at the 2-7 year postpartum in-person visit. During the visit, participants provided information on their medical history, mental health, nutrition, physical activity, and sleep by completing questionnaires, and also underwent clinical measurements and biospecimen

collection (Table S6). Long-term HTN 2-7 years post the first pregnancy was defined by a SBP and DBP reading of ≥ 130/80 mmHg or the use of antihypertensive medication.

Analytical Workflow for Quality Control (QC) and GWAS with Multi-Ethnic Data

Given the ethnic diversity of the nuMoM2b cohort, careful adjustment for potential bias due to population stratification is crucial. In multi-ethnic GWAS, it is common to split the total sample into separate populations, perform genotype data QC, imputation, and GWAS in each population separately, followed by a meta-analysis. However, in our study, the sample size for GWAS of proteins is relatively small, at around two-thousand individuals. Further splitting the data into separate populations would make GWAS unreasonable for certain populations with limited sample sizes, particularly those that cannot be classified into any group. The PCA revealed a continuous population structure in the nuMoM2b cohort (Figure S4). Consequently, we have developed an analytical workflow for QC and GWAS with multi-ethnic data that maximizes the sample size while minimizing bias due to population stratification (Figure S15).

Genotype data QC and imputation

To conduct pre-imputation QC, we initially excluded SNPs and women with a missing rate greater than 10% in the nuMoM2b cohort. Subsequently, we identified SNPs with Hardy-Weinberg equilibrium (HWE) $P < 1\times10^{-6}$ or MAF < 0.01 in self-reported White, Black, and Hispanic populations, which constitute the three major ethnic groups. These three major ethnic groups were used as proxies for genetic ancestries in our study. These SNPs were then removed from the full dataset. We subsequently performed PCA using PC-AiR[53] to account for related individuals, with kinship coefficients estimated by KING[54]. Next, we assessed sex concordance and autosomal heterozygosity, accounting for population structure with the first six PCs, following the approach described in Bycroft et al. S3.5.2[55]. In brief, we fitted the following linear regression model for the raw sex concordance or autosomal heterozygosity value, F,

$$F = F_0 + \sum_{i=1}^{6} \beta_i PC_i + \sum_{i=1}^{6} \sum_{j=i}^{6} \beta_{ij} PC_i PC_j + \varepsilon$$

where the fitted F_0 is the PC-adjusted (i.e., ancestry-corrected) sex concordance or autosomal heterozygosity F-value. We then excluded women from the data who violated sex concordance (adjusted F-value > 0.2) or autosomal heterozygosity (adjusted |F-value| > 0.15) (Figure S16). As

a result, a few individuals were removed from the data. We then repeated the previously mentioned steps for HWE and MAF filters and recalculated the PCA (Figure S4).

Next, we used the TOPMed Imputation Server[56] to perform genotype imputation, using the TOPMed (Version R2 on GRC38) as the reference panel, which includes all populations in our cohort. Joint imputation was applied to all individuals, as it is expected to perform as well as splitting the reference panel to match the target population[57]. We retained genotyped and imputed SNPs with imputation quality $r^2 > 0.3$ and repeated HWE and MAF filters. Following QC, our nuMoM2b cohort comprised 9,742 women. For protein GWAS, we analyzed 2,263, 2,134, and 2,045 women for visit 1, visit 2, and visit 2-1 analyses, respectively. Due to our moderate sample size, we focused on SNPs with MAF > 0.05.

GWAS adjusting for population structure

To account for population stratification and genetic relatedness among individuals, we employed a mixed-model approach to conduct GWAS of proteins. This approach is advantageous because it includes all individuals, regardless of familial and ancestral relatedness, and can provide greater statistical power by controlling for the variance attributed to genetic relatedness[58]. We used the GENESIS R/Bioconductor package[59,60] to fit linear mixed models that integrated a random effect to control for genetic relatedness, with the genetic relationship matrix (GRM) computed by PC-Relate[61]. The covariates considered in the analysis included age and age-squared at visit 1, first 10 PCs calculated by genotypic data, self-reported race, clinical sites, and status for any APOs. To model the continuous log-transformed levels of each protein, y, we used a linear mixed model:

$$y = \beta_0 + \beta_1 \mathrm{SNP} + \boldsymbol{\beta}_{\mathrm{cov}} \mathbf{covariates} + u + \varepsilon$$

where u is the random variable accounting for genetic relatedness, and ε is the random error. Although mixed models are advantageous for controlling population stratification, their effectiveness may be limited in a diverse cohort. This is because the GRM, which is estimated based on genome-wide data, assumes that all SNPs have similar population-level deviations. However, some SNPs may have larger or smaller deviations between populations than the genome-wide average, leading to inadequate control of false positive rates[62]. As a result, further adjustments are necessary to address this issue.

Hence, we conducted supplementary genome-wide interaction studies (GWIS) to explore the association between SNP-by-PCs and protein levels,

$$y = \beta_0 + \beta_1 \mathrm{SNP} + \boldsymbol{\beta}_{\mathrm{int}} \mathrm{SNP} \times \mathbf{PCs} + \boldsymbol{\beta}_{\mathrm{cov}} \mathbf{covariates} + u + \varepsilon$$

where **PCs** are the first 10 PCs that capture the ancestry information and thus $\boldsymbol{\beta}_{\mathrm{int}}$ includes 10 parameters. We then used a 10 degree of freedom test to evaluate the null hypothesis H_0: $\boldsymbol{\beta}_{\mathrm{int}} = 0$. A significant $\boldsymbol{\beta}_{\mathrm{int}}$ indicates that the effects of a particular SNP differ across ancestries. The GWIS were also performed using the GENESIS package[59,60]. To ensure the retention of only SNPs with consistent effects, we removed SNPs with a *P*-value of $\boldsymbol{\beta}_{\mathrm{int}} < 0.01$ from the GWAS results. It is worth noting that the threshold of 0.01 is conservative and excludes any potentially ancestry-specific SNPs.

Collider bias consideration

The GWAS of proteins were based on a combination of all women with APOs and a random subset. As a result, the protein levels might not accurately reflect the distributions in the general population. Although women in the nuMoM2b cohort had not progressed to APOs during visit 1 and visit 2, we adjusted for the APO status in our GWAS of proteins. However, if both the protein and SNP are independent causes of APO status, there is a possibility of collider bias (as illustrated in Figure S17). To address this issue, we performed a GWAS of APO status and examined the association between proteins and APO status, while adjusting for the same covariates as before except for the APO status, which was the binary outcome in this analysis. We identified the protein-SNP pair as a potential collider bias and excluded it when the *P*-value was less than 0.01 in the GWAS of APO status and when the *P*-value of the association between the protein and APO status was less than 0.05.

Using the proposed GWAS pipeline, we examined genetic associations with nine placental proteins, which were measured during visit 1 and visit 2, as well as the differences between these two time points. We set the genome-wide significance level at $P < 5.6\times10^{-9}$ (Bonferroni-adjusted for nine proteins: $5\times10^{-8}/9 = 5.6\times10^{-9}$) to identify significant SNPs, which were annotated using ANNOVAR[63]. Additionally, we performed sensitivity analyses by restricting the GWAS to self-reported Whites only. We conducted the analyses using PLINK2[64] without considering a random effect that controls for genetic relatedness. We further conducted a TWAS using the Fusion

software[65] as a post-GWAS analysis to search for potential regulatory effects. The TWAS analysis utilized expression reference weights for liver, pancreas, vagina, uterus, and whole blood from the GTEx v8 multi-tissue expression data, with LD calculated based on nuMoM2b individuals. Accounting for multiple testing, we set a significance level of *FDR P* < 0.05.

MR Causal Inference

Figure S18 illustrates the design of the MR study for causal inference. The MR approach was based on three assumptions: (1) the IV SNPs are linked to the exposure, i.e., protein levels; (2) there are no unobserved confounders influencing the associations between IV SNPs and outcome; and (3) the IV SNPs affect the outcome only through changes in the exposure, i.e., no horizontal pleiotropy. To investigate the causal relationships between proteins and PE/gHTN, as well as PE/gHTN and long-term postpartum HTN, we used a two-sample MR framework that incorporates sensitivity analyses. We also conducted a causal mediation analysis for proteins → PE/gHTN → long-term postpartum HTN. Two-sample MR has a major advantage over one-sample MR as it only requires GWAS summary statistics, rather than individual-level data. Additionally, two-sample MR is typically considered more conservative and unbiased than one-sample MR because it allows for separate cohorts for exposure and outcome data, while in one-sample MR, both exposure and outcome are from the same cohort[23].

Proteins → PE/gHTN

We used independent SNPs with a $P < 1\times10^{-5}$ from the GWAS of proteins as IVs. Our rationale for using this threshold was to identify more independent IV SNPs, as this could promote balanced pleiotropy, which helps mitigate bias due to horizontal pleiotropy. Previous studies have shown that this liberal threshold can result in better performance than a conservative threshold of 5×10^{-8} [66]. To select independent IV SNPs, we used a stepwise selection strategy for each chromosome. This approach involved: (1) selecting SNPs with a GWAS $P < 1\times10^{-5}$, ordering them by *P*-values, and selecting the top SNP with the lowest *P*-value; (2) running the genetic regression models again on the remaining SNPs, with additional adjustment on the saved top SNP; (3) selecting SNPs with a conditional $P < 1\times10^{-5}$, ordering them by *P*-values, and saving the top conditional SNP; (4) rerunning the genetic regression models again on the remaining SNPs, with additional adjustment

on the last saved top conditional SNP; and (5) repeating steps (3) and (4) until no SNPs remained in the remaining set.

We obtained genetic association estimates for PE and gHTN from a recent multi-ancestry meta-analysis of GWAS, which included 17,150 PE cases and 451,241 controls, and 8,961 gHTN cases and 184,925 controls in the discovery analysis. It's important to note that the nuMoM2b cohort was not included in the discovery analysis, but rather treated as a follow-up cohort. A detailed description of the study design and participant characteristics can be found in the original publication[24]. After extracting the genetic effect estimates from the GWAS of PE/gHTN, we switched the effect directions and test alleles to ensure consistency with the results from the GWAS of proteins. Ideally, both studies in a two-sample MR should include the same populations, as some SNPs are expected to have ancestry-specific effects. However, since we excluded SNPs from the GWAS of proteins that had ancestry-specific effects on placental proteins, the genetic results were considered to be generic, and therefore, there were fewer concerns about population compatibility with a second study when conducting a two-sample MR in this current study.

We used MR-RAPS[31] as the primary method, with squared error loss, as it can account for weak instrument bias which is particularly relevant in our study with a limited number of strong IV SNPs. We also used the commonly used MR methods, including IVW random-effect (RE)[32], and MR-PRESSO[33], which corrects pleiotropy via outlier IV removal. However, we did not use MR-Egger in this study, as it is a conservative method[67], especially when only a few genetic loci are associated with placental proteins in our GWAS. The reason for using these different methods is that they make different assumptions, and if they produce similar effect estimates, this provides greater confidence in any causal claims. We also assessed the robustness of our results by conducting analyses to identify potential violations of MR assumptions, including heterogeneity measured by Cochran's Q statistic for IVW analyses[32] and horizontal pleiotropy measured by the MR-PRESSO global pleiotropy test[33]. To ensure statistical significance, we set the threshold at $P < 5.6\times10^{-3}$ for the primary analysis, Bonferroni-adjusted for nine proteins ($0.05/9 = 5.6\times10^{-3}$).

Sensitivity analyses

As sensitivity analyses, we conducted one-sample MR to assess the robustness of our findings, focusing on ADAM-12 due to its strongest causal indication. We conducted GWAS of PE (528 cases and 6,690 controls), gHTN (2,020 cases and 6,690 controls), third trimester (v3: 22-29 weeks) SBP (n = 8,861), and third trimester DBP (n = 8,861) using the nuMoM2b cohort, following the proposed GWAS pipeline. In the one-sample setting, the same covariate adjustments could be ensured. The same IV SNPs and their effect estimates extracted from the above ADAM-12 GWAS were used for the one-sample MR analysis, employing the IVW (RE) method, which is asymptotically equivalent to the two-stage least square regression (2SLS) method in a one-sample MR analysis[68].

Up to this point, we have exclusively used common variants (MAF > 0.05) as IVs. However, a one-sample setting makes it easier to harmonize genetic associations with the exposure and outcome, providing an opportunity to include rare variants (MAF < 0.05) as IVs. Therefore, we additionally examined directly genotyped rare variants in the analysis, excluding imputed rare variants due to potential suboptimal imputation quality. It should be noted that as our data was genotyped using microarray, only a small portion of rare variants were covered. For the analysis of ADAM-12, we used genome-wide significant rare variants collapsed into gene-based burden statistics as additional IVs (burden $P < 1\times10^{-5}$), which was conducted using the GENESIS package[59,60]. Similarly, we conducted burden analyses on PE, gHTN, SBP (v3), and DBP (v3) using the nuMoM2b cohort. We considered genes with at least 5 rare variants. Using both common and rare variants as IVs can enhance the credibility of the MR analysis and provide a convincing result.

PE/gHTN → long-term postpartum HTN

We used SNPs with a $P < 5\times10^{-8}$ from a recent multi-ancestry meta-analysis of GWAS of PE/gHTN as IVs. To investigate genetic associations with long-term postpartum HTN, we conducted a GWAS of HTN occurring 2-7 years after the first pregnancy (972 cases and 3,409 controls) using the nuMoM2b-HHS cohort according to the proposed GWAS pipeline. Long-term HTN was defined as SBP/DBP ≥ 130/80 mmHg or use of antihypertensive medication 2-7 years after the first pregnancy. Subsequently, we conducted a two-sample MR analysis using MR-RAPS,

IVW (RE), and MR-PRESSO to assess the causal relationships between PE/gHTN during the first pregnancy and long-term postpartum HTN.

Proteins → PE/gHTN → long-term HTN

The two MR analyses we performed, one for proteins → PE/gHTN and another for PE/gHTN → long-term postpartum HTN, can be used for mediation analysis[69]. We used the product of coefficients method to estimate the mediated effect ($\alpha\beta$), where α is the causal effect of proteins → PE/gHTN and β is the causal effect of PE/gHTN → long-term postpartum HTN. To test the composite null hypothesis $H_0: \alpha\beta = 0$, we estimated the standard error of $\alpha\beta$ using the multivariate delta method based on a first-order Taylor series approximation, $\sigma_{\alpha\beta} = \sqrt{\alpha^2\sigma_\beta^2 + \beta^2\sigma_\alpha^2}$ [70]. This analysis was conducted solely on ADAM-12 at visit 1 due to the strong causal association it showed with PE/gHTN.

Observational associations

We conducted logistic regression analysis to investigate the association between nine placental proteins (at visit 1, visit 2, and visit 2-1) and PE or gHTN, respectively, using the nuMoM2b cohort. The same covariate adjustments were applied as in the GWAS of proteins, except for APO status. Additionally, we used logistic regression analysis to examine the associations of PE and gHTN with long-term postpartum HTN, using the nuMoM2b-HHS cohort, adjusting for the same covariates.

DATA AVAILABILITY

The individual-level phenotype and genotype data have been reported earlier[51,52] and are available upon request.

FUNDING

Support for performing DNA extraction and GWAS from the Indiana University Grand Challenges Precision Diabetes project funding. D.M.H. was partially funded by R01HD101246 from the U.S. National Institutes of Health (NIH). M.C.H. was supported by K08HL166687 from the U.S. NIH

and American Heart Association (940166, 979465). Original funding for nuMoM2b sample and data collection are noted in the study methods papers[29,51].

CONFLICT OF INTEREST

M.C.H. reports consulting fees from CRISPR Therapeutics; advisory board service for Miga Health; and grant support from Genentech (all unrelated to the present work).

REFERENCE

1. Huppertz, B. (2008). Placental origins of preeclampsia: challenging the current hypothesis. Hypertension *51*, 970-975. 10.1161/HYPERTENSIONAHA.107.107607.
2. De Falco, S. (2012). The discovery of placenta growth factor and its biological activity. Exp Mol Med *44*, 1-9. 10.3858/emm.2012.44.1.025.
3. Levine, R.J., Maynard, S.E., Qian, C., Lim, K.H., England, L.J., Yu, K.F., Schisterman, E.F., Thadhani, R., Sachs, B.P., Epstein, F.H., et al. (2004). Circulating angiogenic factors and the risk of preeclampsia. N Engl J Med *350*, 672-683. 10.1056/NEJMoa031884.
4. Maynard, S.E., Min, J.Y., Merchan, J., Lim, K.H., Li, J., Mondal, S., Libermann, T.A., Morgan, J.P., Sellke, F.W., Stillman, I.E., et al. (2003). Excess placental soluble fms-like tyrosine kinase 1 (sFlt1) may contribute to endothelial dysfunction, hypertension, and proteinuria in preeclampsia. J Clin Invest *111*, 649-658. 10.1172/JCI17189.
5. Myatt, L. (2002). Role of placenta in preeclampsia. Endocrine *19*, 103-111. 10.1385/ENDO:19:1:103.
6. Wang, J., Dong, X., Wu, H.Y., Wu, N., Zhang, X.J., Wang, X., and Shang, L.X. (2016). Relationship of Liver X Receptors alpha and Endoglin Levels in Serum and Placenta with Preeclampsia. PLoS One *11*, e0163742. 10.1371/journal.pone.0163742.
7. Ciobanu, A., Rouvali, A., Syngelaki, A., Akolekar, R., and Nicolaides, K.H. (2019). Prediction of small for gestational age neonates: screening by maternal factors, fetal biometry, and biomarkers at 35-37 weeks' gestation. Am J Obstet Gynecol *220*, 486 e481-486 e411. 10.1016/j.ajog.2019.01.227.
8. Korzeniewski, S.J., Romero, R., Chaiworapongsa, T., Chaemsaithong, P., Kim, C.J., Kim, Y.M., Kim, J.S., Yoon, B.H., Hassan, S.S., and Yeo, L. (2016). Maternal plasma angiogenic index-1 (placental growth factor/soluble vascular endothelial growth factor receptor-1) is a biomarker for the burden of placental lesions consistent with uteroplacental underperfusion: a longitudinal case-cohort study. Am J Obstet Gynecol *214*, 629 e621-629 e617. 10.1016/j.ajog.2015.11.015.
9. Kim, M., Park, H.J., Seol, J.W., Jang, J.Y., Cho, Y.S., Kim, K.R., Choi, Y., Lydon, J.P., Demayo, F.J., Shibuya, M., et al. (2013). VEGF-A regulated by progesterone governs uterine angiogenesis and vascular remodelling during pregnancy. EMBO Mol Med *5*, 1415-1430. 10.1002/emmm.201302618.
10. Christians, J.K., and Beristain, A.G. (2016). ADAM12 and PAPP-A: Candidate regulators of trophoblast invasion and first trimester markers of healthy trophoblasts. Cell Adh Migr *10*, 147-153. 10.1080/19336918.2015.1083668.
11. Serra, B., Mendoza, M., Scazzocchio, E., Meler, E., Nolla, M., Sabria, E., Rodriguez, I., and Carreras, E. (2020). A new model for screening for early-onset preeclampsia. Am J Obstet Gynecol *222*, 608 e601-608 e618. 10.1016/j.ajog.2020.01.020.
12. Wald, N.J., Watt, H.C., and Hackshaw, A.K. (1999). Integrated screening for Down's syndrome based on tests performed during the first and second trimesters. N Engl J Med *341*, 461-467. 10.1056/NEJM199908123410701.
13. Jelliffe-Pawlowski, L.L., Baer, R.J., Blumenfeld, Y.J., Ryckman, K.K., O'Brodovich, H.M., Gould, J.B., Druzin, M.L., El-Sayed, Y.Y., Lyell, D.J., Stevenson, D.K., et al. (2015). Maternal characteristics and mid-pregnancy serum biomarkers as risk factors for subtypes of preterm birth. BJOG *122*, 1484-1493. 10.1111/1471-0528.13495.
14. Parry, S., Carper, B.A., Grobman, W.A., Wapner, R.J., Chung, J.H., Haas, D.M., Mercer, B., Silver, R.M., Simhan, H.N., Saade, G.R., et al. (2022). Placental protein levels in

maternal serum are associated with adverse pregnancy outcomes in nulliparous patients. Am J Obstet Gynecol *227*, 497 e491-497 e413. 10.1016/j.ajog.2022.03.064.

15. Botkjaer, J.A., Borgbo, T., Kloverpris, S., Noer, P.R., Oxvig, C., and Andersen, C.Y. (2016). Effect of pregnancy-associated plasma protein-A (PAPP-A) single-nucleotide polymorphisms on the level and activity of PAPP-A and the hormone profile in fluid from normal human small antral follicles. Fertil Steril *106*, 1778-1786 e1778. 10.1016/j.fertnstert.2016.09.008.
16. Botkjaer, J.A., Noer, P.R., Oxvig, C., and Andersen, C.Y. (2019). Author Correction: A common variant of the pregnancy-associated plasma protein-A (PAPPA) gene encodes a protein with reduced proteolytic activity towards IGF-binding proteins. Sci Rep *9*, 17523. 10.1038/s41598-019-53957-x.
17. Ruggiero, D., Nutile, T., Nappo, S., Tirozzi, A., Bellenguez, C., Leutenegger, A.L., and Ciullo, M. (2021). Genetics of PlGF plasma levels highlights a role of its receptors and supports the link between angiogenesis and immunity. Sci Rep *11*, 16821. 10.1038/s41598-021-96256-0.
18. Ahola-Olli, A.V., Wurtz, P., Havulinna, A.S., Aalto, K., Pitkanen, N., Lehtimaki, T., Kahonen, M., Lyytikainen, L.P., Raitoharju, E., Seppala, I., et al. (2017). Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors. Am J Hum Genet *100*, 40-50. 10.1016/j.ajhg.2016.11.007.
19. Choi, S.H., Ruggiero, D., Sorice, R., Song, C., Nutile, T., Vernon Smith, A., Concas, M.P., Traglia, M., Barbieri, C., Ndiaye, N.C., et al. (2016). Six Novel Loci Associated with Circulating VEGF Levels Identified by a Meta-analysis of Genome-Wide Association Studies. PLoS Genet *12*, e1005874. 10.1371/journal.pgen.1005874.
20. Debette, S., Visvikis-Siest, S., Chen, M.H., Ndiaye, N.C., Song, C., Destefano, A., Safa, R., Azimi Nezhad, M., Sawyer, D., Marteau, J.B., et al. (2011). Identification of cis- and trans-acting genetic variants explaining up to half the variation in circulating vascular endothelial growth factor levels. Circ Res *109*, 554-563. 10.1161/CIRCRESAHA.111.243790.
21. Sliz, E., Kalaoja, M., Ahola-Olli, A., Raitakari, O., Perola, M., Salomaa, V., Lehtimaki, T., Karhu, T., Viinamaki, H., Salmi, M., et al. (2019). Genome-wide association study identifies seven novel loci associating with circulating cytokines and cell adhesion molecules in Finns. J Med Genet *56*, 607-616. 10.1136/jmedgenet-2018-105965.
22. Burgess, S., Timpson, N.J., Ebrahim, S., and Davey Smith, G. (2015). Mendelian randomization: where are we now and where are we going? Int J Epidemiol *44*, 379-388. 10.1093/ije/dyv108.
23. Burgess, S., Davey Smith, G., Davies, N.M., Dudbridge, F., Gill, D., Glymour, M.M., Hartwig, F.P., Holmes, M.V., Minelli, C., Relton, C.L., and Theodoratou, E. (2019). Guidelines for performing Mendelian randomization investigations. Wellcome Open Res *4*, 186. 10.12688/wellcomeopenres.15555.2.
24. Honigberg, M.C., Truong, B., Khan, R.R., Xiao, B., Bhatta, L., Vy, T.H., Guerrero, R.F., Schuermans, A., Selvaraj, M.S., Patel, A.P., et al. (2022). Genome-wide meta-analysis identifies novel maternal risk variants and enables polygenic prediction of preeclampsia and gestational hypertension. medRxiv. 10.1101/2022.11.30.22282929.
25. Magnussen, E.B., Vatten, L.J., Smith, G.D., and Romundstad, P.R. (2009). Hypertensive disorders in pregnancy and subsequently measured cardiovascular risk factors. Obstet Gynecol *114*, 961-970. 10.1097/AOG.0b013e3181bb0dfc.

26. Bellamy, L., Casas, J.P., Hingorani, A.D., and Williams, D.J. (2007). Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis. BMJ *335*, 974. 10.1136/bmj.39335.385301.BE.
27. Vikse, B.E., Irgens, L.M., Bostad, L., and Iversen, B.M. (2006). Adverse perinatal outcome and later kidney biopsy in the mother. J Am Soc Nephrol *17*, 837-845. 10.1681/ASN.2005050492.
28. Vikse, B.E., Irgens, L.M., Leivestad, T., Skjaerven, R., and Iversen, B.M. (2008). Preeclampsia and the risk of end-stage renal disease. N Engl J Med *359*, 800-809. 10.1056/NEJMoa0706790.
29. Haas, D.M., Ehrenthal, D.B., Koch, M.A., Catov, J.M., Barnes, S.E., Facco, F., Parker, C.B., Mercer, B.M., Bairey-Merz, C.N., Silver, R.M., et al. (2016). Pregnancy as a Window to Future Cardiovascular Health: Design and Implementation of the nuMoM2b Heart Health Study. Am J Epidemiol *183*, 519-530. 10.1093/aje/kwv309.
30. Shibuya, M. (2011). Vascular Endothelial Growth Factor (VEGF) and Its Receptor (VEGFR) Signaling in Angiogenesis: A Crucial Target for Anti- and Pro-Angiogenic Therapies. Genes Cancer *2*, 1097-1105. 10.1177/1947601911423031.
31. Zhao, Q., Wang, J., Hemani, G., Bowden, J., and Small, D.S. (2020). Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. Ann Statist *48*, 1742-1769. 10.1214/19-AOS1866.
32. Burgess, S., Butterworth, A., and Thompson, S.G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol *37*, 658-665. 10.1002/gepi.21758.
33. Verbanck, M., Chen, C.Y., Neale, B., and Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet *50*, 693-698. 10.1038/s41588-018-0099-7.
34. Pihl, K., Sorensen, S., and Stener Jorgensen, F. (2020). Prediction of Preeclampsia in Nulliparous Women according to First Trimester Maternal Factors and Serum Markers. Fetal Diagn Ther *47*, 277-283. 10.1159/000503229.
35. Jacobs, M., Nassar, N., Roberts, C.L., Hadfield, R., Morris, J.M., and Ashton, A.W. (2011). Levels of soluble fms-like tyrosine kinase one in first trimester and outcomes of pregnancy: a systematic review. Reprod Biol Endocrinol *9*, 77. 10.1186/1477-7827-9-77.
36. von Schoultz, B., Carlstrom, K., Collste, L., Eriksson, A., Henriksson, P., Pousette, A., and Stege, R. (1989). Estrogen therapy and liver function--metabolic effects of oral and parenteral administration. Prostate *14*, 389-395. 10.1002/pros.2990140410.
37. Oh, J.W., Kim, S.K., Cho, K.C., Kim, M.S., Suh, C.S., Lee, J.R., and Kim, K.P. (2017). Proteomic analysis of human follicular fluid in poor ovarian responders during in vitro fertilization. Proteomics *17*. 10.1002/pmic.201600333.
38. Wyatt, A.R., Cater, J.H., and Ranson, M. (2016). PZP and PAI-2: Structurally-diverse, functionally similar pregnancy proteins? Int J Biochem Cell Biol *79*, 113-117. 10.1016/j.biocel.2016.08.018.
39. Chalasani, N., Guo, X., Loomba, R., Goodarzi, M.O., Haritunians, T., Kwon, S., Cui, J., Taylor, K.D., Wilson, L., Cummings, O.W., et al. (2010). Genome-wide association study identifies variants associated with histologic features of nonalcoholic Fatty liver disease. Gastroenterology *139*, 1567-1576, 1576 e1561-1566. 10.1053/j.gastro.2010.07.057.

40. Eriksson, N., Benton, G.M., Do, C.B., Kiefer, A.K., Mountain, J.L., Hinds, D.A., Francke, U., and Tung, J.Y. (2012). Genetic variants associated with breast size also influence breast cancer risk. BMC Med Genet *13*, 53. 10.1186/1471-2350-13-53.
41. Comuzzie, A.G., Cole, S.A., Laston, S.L., Voruganti, V.S., Haack, K., Gibbs, R.A., and Butte, N.F. (2012). Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. PLoS One *7*, e51954. 10.1371/journal.pone.0051954.
42. Lauc, G., Huffman, J.E., Pucic, M., Zgaga, L., Adamczyk, B., Muzinic, A., Novokmet, M., Polasek, O., Gornik, O., Kristic, J., et al. (2013). Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. PLoS Genet *9*, e1003225. 10.1371/journal.pgen.1003225.
43. Collins, A.M., and Watson, C.T. (2018). Immunoglobulin Light Chain Gene Rearrangements, Receptor Editing and the Development of a Self-Tolerant Antibody Repertoire. Front Immunol *9*, 2249. 10.3389/fimmu.2018.02249.
44. Porcu, E., Medici, M., Pistis, G., Volpato, C.B., Wilson, S.G., Cappola, A.R., Bos, S.D., Deelen, J., den Heijer, M., Freathy, R.M., et al. (2013). A meta-analysis of thyroid-related traits reveals novel loci and gender-specific differences in the regulation of thyroid function. PLoS Genet *9*, e1003266. 10.1371/journal.pgen.1003266.
45. Nolan, J., Campbell, P.J., Brown, S.J., Zhu, G., Gordon, S., Lim, E.M., Joseph, J., Cross, S.M., Panicker, V., Medland, S.E., et al. (2021). Genome-wide analysis of thyroid function in Australian adolescents highlights SERPINA7 and NCOA3. Eur J Endocrinol *185*, 743-753. 10.1530/EJE-21-0614.
46. ACOG Practice Bulletin No. 202: Gestational Hypertension and Preeclampsia. (2019). Obstet Gynecol *133*, 1. 10.1097/AOG.0000000000003018.
47. Koopmans, C.M., Bijlenga, D., Groen, H., Vijgen, S.M., Aarnoudse, J.G., Bekedam, D.J., van den Berg, P.P., de Boer, K., Burggraaff, J.M., Bloemenkamp, K.W., et al. (2009). Induction of labour versus expectant monitoring for gestational hypertension or mild pre-eclampsia after 36 weeks' gestation (HYPITAT): a multicentre, open-label randomised controlled trial. Lancet *374*, 979-988. 10.1016/S0140-6736(09)60736-4.
48. Wu, P., Haththotuwa, R., Kwok, C.S., Babu, A., Kotronias, R.A., Rushton, C., Zaman, A., Fryer, A.A., Kadam, U., Chew-Graham, C.A., and Mamas, M.A. (2017). Preeclampsia and Future Cardiovascular Health: A Systematic Review and Meta-Analysis. Circ Cardiovasc Qual Outcomes *10*. 10.1161/CIRCOUTCOMES.116.003497.
49. Sun, B.B., Maranville, J.C., Peters, J.E., Stacey, D., Staley, J.R., Blackshaw, J., Burgess, S., Jiang, T., Paige, E., Surendran, P., et al. (2018). Genomic atlas of the human plasma proteome. Nature *558*, 73-79. 10.1038/s41586-018-0175-2.
50. Chen, L., Peters, J.E., Prins, B., Persyn, E., Traylor, M., Surendran, P., Karthikeyan, S., Yonova-Doing, E., Di Angelantonio, E., Roberts, D.J., et al. (2022). Systematic Mendelian randomization using the human plasma proteome to discover potential therapeutic targets for stroke. Nat Commun *13*, 6143. 10.1038/s41467-022-33675-1.
51. Haas, D.M., Parker, C.B., Wing, D.A., Parry, S., Grobman, W.A., Mercer, B.M., Simhan, H.N., Hoffman, M.K., Silver, R.M., Wadhwa, P., et al. (2015). A description of the methods of the Nulliparous Pregnancy Outcomes Study: monitoring mothers-to-be (nuMoM2b). Am J Obstet Gynecol *212*, 539 e531-539 e524. 10.1016/j.ajog.2015.01.019.
52. Guerrero, R.F., Khan, R.R., Wapner, R.J., Hahn, M.W., Raja, A., Salleb-Aouissi, A., Grobman, W.A., Simhan, H., Silver, R.M., Chung, J.H., et al. (2022). Genetic

Polymorphisms Associated with Adverse Pregnancy Outcomes in Nulliparas. medRxiv. https://doi.org/10.1101/2022.02.28.22271641.
53. Conomos, M.P., Miller, M.B., and Thornton, T.A. (2015). Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet Epidemiol *39*, 276-293. 10.1002/gepi.21896.
54. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867-2873. 10.1093/bioinformatics/btq559.
55. 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. 10.1038/s41586-018-0579-z.
56. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature *590*, 290-299. 10.1038/s41586-021-03205-y.
57. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet *44*, 955-959. 10.1038/ng.2354.
58. Peterson, R.E., Kuchenbaecker, K., Walters, R.K., Chen, C.Y., Popejoy, A.B., Periyasamy, S., Lam, M., Iyegbe, C., Strawbridge, R.J., Brick, L., et al. (2019). Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. Cell *179*, 589-603. 10.1016/j.cell.2019.08.051.
59. Gogarten, S.M., Sofer, T., Chen, H., Yu, C., Brody, J.A., Thornton, T.A., Rice, K.M., and Conomos, M.P. (2019). Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics *35*, 5346-5348. 10.1093/bioinformatics/btz567.
60. Wojcik, G.L., Graff, M., Nishimura, K.K., Tao, R., Haessler, J., Gignoux, C.R., Highland, H.M., Patel, Y.M., Sorokin, E.P., Avery, C.L., et al. (2019). Genetic analyses of diverse populations improves discovery for complex traits. Nature *570*, 514-518. 10.1038/s41586-019-1310-4.
61. Conomos, M.P., Reiner, A.P., Weir, B.S., and Thornton, T.A. (2016). Model-free Estimation of Recent Genetic Relatedness. Am J Hum Genet *98*, 127-148. 10.1016/j.ajhg.2015.11.022.
62. Conomos, M.P., Reiner, A.P., McPeek, M.S., and Thornton, T.A. (2018). Genome-wide control of population structure and relatedness in genetic association studies via linear mixed models with orthogonally partitioned structure. bioRxiv. 10.1101/409953.
63. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res *38*, e164. 10.1093/nar/gkq603.
64. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7. 10.1186/s13742-015-0047-8.
65. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W., Jansen, R., de Geus, E.J., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet *48*, 245-252. 10.1038/ng.3506.
66. Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. PLoS Genet *9*, e1003348. 10.1371/journal.pgen.1003348.

67. Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol *44*, 512-525. 10.1093/ije/dyv080.
68. Burgess, S., Davies, N.M., and Thompson, S.G. (2016). Bias due to participant overlap in two-sample Mendelian randomization. Genet Epidemiol *40*, 597-608. 10.1002/gepi.21998.
69. Carter, A.R., Sanderson, E., Hammerton, G., Richmond, R.C., Davey Smith, G., Heron, J., Taylor, A.E., Davies, N.M., and Howe, L.D. (2021). Mendelian randomisation for mediation analysis: current methods and challenges for implementation. Eur J Epidemiol *36*, 465-478. 10.1007/s10654-021-00757-1.
70. MacKinnon, D.P., Lockwood, C.M., Hoffman, J.M., West, S.G., and Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. Psychol Methods *7*, 83-104. 10.1037/1082-989x.7.1.83.