BMC Pregnancy and Childbirth

RESEARCH

Open Access

Early prediction of preeclampsia from clinical, multi-omics and laboratory data using random forest model

Qiang Zhao[1,2†], Jia Li[3,4†], Zhuo Diao[4†], Xiao Zhang[3,4†], Suihua Feng[1†], Guixue Hou[4], Wenqiu Xu[3,4], Zhiguang Zhao[3,4], Zhixu Qiu[3,4], Wenzhi Yang[3,4], Si Zhou[3,4], Peirun Tian[4], Qun Zhang[1], Weiping Chen[1], Huahua Li[1], Gefei Xiao[5], Jie Qin[5], Liqing Hu[5], Zhongzhe Li[6], Liang Lin[4], Shunyao Wang[4,7], Ruyun Gao[8], Wuyan Huang[5], Xiaohong Ruan[1,2*], Sufen Zhang[5*], Jianguo Zhang[3,4,9*], Lijian Zhao[3,4,10*] and Rui Zhang[11*]

Abstract

Background Predicting preeclampsia (PE) within the first 16 weeks of gestation is difficult due to various risk factors, poorly understood causes and likely multiple pathogenic phenotypes of preeclampsia.

Objectives In this study, we aimed to develop prediction models for early-onset preeclampsia (EPE) and late-onset preeclampsia (LPE) respectively using clinical data, metabolome and proteome analyses on plasma samples and laboratory data.

Methods We retrospectively recruited 56 EPE, 50 LPE patients and 92 normotensive controls from three tertiary hospitals and used clinical and laboratory data in early pregnancy. Models for EPE and LPE were fitted with the use of patient' clinical, multi-omics and laboratory data.

Results By comparing multi-omics and laboratory test variables between EPE, LPE and healthy controls, we identified sets of differentially expressed biomarkers, including 49 and 33 metabolites, 28 and 36 proteins as well as 5 and 7 laboratory variables associated with EPE and LPE respectively. Using the random forest algorithm, we developed a prediction model using seven clinical factors, seven metabolites, five laboratory test variables. The model yielded the highest accuracy for EPE prediction with good sensitivity (87.5%, 95% confidence interval [CI]: 67.64%-97.34%) and specificity (94.1%, 95% CI: 80.32%-99.28%). We also developed a prediction model that exhibited high accuracy in separating LPE from controls (sensitivity: 66.67%, 95% CI: 43.03%-85.41%; specificity: 94.12%, 95% CI: 80.32%-99.28%) using seven clinical factors, five metabolites and eight proteins.

Conclusion Our study has identified a set of significant omics and laboratory features for PE prediction. The established models yielded high prediction performance for preeclampsia risk from clinical, multi-omics and laboratory information.

Keywords Preeclampsia, Proteome, Metabolome, Laboratory data, Random forest

†Qiang Zhao, Jia Li, Zhuo Diao, Xiao Zhang and Suihua Feng contribute equally to this study and share first authorship.

*Correspondence:
Xiaohong Ruan
13924680902@139.com
Sufen Zhang
chouchouxinran@126.com
Jianguo Zhang
zhangjg@bgi.com
Lijian Zhao
zhaolijian@bgi.com
Rui Zhang
zhangrui0829@163.com
Full list of author information is available at the end of the article

Introduction

Pre-eclampsia (PE) is a complex and heterogeneous multisystem disease characterized by new development of hypertension after 20 weeks of gestation and one of the PE-related complications, including proteinuria, maternal organ dysfunction or uteroplacental dysfunction, such as angiogenic imbalance or fetal growth restriction. The global incidence of PE is approximately 4.6%, ranging from 1% to 5.6% [1]. Pre-eclampsia is one of the most severe complications during pregnancy, leading to a large number of maternal and perinatal morbidities and mortalities [2]. An estimated 4 million pre-eclampsia cases occur annually, leading to the mortalities of over 70,000 women and 500,000 babies worldwide [2, 3]. PE is commonly classified into two subgroups, including the early-onset PE (EPE) and late-onset PE (LPE), according to the gestational weeks at clinical presentation (34 weeks of gestation). There is an emerging evidence that EPE is thought to be a consequence of impaired placentation [4], whereas the metabolic syndrome with increased insulin resistance is the main pathophysiological processes in LPE [5, 6].

Over the past three decades, though numerous studies have been performed to investigate the pathophysiology of preeclampsia and the understanding of the disease has been remarkably improved, it remains not completely understood with respect to specific biological processes implicated in the development of PE. Accurate prediction of PE in early pregnancy has remained highly challenging, possible reasons might include incomplete understanding of the disease, a variety of risk factors, and likely different pathogenic phenotypes of PE [7, 8]. The rapid development of high-throughput omics assays has enabled integrated analyses of the high-dimensional multi-omics data [9, 10] and may capture complex dynamic processes implicated in the preeclampsia. Furthermore, the most predictive features are probably discovered from high-dimensional multi-omics using machine-learning methods, which might accelerate the development of more precise prediction models. Two recent studies have investigated the predictive value of laboratory data for PE and their models presented a relatively poor performance for PE screening [11, 12], however, these studies didn't include liver and kidney dysfunction markers which are important predictors for PE in early pregnancy [13–15]. So far, it remains not completely understood regarding the predictive values of the laboratory markers alone and in combination with multi-omics markers.

In this study, we performed proteome, metabolome assays on a set of biospecimens collected retrospectively from preeclamptic and normotensive pregnant women and a multi-omics data analysis to discover sets of metabolites and proteins predictive of PE in early pregnancy; and then, we performed a joint analysis of the multi-omics data with the available clinical/laboratory data to establish integrated predictive models based on a small number of clinical characteristics, protein and metabolite biomarkers and laboratory test variables. Finally, we compared prediction capabilities of different combinations of predictors to achieve the best accuracy for early and accurate detection of PE in pregnant women and eventually guide therapeutic intervention.

Methods and materials

Participants and study design

Maternal peripheral blood (5 mL) was collected in Streck Cell Free DNA BCT ® blood collection tubes (Streck, La Vista, NE, USA) and stored in the refrigerators at 4 degrees for non-invasive prenatal test (NIPT) and processed within four days. The remaining plasma samples were stored in the refrigerators at − 20 degree. We retrospectively reviewed the medical records of all pregnant women who underwent NIPT tests at $11–15^{+6}$ gestational weeks in Zhuhai Maternal and Child Health Hospital, Shenzhen Bao'an District Maternal and Child Health Hospital, Jiangmen Central Hospital between January 1, 2019, and December 30, 2021 and recruited 56 EPE, 50 LPE patients and 92 normotensive controls. Maternal characteristics, demographics, gestational ages at delivery and birth weight were retrospectively retrieved from medical records by the physicians. The participants were randomly split into two datasets, including the training set and test set at a ratio of 3:2. The training set was used to identify and select potential protein and metabolite biomarkers and train the models for predicting PE. The test set was utilized to confirm the proteomic and metabolomics results and assess the performance of the established models (Fig. 1). Informed consent was waived by the Ethics Committees of Beijing Genomics Institute. This study was approved by the Ethics Committees of Beijing Genomics Institute (BGI-IRB 22026). All methods were performed in accordance with the Declaration of Helsinki.

The PE patients were diagnosed and enrolled in this study following the guidelines of the International Society for the Study of Hypertension in Pregnancy [2]. These patients developed high blood pressure after 20 gestational weeks, with the systolic blood pressure above 140 mm Hg and/or the diastolic blood pressure above 90 mm Hg on at least two occasions 4 h apart; PE patients also presented proteinuria of 300 mg or more in 24 h, in case of lack of 24 h urine protein quantitation, two readings of at least + + on dipstick analysis of urine specimens were required for the diagnosis of PE. EPE and LPE were classified based on the clinical manifestations of preeclampsia developed before and after 34 weeks of gestation,

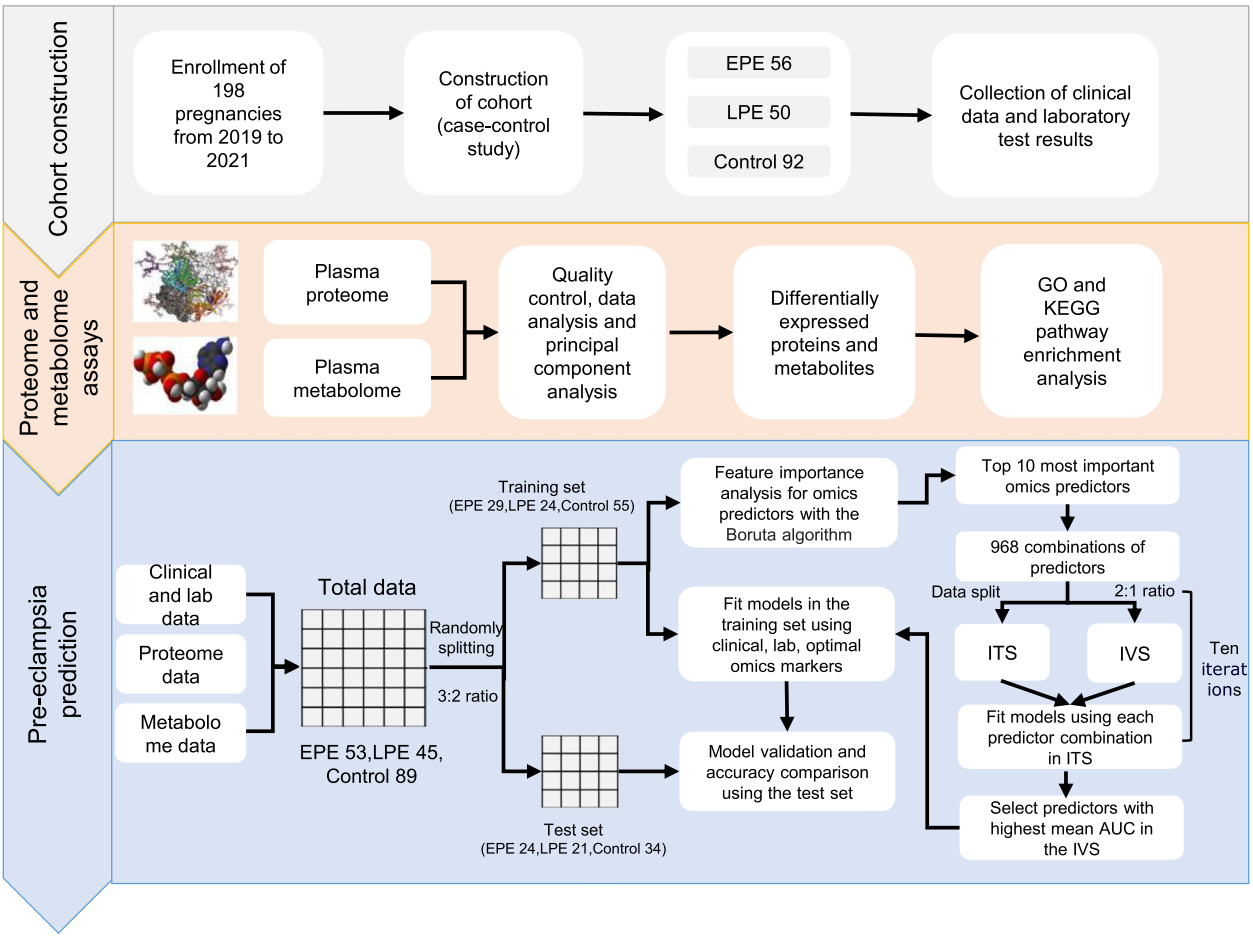

Fig. 1 The schematic workflow of this study. 198 pregnancies were enrolled from 2019 to 2021 in this study, these participants comprise 56 EPE, 50 LPE and 92 healthy controls. Clinical and laboratory data of each participant were retrieved from health information system by physicians. Plasma samples were collected and underwent proteome and metabolome assays according to the manufacturer's instructions. Protein and metabolite expression were analyzed using the Spectronaut software with the default parameters. Then differentially expressed proteins and metabolites were identified, followed by GO and KEGG pathway enrichment analysis. The total data were split into the training and test sets at a ratio of 3:2, then feature importance analysis was performed on all proteomic and metabolic biomarkers in the training dataset. The top ten most important metabolic or proteomic biomarkers were selected to build 968 predictor combinations (> 2 biomarkers) separately. The training dataset was randomly split into an internal training set (ITS) and internal validation set (IVS) at a ratio of 2:1. For each combination of metabolic or proteomic predictors, a random forest model was built in the ITS and validated in the IVS. The process was repeated 10 times, generating 10 prediction models and their corresponding area under the curves (AUC) values. The feature combination with highest mean AUC value in the IVS was considered as the optimal set of biomarkers for the construction of the prediction models in the training set. Lastly, the final models were established using different combinations of clinical factors, the optimal combination of proteomic, metabolic biomarkers and laboratory test variables in the training dataset, the performances of the established models were independently evaluated in the test dataset

respectively. Healthy control was defined as a full-term pregnancy without obstetric, medical or surgical complications during pregnancy.

Plasma proteome profiling by LC–MS/MS

Plasma samples were processed with SPE columns (Waters, USA) for enrichment of low-abundance proteins as mentioned in previous reports [16, 17]. Proteins were subsequently reduced by dithiothreitol in 56 °C water bath for 30 min and alkylated by iodoacetamide in the darkroom at room temperature for 30 min. After dilution, proteins were digested by trypsin (Promega, USA) and desalting using Strata-X column (Agela, China). All samples were then conducted by Orbitrap Lumos mass spectrometer (Thermo Scientific, San Jose, USA) coupled with an Ultimate 3000 UHPLC liquid chromatography (Thermo Scientific, San Jose, USA). Peptide separation was performed using a self-packed analytical column (1.7 μm, 150 mm x 30 cm) at a flow rate of 500 nL/min. The mobile phase consisted of two

steps: phase A containing 0.1% formic acid and 2% acetonitrile in water, and phase B comprising 0.1% formic acid in acetonitrile with a 50-min elution gradient. The settings were as follows: 0–5 min, 5% B; 5–45 min, 5–25% B; 45–55 min, 25–35% B. The MS1 mass range was set at 400–1250 m/z with a resolution of 60,000 and a maximum injection time of 50 ms. The mass range of 400–1250 m/z was split into 45 continuous windows for MS2 scans with resolution setting as 30,000 and automatic gain control (AGC) of 1E6 for the DIA setting. Normalized collision energy of MS2 was assigned to 22.5, 25, and 27.5. The DIA-NN software [18] was used for DIA data analysis with self-built plasma spectral library using plasma from pregnant subjects containing 4,979 proteins and 34,268 precursors). The FDR cutoff was set at 1% for both peptide and protein levels.

Targeted metabolites quantification by LC–MS/MS

Sample preparation

Targeted metabolites quantitative detection was performed using BGI HM400 kit. Calibrator I was reconstituted with 150 μL of 50% methanol solution and mixed with 150 μL Calibrator II, and the mixture was shaken at room temperature for 20 min at 1200 rpm. 100 μL of the mixture was taken out and diluted with 75% methanol solution as follows: 1/2, 1/4, 1/8, 1/16, 1/32, 1/160, 1/320, 1/640, 1/1280, 1/2560 to obtain 11 concentration gradient calibrator mixtures. 20 μL of ultrapure water was added to well A1 of a 96-well plate; 20 μL calibrator mixed solutions of 11 concentrations were added to wells A2 to A12 according to the concentration from low to high; 20 μL of plasma samples were added to other wells. The internal standard I was reconstituted with 1 mL of 50% methanol, and then added to 13 mL of methanol to obtain the sample release agent. Then, 120μL of sample release agent was added to each of the above wells. The plate was shaken at 10 °C 600 rpm for 20 min and centrifuged at 4000 g for 20 min at 4 °C. After centrifugation, 30 μL of the supernatant was transferred to a new 96-well plate.

3 mL of derivatization reagent diluent was added into the derivatization reagent bottle, the mixture was shaken to dissolve, the derivatization reagent working solution was obtained. 3 mL of the EDC diluent was added to the EDC reagent bottle, the powder was dissolved to obtain the EDC working solution. 20 μL of derivatization reagent working solution and 20 μL of EDC working solution were added sequentially to each well of the new 96-well plate containing supernatant. The plate was covered with an aluminum film, placed in a constant temperature shaker and shaken at 40 °C 1200 rpm for 60 min. After the reaction was completed, the plate was cooled to room temperature and centrifuged at 2000 g for 5 min. 30 μL reaction solution was transferred to another new 96-well plate, 90 μL of 50% methanol was added to each well. The plate was covered with an aluminum film and mixed at 10 °C 600 rpm for 5 min.

LC–MS Acquisition and data quantification

Metabolites extracted from plasma samples and derivatized were detected and quantified using targeted profiling strategies analysis strategy by LC–MS platform [19]. The 96-well plate finally obtained by the above preparation, which is covered by aluminum film can be directly used for LC–MS detection. QC samples were prepared by pooling equal volumes of each sample to evaluate the reproducibility of the analysis. Quantification of samples were then conducted by SCIEX Triple Quad 6500 mass spectrometer coupled with an Waters ACQUITY ICLASS UPLC using the MRM mode, and chromatographic separation was performed on a Waters ACQUITY UPLC BEH C18 1.7 μm 2.1 × 100 mm column at a flow rate of 0.4 mL/min. The mobile phase consisted of two parts: phase A containing 0.1% formic acid in water; phase B consisting of 70% acetonitrile and 30% isopropanol with a 18-min elution gradient. The settings were as follows: 0–1 min, 5% B; 1–5 min, 5–30%B; 5–9 min, 30–50% B; 9–11 min, 50–78%B; 11–13.5 min, 78–95%B; 13.5–14 min; 95–100% B; 14–16 min, 100%B; 16–16.1 min, 100–5%B; 16.1–18 min, 5%B. The mass spectrometry method included positive/negative ion methods. All ion transitions and corresponding parameters were set according to the methods provided by the kit. Batch sequence was edited according to the sample format in instruction manual. After detection, format conversion of the wiff data was performed. Mass spectrometry data were quantified using the HMQuant quantitative software.

Differential expression analysis and pathway enrichment analysis

Principal components analysis on the proteome or metabolome data matrix was performed using the prcomp function in R 3.6.0. Samples whose principal component 1 and 2 values are within mean ± 3*standard deviation (SD) of principal component 1 and 2 values were included in the downstream analysis. This step eliminated 3 EPE, 5 LPE and 3 healthy control samples from five mass spectrometry batches (Figure S1). Differentially expressed metabolites and proteins were identified using Wilcoxon rank sum test following the cutoff of $P < 0.05$. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis on identified metabolites was carried out using MetaboAnalyst [20]. Over-representation analysis of the metabolites was performed using the hypergeometric test in MetaboAnalyst. P value < 0.05 was considered as statistically significant.

Gene ontology (GO) term enrichment analysis on the identified differentially expressed proteins was conducted using topology-based Gene Ontology scoring (topGo) [21], an R software package. *P* value < 0.05 was considered as statistically significant. Proteome pathway enrichment analysis was performed using the online tool g:profiler [22]. Adjusted *p* value < 0.05 was considered as statistically significant. Correlations between metabolites and proteins predicted significantly differentially expressed between EPE and LPE samples were analyzed using cor.test. Metabolite-protein interaction network was built with *igraph* using the metabolite-protein interactions with correlation coefficient above 0.2.

Analysis of clinical and laboratory data

All participants underwent routine laboratory tests in early pregnancy, their clinical and laboratory data were retrieved from health information system by physicians. The clinical data analyzed in the study consist of participant' age, body mass index (BMI), diastolic blood pressure(DBP), systolic blood pressure(SBP), birth times, recurrent pregnancy loss (RPL, > 2 times), twin pregnancy, in vitro fertilization(IVF), past medical history (PMH) which involves at least one of the following complications: diabetes mellitus, PE, family history of PE, chronic hypertension, systemic lupus erythematosus and antiphospholipid syndrome, birth weight and gestational weeks at birth. BMI, DBP and SBP values were measured between 11 and 15 + 6 weeks of gestation. Mean arterial pressure (MAP) was calculated following the equation below [23]:

$$\mathrm{MAP} = \mathrm{DBP} + \mathrm{PP} \times \frac{(27.07 + 0.181 \times \mathrm{DBP} + 2.303)}{100}$$

in which, PP is the difference between systolic and diastolic blood pressure. The laboratory data include 46 routine prenatal laboratory test results from routine blood test, hepatic and renal function tests, routine urine test, urine sediments analysis, hepatitis B antigens and antibodies. The laboratory data have an average missing rate of 23.66% across 46 prenatal laboratory test results. Missing values were replaced with the median value of each laboratory test variable. Fisher exact test and Wilcoxon sum rank test were utilized to investigate categorical and continuous variables respectively. P < 0.05 was considered statistically significant.

Feature selection with the Boruta algorithm

The Boruta algorithm was established to investigate the predictive importance of variables in a classification framework. It involves duplicating and shuffling features to eliminate correlations with the response. A random forest classifier is then applied to the extended dataset, collecting z-score values for variable importance. A two-sided test compares the importance of each real variable with the maximum z-score value of shadow variables (MZSA). Variables showing significantly higher importance than MZSA are considered important, while those with significantly lower importance are deemed unimportant. Unimportant features and shadow attributes are permanently removed from the analysis. These steps are repeated until all attributes have their importance values [24].

The establishment and validation of random forest models

Machine-learning analyses consist of two main steps, including the predictor selection step and model development step. Firstly, in order to minimize the impact of the predictor scale difference on prediction models, patient' age, BMI, MAP, omics and laboratory test variables were normalized by division of raw values by their corresponding median values of healthy controls. Then, the top ten most important proteomic or metabolic biomarkers evaluated by the Boruta algorithm were selected and combined randomly to establish 968 combinations of proteomic or metabolic biomarkers (> 2 biomarkers) respectively. The training set was split into an internal training dataset and internal validation dataset at a ratio of 2:1. Each combination of omics biomarkers was utilized to establish a random forest model in the internal training set and its performance was evaluated in the internal validation set. The process was repeated 10 times, generating 10 prediction models and their corresponding area under the curves (AUC) values. Proteomic/metabolic predictors with highest mean AUC value in the internal validation set was deemed most predictive for PE. Secondly, prediction models were developed for clinical factors dataset, to investigate possible gains from integration of clinical, proteomic, metabolic biomarkers and laboratory test predictors, random forest models that take different combinations of clinical characteristics, omics and laboratory test variables as input were fit and independently verified in the test set. The receiver operating characteristic curves (ROC) of the classifiers were drawn and AUC values were calculated with the python package sklearn [25]. Correlations between predictors and risk scores predicted by the random forest models were analyzed using cor.test and visualized using the R package pheatmap.

Results

Identification of clinical risk factors for PE

Maternal characteristics, demographics, birth weight and gestational ages at delivery are shown in Table 1. There

Table 1 Comparison of maternal obstetric characteristics and pregnancy outcome of the women who did and did not develop PE and healthy pregnant women

Clinical feature		EPE ($n = 53$)	LPE ($n = 45$)	Control ($n = 89$)
Gestational weeks at sampling		13.28 ± 0.94**	13.73 ± 1.07	13.89 ± 1.18
Age at sampling		32.47 ± 5.21***	30.78 ± 4.56	29.5 ± 3.71
BMI		22.19 ± 3.24***	22.07 ± 5.6*	19.77 ± 4.21
MAP		92.91 ± 10.05***	89.93 ± 9.15***	81.83 ± 8.69
Primiparous woman	No	22(41.51%)	13(28.89%)	41(46.07%)
	Yes	31(58.49%)	32(71.11%)	48(53.93%)
Recurrent pregnancy loss	No	42**(79.25%)	39(86.67%)	85(95.51%)
	Yes	11(20.75%)	6(13.33%)	4(4.49%)
Past medical history	No	50(94.34%)	45(100%)	89(100%)
	Yes	3(5.66%)	0	0
IVF	No	45 (84.91%)	37 (82.22%)	83 (93.%)
	Yes	8 (15.09%)	8 (17.78%)	6 (6.74%)
Twin pregnancy	No	46*** (86.79%)	42* (93.33%)	89 (100%)
	Yes	7 (13.21%)	3 (6.67%)	0
Gestational age at delivery (weeks)		32.06 ± 3.03***	38.02 ± 1.55***	39.02 ± 1.03
Birthweight (g)		1558.3 ± 615.77***	2760.67 ± 480.19***	3185 ± 335.61

Comparison of maternal obstetric characteristics and pregnancy outcome was carried out between each PE type and healthy control

*, **, *** denotes *P* value < 0.05, < 0.01 and < 0.001 respectively. Recurrent pregnancy loss is defined as a woman who has more than 2 pregnancy losses. Past medical history denotes a woman has at least one of the following complications: medical history of pregnancy diabetes, pre-eclampsia, family history of pre-eclampsia, chronic hypertension, systemic lupus erythematosus and antiphospholipid syndrome

was no significant difference in gestational weeks at sampling between LPE and control groups (Table 1, $P > 0.05$, Wilcoxon rank sum test). As compared with healthy controls, the EPE and LPE patients presented an older age, higher BMI and MAP values, higher prevalence of twin pregnancies-all known risks for PE (Table 1, $P < 0.05$ for all cases, Wilcoxon rank sum test or Fisher exact test). Additionally, we observed the participants with RPL, PMH and IVF were more likely to develop PE, however, the result didn't reach statistical significance, probably because the sample size of this study is relatively small. As for the comparison between the EPE and LPE women, we didn't observe there were significant differences in age, BMI, MAP, and other clinical factors except birth weight and gestational ages at delivery between the two subtypes of PE samples (Table S1).

Identification of metabolic biomarkers for PE

First, we aimed to identify the PE-associated metabolites and performed differential expression analysis in the training set. In the metabolome set containing 165 metabolites, 43 metabolites were significantly differentially expressed between EPE and heathy controls, with 35 upregulated and 8 downregulated (P value < 0.05, Wilcoxon rank-sum test, Fig. 2A). L-Malic acid, erythronic acid, palmitoylcarnitine, ornithine, 2-Hydroxy- 3-methylbutyric acid were the top five most significantly differentially expressed metabolites (Fig. 2B-F), Pathway enrichment analysis on these metabolic markers uncovered the following pathways ($p < 0.05$): Arginine biosynthesis, Tyrosine metabolism, Citrate cycle (TCA cycle), Alanine, aspartate and glutamate metabolism, beta-Alanine metabolism (Fig. 2G, the hypergeometric test, P value < 0.05). With respect to the metabolites associated with LPE, we identified 33 metabolites showing significant difference in metabolite expression between LPE and heathy controls (Fig. 2H, P value < 0.05, Wilcoxon rank sum test), of which, Indole- 3-butyric acid, tartaric acid, levulinic acid, 2-Hydroxy- 2-methylbutyric acid and m-Coumaric acid rank the top five (Fig. 2I-M). These metabolites were significantly enriched in 12 KEGG pathways, such as Alanine, aspartate and glutamate metabolism, Citrate cycle (TCA cycle), Arginine and proline metabolism, beta-Alanine metabolism and Phenylalanine, tyrosine and tryptophan biosynthesis (Fig. 2N, the hypergeometric test, P value < 0.05). Further analysis of the differentially expressed metabolites revealed 17 metabolites were differentially expressed in both EPE and LPE, such as ornithine, 2-Hydroxy-3-methylbutyric acid, 2-Hydroxy- 2-methylbutyric acid, homovanillic acid. While, 26 and 16 metabolites were differentially expressed in only EPE, such as palmitoylcarnitine and stearylcarnitine(C18) and LPE respectively, such as l-Pipecolic acid and tartaric acid (Figure S2). We also compared differences in the identified metabolic markers between the EPE and LPE samples, Gentisic acid,

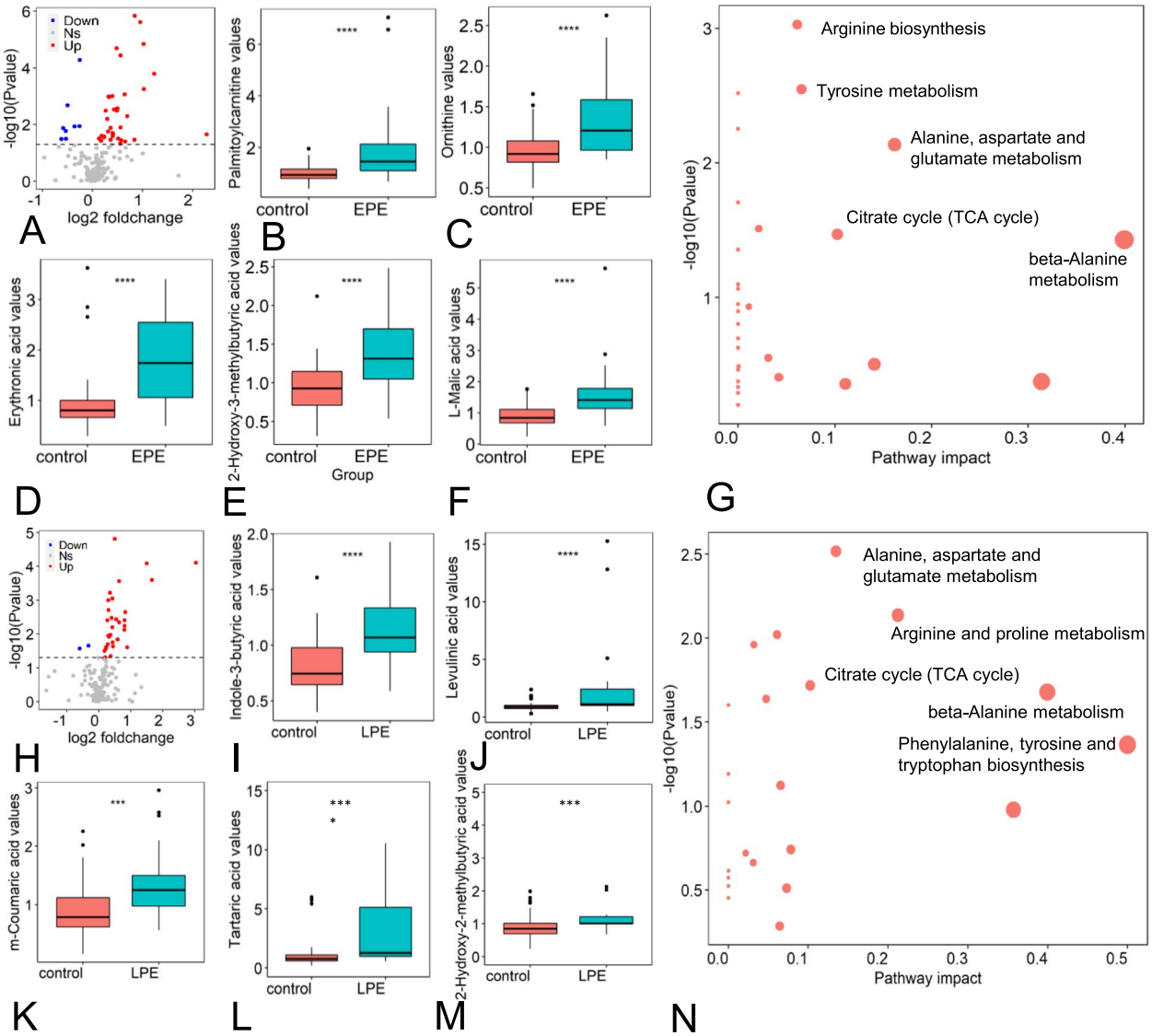

Fig. 2 Analysis of differentially expressed metabolites related with PE. **A**. Volcano plot of differentially expressed metabolites and associated *P* values and log2 fold change values for EPE patients. The dashed line represents *P* value < 0.05. Down, Ns, Up denote down-regulated, not significant, up-regulated metabolites respectively. ****: $p < 0.0001$. **B**-**F**. The expression differences of top five most differentially expressed metabolites between EPE patients and controls. **G**. the KEGG pathways significantly enriched for differentially expressed metabolites in the EPE cohort. **H**. Volcano plot of differentially expressed metabolites and associated *P* values and log2 fold change values for LPE patients. **I**-**M**. The expression differences of top five most differentially expressed metabolites between LPE patients and healthy controls (***: $p < 0.001$, ****: $p < 0.0001$). N. the KEGG pathways significantly enriched for differentially expressed metabolites in the LPE cohort

D-Glucose, Tartaric acid and L-Glutamine were significantly up-regulated, Trehalose, Palmitoylcarnitine and Stearylcarnitine(C18) were significantly down-regulated in LPE samples as compared to EPE ones ($P < 0.05$ for all cases, Wilcoxon rank sum test, Figure S3).

Identification of proteomic biomarkers for PE

In the proteome set containing 474 proteins, in early pregnancy, 28 proteins (15 upregulated and 13 downregulated) exhibited expression changes significantly associated with EPE in the training dataset as compared to healthy controls (*P* value < 0.05, Wilcoxon rank sum test, Fig. 3A). Superoxide dismutase 3 (SOD3), Macrophage migration inhibitory (MIF) factor, Neurogranin (NRGN), Hemoglobin Subunit Delta (HBD), Vasorin (VASN) were the top five most significantly differentially expressed proteins (Fig. 3B-F). GO term enrichment analysis on these proteins identified

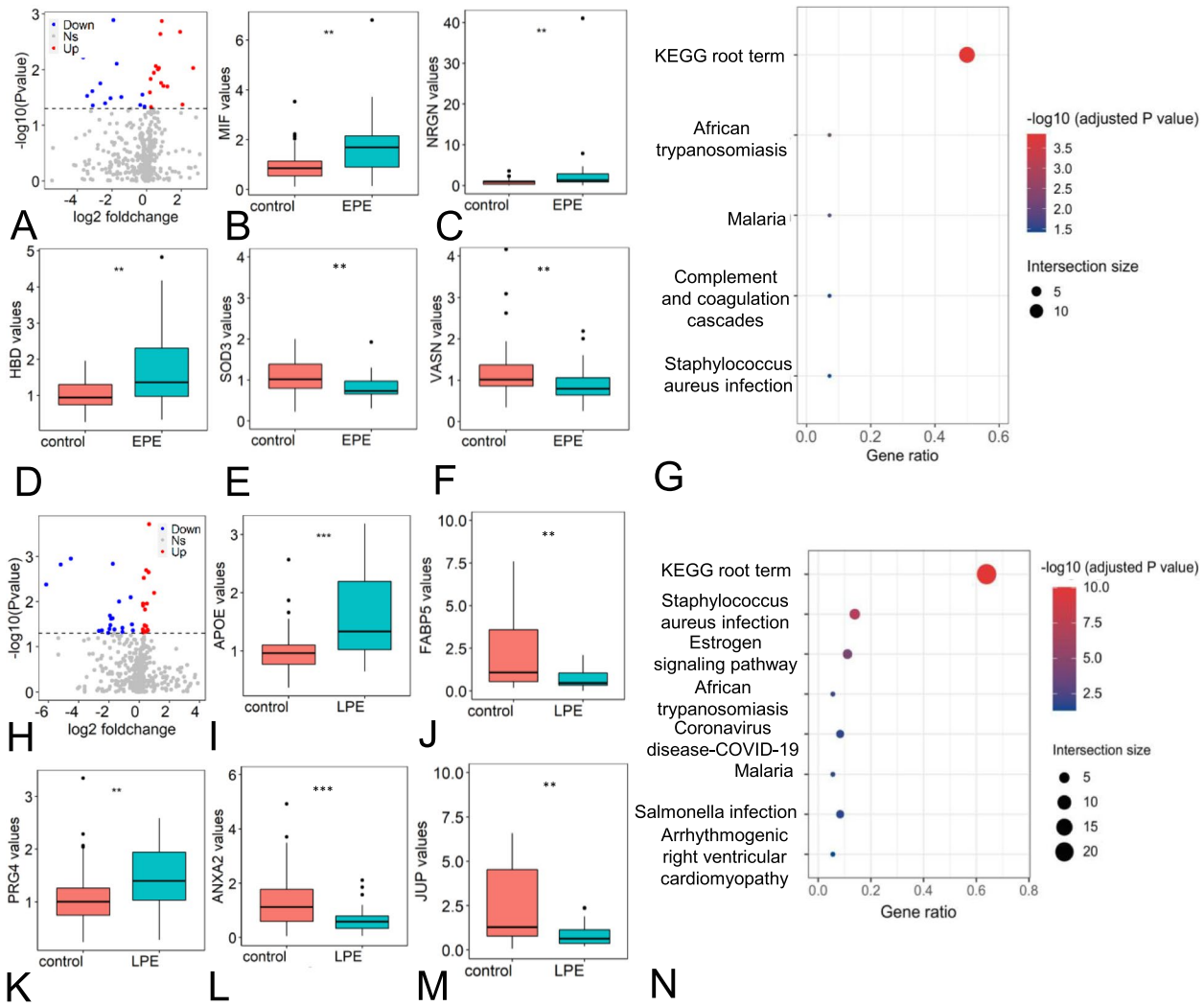

Fig. 3 Analysis of differentially expressed proteins related with PE. **A**. Volcano plot of differentially expressed proteins and associated *P* values and log2 fold change values for EPE patients. **B**-F. The expression differences of top five most differentially expressed proteins between EPE patients and controls. **: $p < 0.01$. **G**. the KEGG pathways significantly enriched for differentially expressed proteins in the EPE cohort. **H**. Volcano plot of differentially expressed proteins and associated *P* values and log2 fold change values for LPE patients. The dashed line represents *P* value < 0.05. **I**-**M**. The expression differences of top five most differentially expressed proteins between LPE patients and healthy controls (**: $p < 0.01$, ***: $p < 0.001$). N. The KEGG pathways significantly enriched for differentially expressed proteins in the LPE cohort

3 significant GO terms, cellular response to stimulus (GO:0051716), ion transport (GO:0006811), response to stress (GO:0006950) (Figure S4, Fisher's exact test, *P* value < 0.05). Further analysis of the 3 GO terms revealed that cellular response to stimulus (GO:0051716) and response to stress (GO:0006950) were downstream biological responses to stimulus (GO:0050896) (Figure S4). Pathway enrichment analysis on these proteins uncovered the following pathways: KEGG root term, African trypanosomiasis, Malaria, Complement and coagulation cascades, Staphylococcus aureus infection (Fig. 3G, Fisher's exact test, *P* value < 0.05). With respect to the proteins associated with LPE, we identified 36 proteins showing significant difference in protein expression between LPE and heathy controls, of which, Apolipoprotein E (Apo-E), Junction Plakoglobin (JUP), Annexin A2(ANXA2), Fatty acid binding protein 5 (FABP5) and Proteoglycan 4 (PRG4) rank the top five (Fig. 3H-M, *P* value < 0.05, Wilcoxon rank sum test). These proteins were significantly enriched in 55 GO terms (Figure S5, Fisher's exact test, *P* value < 0.05) and 8 KEGG pathways (Fig. 3N, Fisher's exact test, *P* value < 0.05), such as KEGG root term, Staphylococcus aureus infection, Estrogen signaling pathway, African trypanosomiasis and Coronavirus disease—COVID- 19. Further analysis showed that 10 proteins were differentially expressed in both EPE and LPE, such as MIF, Hemoglobin Subunit Alpha 1 (HBA1), Hemoglobin Subunit Beta (HBB), HBD,

and Complement Factor D (CFD). Additionally, 18 and 26 proteins were differentially expressed in only EPE and LPE, respectively (Figure S6). Additionally, we identified five up-regulated proteins, including APOE, GP5, PRG4, PSG4, SOD3, and four down-regulated proteins, including FABP5, NRGN, RPLP2, TXN in LPE samples as compared to EPE ones ($P < 0.05$ for all cases, Wilcoxon rank sum test, Figure S7). We performed interaction analysis between the above mentioned 7 metabolomic and 9 proteomic markers significantly differentially expressed between EPE and LPE samples. FABP5 and D-Glucose presented higher frequencies of interactions with surrounding proteins and metabolites separately, suggesting they may play an important role in the pathophysiological mechanism of PE (Figure S8).

Identifying laboratory test variables for PE

Forty-six prenatal laboratory test results were obtained from routine prenatal laboratory data in early pregnancy. A total of 5 laboratory test variables were significantly different between EPE and healthy controls ($P < 0.05$, Wilcoxon rank sum test or Fisher exact test, Table S2), Creatinine (CRE), Eosinophils (EO), monocytes (MO) are the top three lab test variables showing the largest difference. Moreover, 7 clinical laboratory test results were found associated with LPE, with CRE, MO and hematocrit (Hct) ranking the top three ($P < 0.05$, Wilcoxon rank sum test or Fisher exact test, Table S3). Four laboratory test variables, CRE, EO, lymphocytes (LY), MO, were differentially expressed in both EPE and LPE, while, Hepatitis B surface antigen (HBsAg) and three laboratory variables consisting of Hct, white blood cell (WBC), crystals (XTAL) were only differentially expressed in EPE and LPE respectively (Figure S9). As for the comparison of lab parameters between the EPE and LPE women, we didn't observe there were significant differences in lab markers between the two subtypes of PE samples (Table S4). On the basis of the above findings, missing values of the identified laboratory markers were replaced with medians values (Table S5) and used to build the predictive models.

Prediction of preeclampsia in early pregnancy

First, we analyzed feature importance of differentially expressed metabolites and proteins using the Boruta algorithm in the training dataset. L-Glutamine, erythronic acid, 3-Indolebutyric acid, l-Malic acid, levulinic acid, l-Alpha-aminobutyric acid, 2-Hydroxy- 3-methylbutyric acid, stearylcarnitine(C18), ornithine and palmitoylcarnitine were top ten most informative metabolites selected by the Boruta algorithm (Figure S10 A). HBD, CFD, MIF, VASN, Tenascin C (TNC), NRGN, Alpha Hemoglobin Stabilizing Protein (AHSP), Pregnancy Specific Beta- 1-Glycoprotein 4(PSG4), Coiled-Coil Domain Containing 126(CCDC126), SOD3 were top ten most important proteomic biomarkers for EPE prediction (Figure S10B). Then, we selected the top ten omics biomarkers and established separately 968 random combinations to investigate the optimal combination of omics biomarkers to predict EPE using a three-fold cross validation method (Fig. 4A). The AUC values of 968 random forest models followed a normal distribution in the internal validation set. The random forest model with metabolic predictors comprising stearylcarnitine(C18), 2-Hydroxy- 3-methylbutyric acid, levulinic acid, l-Malic acid, 3-Indolebutyric acid, l-Glutamine and ornithine presented the highest mean AUC value in the internal validation set (Fig. 4B). The seven proteins consisting of TNC, VASN, MIF, CFD, HBD, AHSP, SOD3 were the optimal combination of proteomic predictors (Fig. 4B). Secondly, we established random forest models which take different combinations of predictors including seven clinical variables (age, BMI, IVF, RPL, PMH, twin pregnancy, MAP), the most predictive omics biomarkers and laboratory test results and evaluated their performances in the test set. The model that incorporated clinical factors, metabolic and laboratory test biomarkers (herein after referred to as the EPE model) presented the highest mean AUC value (mean AUC ± SD = 0.8816 ± 0.0077, Fig. 4C and D), outperforming predictions from each separate and combined model (Fig. 4C and D). The EPE model distinguished EPE patients from controls in early pregnancy with good sensitivity (87.5%, 95% confidence

(See figure on next page.)

Fig. 4 The establishment and validation of the EPE models. **A**. The top ten most important metabolic and proteomic biomarkers were selected to build 968 predictor combinations separately, the three-fold cross validation method was utilized to develop and validate random forest model for each combination of omics predictors. **B**. QQ plot shows the AUC values of random forest models for all combinations of omics predictors. The metabolic predictors comprising stearylcarnitine(C18), 2-Hydroxy- 3-methylbutyric acid, levulinic acid, l-Malic acid, 3-Indolebutyric acid, l-Glutamine and ornithine showed the highest mean AUC value in the internal validation set, the seven proteins consisting of TNC, VASN, MIF, CFD, HBD, AHSP, SOD3 were the optimal combination of proteomic predictors for EPE prediction. **C**. Comparison of performance of machine-learning models in terms of the AUC values in the training and testsets. Clin: Clinical factors, Lab: laboratory test variables, met: metabolites, pro: proteins. **D**. ROC curves for the optimal EPE model in the training and test datasets. **E**. Spearman correlation between predictors and prediction scores obtained from the EPE model in the whole dataset. The vertical bar represents correlation coefficients, with red and blue showing high and low correlation respectively

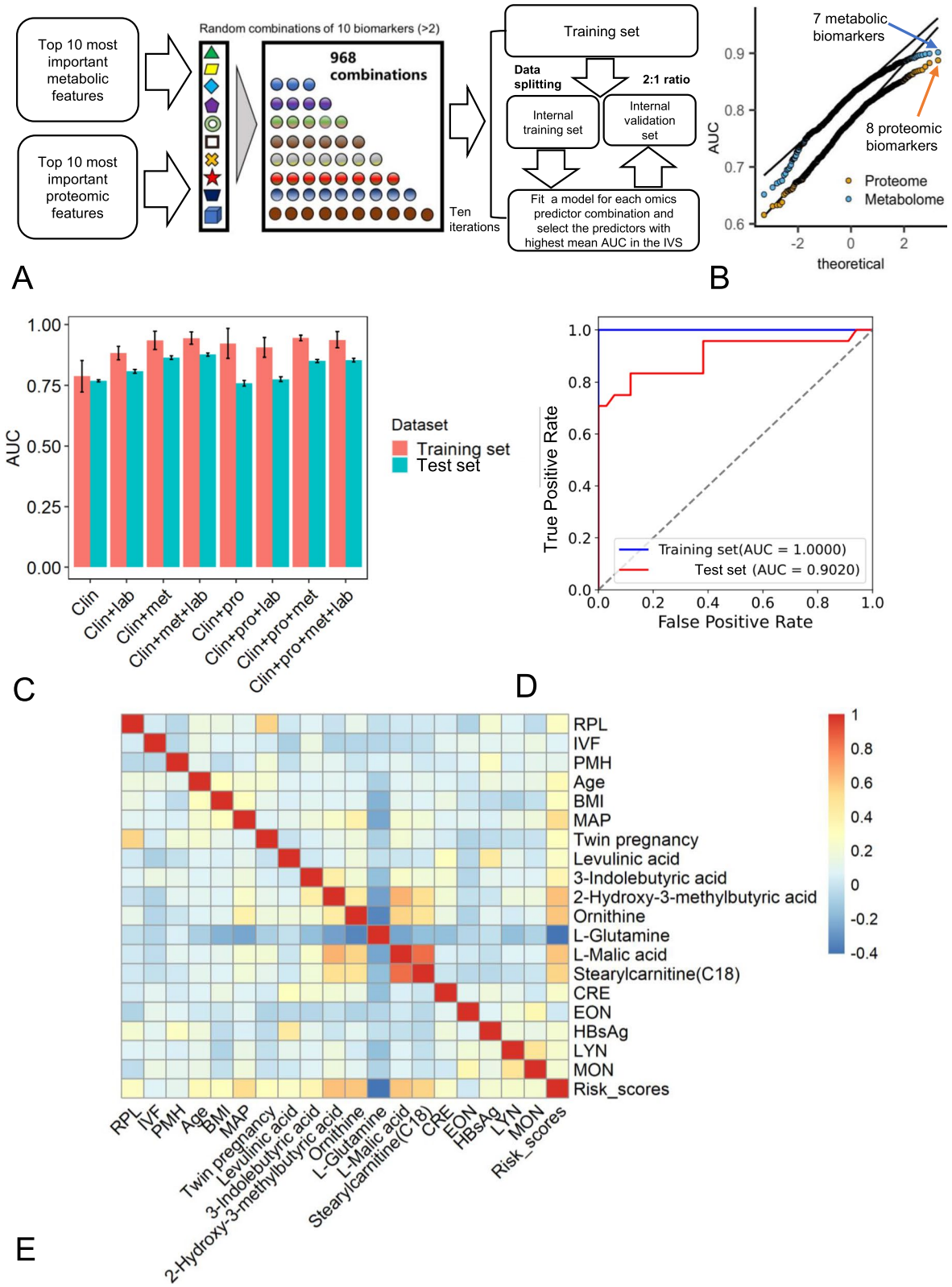

Fig. 4 (See legend on previous page.)

interval [CI]: 67.64%– 97.34%) and specificity (94.1%, 95% CI: 80.32%– 99.28%, Table 2, Table S6) in the test set. Moreover, the correlations between each predictor and prediction model scores were analyzed, the highest correlation was with 2-Hydroxy- 3-methylbutyric acid, followed by l-Malic acid, ornithine, stearylcarnitine(C18) and MAP ($r > 0.5$, $p < 0.001$ for all cases, Fig. 4E), suggesting the model captures mostly metabolite expression differences.

Following the same strategy, we identified top ten most important omics predictors associated with LPE (Figure S11). The levulinic acid, tartaric acid, l-Malic acid, 2-Hydroxy- 2-methylbutyric acid and l-Pipecolic acid were the optimal combination of predictors for LPE prediction (Figure S11 A and Figure S12 A). The model comprising eight proteomic markers, APOE, S100 Calcium Binding Protein A4 (S100 A4), Ribosomal Protein Lateral Stalk Subunit P2(RPLP2), PRG4, Neuraminidase 2(NEU2), JUP, ATPase 13 A3(ATP13 A3) and FABP5, presented the highest mean AUC value (Figure S12 A). Then, we established various LPE models using clinical factors, omics biomarkers and lab test variables, and identified the model consisting of clinical factors, the optimal metabolic and proteomic biomarkers (herein after referred to as the LPE model) performed best in the prediction of LPE in the test set (mean AUC ± SD: 0.8793 ± 0.0114, Figure S12B and C). The LPE model exhibited high accuracy in classifying LPE patients from controls in early pregnancy (sensitivity: 66.67%, 95% CI: 43.03%– 85.41%; specificity: 94.12%, 95% CI: 80.32%– 99.28%, Table S7) in the test set. The predictor APOE, exhibited highest correlation with risk score predicted by the LPE model, followed by MAP, S100 A4, l-Malic acid, PRG4 ($r > 0.40$, $p < 0.001$ for all cases, Figure S12D), indicating the model captures largely differences in multi-omics biomarker expression and maternal characteristics. Lastly, we evaluated the feature importance of the final models, as shown in Figure S13, the top 5 most important predictors were Stearylcarnitine(C18), L-Malic acid, Levulinic acid, MAP, 2-Hydroxy- 2-methylbutyric acid for the EPE model and 2-Hydroxy- 2-methylbutyric acid, MAP, Levulinic acid, Tartaric acid, RPLP2 for the LPE model. 2-Hydroxy- 2-methylbutyric acid, MAP, Levulinic acid were critical predictors for both EPE and LPE models, while, Stearylcarnitine(C18) and Tartaric acid were specifically predictive of EPE and LPE models respectively.

Table 2 The confusion matrices of binary results of the EPE model in the test set of 24 EPE and 34 healthy participants

	EPE	Control
Predicted EPE	21	2
Predicted control	3	32
Sensitivity	87.5% (95% CI, 67.64%– 97.34%)	
Specificity	94.12% (95% CI, 80.32%– 99.28%)	
Positive predictive value	91.3% (95% CI, 73.07%– 97.60%)	
Negative predictive value	91.43% (95% CI, 78.67%– 96.86%)	
Cutoff	0.34	

Discussion

In the present study, we systematically investigated the predictive values of various clinical characteristics and routine prenatal laboratory test parameters for different subtypes of PE in early pregnancy using all available clinical and laboratory data from six hospitals. We confirmed that pregnant women with higher MAP and BMI, IVF, all known PE risk factors [3], had a significantly higher risk for PE than those with lower MAP, BMI and without IVF. The incidence of PE is approximately 9% in twin pregnancies, representing a three-fold increase compared to singleton pregnancies [26], which is in line with our study. Our study also found that participants with RPL are more likely to develop EPE. Several studies have investigated the relationship between RPL and PE. These studies revealed that RPL is strongly associated with preterm PE [27, 28]. Trogstad et al. reported a significantly elevated risk of PE in cases of RPL, only when there was a history of assisted reproduction [29]. These results suggest twin pregnancies and RPL are risk factors for PE development.

In the present study, we identified sets of metabolites that exhibited significantly different concentrations in PE cases relative to normal controls. Several metabolites are informative for both EPE and LPE prediction and known metabolic biomarkers in PE, such as 2-Hydroxy- 3-methylbutyric acid [30], ornithine [31]. Some metabolic biomarkers have been uncovered for the first time in this study, including levulinic acid, l-Malic acid, 3-Indolebutyric acid, homovanillic acid. The arginine biosynthesis pathway has been identified as one of the main pathways associated with preeclampsia, as arginine is a precursor of nitric oxide, a potent endothelial-derived vasodilator that is implicated in the pathophysiology of preeclampsia [32]. Additionally, the alanine, aspartate, glutamate, and glutamine metabolic pathway was found to be another significant metabolic pathway in PE. The glutamine-cycling pathway is a major factor in the development of metabolic risk [33]. Abnormalities in glutamate metabolism indicate the involvement of liver in global metabolic regulation, due to its relatedness with aminotransferase reactions that initiate the metabolism of the majority of amino acids [34, 35]. 26 metabolites, such as palmitoylcarnitine and stearylcarnitine(C18), were only differentially expressed in EPE, 16 metabolite markers, such as l-Pipecolic acid and tartaric acid, were only differentially

expressed in LPE, demonstrating the metabolic processes involved in the pathogenesis of EPE may largely differ from those of LPE.

The proteome analysis uncovered 28 proteins significantly differentially expressed between EPE and normal controls, seven proteins identified by the Boruta model are candidate biomarkers of PE. MIF, HBA1, HBB, HBD were found to be significantly increased in both EPE and LPE as compared to controls. MIF is a proinflammatory cytokine, which plays a critical role in the regulation of the innate immune response and normal placental development processes [36]. MIF serum expression was significantly up-regulated in preeclamptic pregnancies than in control group, which is in line with previous studies [37, 38]. HBA1, HBB, HBD encode alpha 1, beta and delta subunits of hemoglobin and play a key role in oxygen carrier activity and oxygen binding [39]. Placental hypoxia is a major characteristics of preeclampsia, which may stimulate the increased expression of HBA1, HBB, HBD in the plasma from PE patients [40]. Pregnancy Specific Glycoprotein (PSG)4 and PSG9 were significantly down-regulated in EPE samples but not in LPE samples. PSG9 stimulates increase in FoxP3 + regulatory T-Cells through the TGF-β1 pathway [41] and regulates platelet-fibrinogen interactions and has antiplatelet activity [42], supporting the role of PSG9 in immune regulation. APOE was significantly up-regulated and specific biomarker for LPE. APOE is well-known for its protective role in atherosclerosis, and APOE-knockout mouse model is often used as pre-clinical atherosclerosis model and more relevant to LPE [43, 44]. The analysis of laboratory test variables identified 5 and 7 test results significantly different between EPE, LPE and healthy controls respectively. Many variables have been reported for their association with PE, such as LY, WBC, MO, CRE, HCT [11, 12, 45]. Furthermore, the addition of these laboratory variables to the EPE model did further improve model performance, suggesting these laboratory variables provide additional value in PE prediction.

Previous studies show that an integrated multi-omics model further improved prediction accuracy as compared to single omics models for PE patients [46, 47]. In this study, the PE model that integrated clinical factors, multi-omics biomarkers outperformed clinical factors-only and single omics models for PE prediction, further validating the results. Furthermore, we identified the addition of laboratory test variables to the prediction models yielded highest prediction accuracies early in pregnancy, suggesting they provide additional value to PE prediction.

Previous models to early predict preeclampsia have incorporated maternal characteristics, uterine artery Doppler measurements and specific protein biomarkers, including Placental growth factor (PlGF) and Pregnancy-associated plasma protein-A (PAPP-A) [48–50]. However, the model shows poorer performance in screening PE in Asian population than in Western pregnant women [51–54]. Cheng, et al. reported that the combined model showed detection rates of 72 and 55% for early and late PE, respectively, for a 10% false positive rate, which demonstrates poorer performance than our models [51]. PE is an extremely heterogenous disorder, making it's biologically implausible to distinguish this disorder from normal using a single biomarker or a single omics data. Our integrated models have successfully captured variations in maternal characteristics, multi-omics expression, and laboratory test variables, resulting in superior performance in predicting PE compared to models that only consider clinical factors or single omics data. The metabolic and proteomic markers identified in this study can be easily measured using LC–MS/MS technology in clinical settings. Additionally, the laboratory test variables are derived from routine prenatal tests, making them readily accessible. By incorporating metabolic and proteomic markers alongside laboratory test results, the EPE and LPE models offer novel approaches for predicting PE in early pregnancy. These models can identify pregnant women who are at a higher risk of developing PE, enabling timely intervention. It is recommended that women predicted to be at a high risk receive low-dose aspirin treatment, which may significantly reduce the incidence of PE and enhance pregnancy outcomes for both mother and fetus.

Our integrated models have demonstrated the highest accuracies in distinguishing PE patients from healthy controls, surpassing the performance of clinical factors-only and single omics models. These results further emphasize the strength of multi-omics biomarkers in predicting PE. Despite progresses, this study has several limitations. First of all, the blood samples used for proteomic and metabolomic measurement were non-fasting, which might impact the results. Given the small number of samples and hospitals and the study's focus on a specific Chinese cohort, the identified multi-omics biomarkers and laboratory test variables might not be stable and generalizable. Future studies will be needed to address the generalizability of these findings to other populations with different demographic or genetic backgrounds. Secondly, the precise role of these changes in metabolites, proteins, and laboratory test parameters in the onset and progression of preeclampsia remains incompletely understood. Further studies will be necessary to fully characterize the functional implications and molecular mechanisms underlying these changes. Thirdly, the developed models require further validation using a

larger cohort of pregnant women, we are going to address these issues in future studies.

Conclusion

In conclusion, we identified a number of potential multi-omics and laboratory test biomarkers for PE prediction. We developed EPE and LPE prediction models based on clinical characteristics, multi-omics and laboratory test variables to screening for PE in early pregnancy. These models have high sensitivity and specificity, showing the potential to further improve early diagnosis of PE and eventually guide therapeutic interventions in clinical settings.

Abbreviations

PE	Preeclampsia
EPE	Early-onset preeclampsia
LPE	Late-onset preeclampsia
CI	Confidence interval
LC–MS-MS	Liquid Chromatography with tandem mass spectrometry
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene ontology
topGo	Topology-based Gene Ontology scoring
SD	Standard deviation
BMI	Body mass index
DBP	Diastolic blood pressure
SBP	Systolic blood pressure
IVF	In vitro fertilization
PMH	Past medical history
RPL	Recurrent pregnancy loss
MAP	Mean arterial pressure
MZSA	Maximum z-score value of shadow variables
AUC	Area under the curves
ROC	The receiver operating characteristic curves
MCHC	Mean corpuscular hemoglobin concentration
FT4	Free Thyroxine
PMO	Proportion of monocytes
EO	Eosinophils count
CRE	Creatinine
URBC	Urine red blood cell
AST	Aspartate aminotransferase
LY	Lymphocytes count
ALT	Alanine Transaminase
SG	Urine specific gravity
PEO	Proportion of Eosinophils
RBC	Red blood cell count
MPV	Mean platelet volume
HBsAg	Hepatitis B surface antigen
MO	Monocytes count
HCT	Hematocrit
Hb	Hemoglobin
BA	Basophils count
UWBC	Urine white blood cell
WBC	White blood cell count
TSH	Thyroid stimulating hormone
XTAL	Crystals
BLD	Urine blood
EV	External validation
NE	Neutrophils
PLT	Platelet count
PDW	Platelet distribution width
PCT	Plateletcrit
MCH	Mean corpuscular hemoglobin
MCV	Mean corpuscular volume
T-BIL	Total Bilirubin Test
TP	Total protein
GLU	Fasting glucose
CRE	Creatinine
UA	Uric acid
UPRO	Urine Protein
TSH	Thyroid-stimulating hormone
FT4	Free thyroxine
TPOAb	Thyroid peroxidase antibody
PlGF	Placental growth factor
PAPPA	Pregnancy-associated plasma protein A
SOD3	Superoxide dismutase 3
MIF	Macrophage migration inhibitory factor
NRGN	Neurogranin
HBD	Hemoglobin Subunit Delta
VASN	Vasorin
APOE	Apolipoprotein E
JUP	Junction Plakoglobin
ANXA2	Annexin A2
FABP5	Fatty acid binding protein 5
PRG4	Proteoglycan 4
HBA1	Hemoglobin Subunit Alpha 1
HBB	Hemoglobin Subunit Beta
CFD	Complement Factor D
TNC	Tenascin C
AHSP	Alpha Hemoglobin Stabilizing Protein
PSG4	Pregnancy Specific Beta-1-Glycoprotein 4
CCDC126	Coiled-Coil Domain Containing 126
S100 A4	S100 Calcium Binding Protein A4
RPLP2	Ribosomal Protein Lateral Stalk Subunit P2
NEU2	Neuraminidase 2
ATP13 A3	ATPase 13A3

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12884-025-07582-4.

Supplementary Material 1

Acknowledgements

None.

Clinical trial

Not applicable.

Authors' contributions

Conceptualization: Xiaohong Ruan, Jie Qin, Sufen Zhang, Jianguo Zhang, Lijian Zhao, Rui Zhang, Liang Lin; Visualization, Investigation: Zhuo Diao, Suihua Feng, Guixue Hou; Methodology, Software: Qiang Zhao, Jia Li, Xiao Zhang, Wenqiu Xu, Zhiguang Zhao, Zhixu Qiu; Writing- Original draft preparation: Jia Li, Xiao Zhang; Writing- Reviewing and Editing: Jia Li, Wenzhi Yang, Peirun Tian, Si Zhou, Qun Zhang, Weiping Chen, Huahua Li, Gefei Xiao, Sufen Zhang, Liqing Hu, Zhongzhe Li, Liang Lin, Shida Zhu, Hui Jiang, Shunyao Wang, Ruyun Gao, Wuyan Huang.

Funding

The study was supported by Zhuhai Social Development Science and Technology Plan Project (project ID:2220004000295 and 2320004000179), Shenzhen Key Medical Discipline Construction Fund (SZXK028), Shenzhen Science and Technology Innovation Committee (KJYY20180703173402020), S&T Program of Shijiazhuang (No: 235790429H) and Key Research and Development Projects of Hebei Province (No:21377720D). Shenzhen Science, Technology, and Innovation Committee, Grant/Award Numbers: JCYJ20220530162412029; Science and Technology Development Special Fund of Shenzhen Longgang District, Grant/Award Numbers: LGKCYLWS2022008.

Data availability

The data that support the findings of this study have been deposited into CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb) with accession number CNP0004774.

Declarations

Ethics approval and consent to participate

Informed consent was waived by the Ethics Committees of Beijing Genomics Institute. This study was approved by the Ethics Committees of Beijing Genomics Institute (BGI-IRB 22026). All methods were performed in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Obstetrics and Gynecology, Jiangmen Central Hospital, Guangdong Province, Jiangmen 529030, China. [2]Clinical Transformation and Application Key Lab for Obstetrics and Gynecology, Pediatrics, and Reproductive Medicine of Jiangmen, Guangdong Province, Jiangmen 529030, China. [3]Clin Lab, Hebei Industrial Technology Research Institute of Genomics in Maternal & Child Health, BGI Genomics, Hebei Province, Shijiazhuang 050011, China. [4]BGI Genomics, Guangdong Province, Shenzhen 518083, China. [5]Institute of Medical Genetics, Zhuhai Center for Maternal and Child Health Care, Guangdong Province, Zhuhai 519000, China. [6]Department of Prevention and Health Care, Zhuhai Center for Maternal and Child Health Care, Guangdong Province, Zhuhai 519000, China. [7]Hunan Provincial Key Laboratory of Regional Hereditary Birth Defects Prevention and Control, Changsha Hospital for Maternal & Child Health CareAffiliated to, Hunan Normal University, Hunan Province, Changsha 410007, China. [8]School of Public Health, Hebei Province Key Laboratory of Environment and Human Health, Hebei Medical University, Hebei Province, Shijiazhuang 050017, China. [9]School of Public Health, Hebei Medical University, Shijiazhuang 050017, China. [10]Medical Technology College, Hebei Medical University, Shijiazhuang 050017, China. [11]Division of Maternal-Fetal Medicine, Shenzhen Baoan Women's and Children's Hospital, Guangdong Province, Shenzhen 518102, China.

Received: 14 January 2025 Accepted: 8 April 2025
Published online: 05 May 2025

References

1. Abalos E, Cuesta C, Grosso AL, Chou D, Say L. Global and regional estimates of preeclampsia and eclampsia: a systematic review. Eur J Obstet Gynecol Reprod Biol. 2013;170(1):1–7.
2. Poon LC, Shennan A, Hyett JA, Kapur A, Hadar E, Divakar H, et al. The International Federation of Gynecology and Obstetrics (FIGO) initiative on pre-eclampsia: A pragmatic guide for first-trimester screening and prevention. Int J Gynecol Obstet. 2019;145(S1):1–33. Available from: https://doi.org/10.1002/ijgo.12802
3. Magee LA, Nicolaides KH, von Dadelszen P. Preeclampsia. N Engl J Med. 2022;386(19):1817–32.
4. Plasencia W, Maiz N, Bonino S, Kaihura C, Nicolaides KH. Uterine artery Doppler at 11 + 0 to 13 + 6 weeks in the prediction of pre-eclampsia. Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol. 2007;30(5):742–9.
5. Skjaerven R, Vatten LJ, Wilcox AJ, Rønning T, Irgens LM, Lie RT. Recurrence of pre-eclampsia across generations: exploring fetal and maternal genetic components in a population based cohort. BMJ. 2005;331(7521):877.
6. D'Anna R, Baviera G, Corrado F, Giordano D, De Vivo A, Nicocia G, et al. Adiponectin and insulin resistance in early- and late-onset pre-eclampsia. BJOG. 2006;113(11):1264–9.
7. Duckitt K, Harrington D. Risk factors for pre-eclampsia at antenatal booking: systematic review of controlled studies. BMJ. 2005;330(7491):565.
8. Tranquilli AL, Brown MA, Zeeman GG, Dekker G, Sibai BM. The definition of severe and early-onset preeclampsia Statements from the International Society for the Study of Hypertension in Pregnancy (ISSHP). Pregnancy Hypertens. 2013;3(1):44–7.
9. Ghaemi MS, DiGiulio DB, Contrepois K, Callahan B, Ngo TTM, Lee-McMullen B, et al. Multiomics modeling of the immunome, transcriptome, microbiome, proteome and metabolome adaptations during human pregnancy. Bioinformatics. 2019;35(1):95–103.
10. Benny PA, Alakwaa FM, Schlueter RJ, Lassiter CB, Garmire LX. A review of omics approaches to study preeclampsia. Placenta. 2020;92:17–27.
11. Marić I, Tsur A, Aghaeepour N, Montanari A, Stevenson DK, Shaw GM, et al. Early prediction of preeclampsia via machine learning. Am J Obstet Gynecol MFM. 2020;2(2):1–17.
12. Li S, Wang Z, Vieira LA, Zheutlin AB, Ru B, Schadt E, et al. Improving preeclampsia risk prediction by modeling pregnancy trajectories from routinely collected electronic medical record data. npj Digit Med. 2022;5(1):1–16.
13. Chen Y, Ou W, Lin D, Lin M, Huang X, Ni S, et al. Increased Uric Acid, Gamma-Glutamyl Transpeptidase and Alkaline Phosphatase in Early-Pregnancy Associated With the Development of Gestational Hypertension and Preeclampsia. Front Cardiovasc Med. 2021;8(October):1–11.
14. Lee SM, Park JS, Han YJ, Kim W, Bang SH, Kim BJ, et al. Elevated alanine aminotransferase in early pregnancy and subsequent development of gestational diabetes and preeclampsia. J Korean Med Sci. 2020;35(26):1–10.
15. Zhang Y, Sheng C, Wang D, Chen X, Jiang Y, Dou Y, et al. High-normal liver enzyme levels in early pregnancy predispose the risk of gestational hypertension and preeclampsia: A prospective cohort study. Front Cardiovasc Med. 2022;9:963957.
16. Li Y, Hou G, Zhou H, Wang Y, Tun HM, Zhu A, et al. Multi-platform omics analysis reveals molecular signature for COVID-19 pathogenesis, prognosis and drug target discovery. Signal Transduct Target Ther. 2021;6(1). Available from: https://doi.org/10.1038/s41392-021-00508-4
17. Josić D, Horn H, Schulz P, Schwinn H, Britsch L. Size-exclusion chromatography of plasma proteins with high molecular masses. J Chromatogr A. 1998;796(2):289–98.
18. Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat Methods. 2020;17(1):41–4.
19. Xie G, Wang L, Chen T, Zhou K, Zhang Z, Li J, et al. A Metabolite Array Technology for Precision Medicine. Anal Chem. 2021;93(14):5709–17. Available from: https://doi.org/10.1021/acs.analchem.0c04686
20. Pang Z, Chong J, Li S, Xia J. Metaboanalystr 3.0: Toward an optimized workflow for global metabolomics. Metabolites. 2020;10(5):186.
21. Alexa A, Rahnenführer J. Gene set enrichment analysis with topGO. Bioconductor Improv. 2007;27. Available from: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Gene+set+enrichment+analysis+with+topGO#0%5Cnftp://mirrors.nic.funet.fi/bioconductor.org/2.7/bioc/vignettes/topGO/inst/doc/topGO.pdf
22. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. G:Profiler-a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. 2007;35(SUPPL.2):193–200.
23. Grillo A, Salvi P, Furlanis G, Baldi C, Rovina M, Salvi L, et al. Mean arterial pressure estimated by brachial pulse wave analysis and comparison with currently used algorithms. J Hypertens. 2020;38(11):2161–8.
24. Kursa MB, Rudnicki WR. Feature Selection with the Boruta Package. J Stat Softw. 2010;36(11 SE-Articles):1–13. Available from: https://www.jstatsoft.org/index.php/jss/article/view/v036i11
25. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-Learn: machine learning in python. J Mach Learn Res. 2011;12(null):2825–30.
26. Francisco C, Wright D, Benkő Z, Syngelaki A, Nicolaides KH. Hidden high rate of pre-eclampsia in twin compared with singleton pregnancy. Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol. 2017;50(1):88–92.
27. Gunnarsdottir J, Stephansson O, Cnattingius S, Åkerud H, Wikström AK. Risk of placental dysfunction disorders after prior miscarriages: A population-based study. Am J Obstet Gynecol. 2014;211(1):34.e1–34.e8. Available from: https://doi.org/10.1016/j.ajog.2014.01.041
28. Roepke ER, Christiansen OB, Källén K, Hansson SR. Women with a history of recurrent pregnancy loss are a high-risk population for adverse obstetrical outcome: A retrospective cohort study. J Clin Med. 2021;10(2):1–12.
29. Trogstad L, Magnus P, Moffett A, Stoltenberg C. The effect of recurrent miscarriage and infertility on the risk of pre-eclampsia. BJOG An Int J Obstet Gynaecol. 2009;116(1):108–13.

30. Kenny LC, Broadhurst D, Brown M, Dunn WB, Redman CWG, Kell DB, et al. Detection and identification of novel metabolomic biomarkers in preeclampsia. Reprod Sci. 2008;15(6):591–7.
31. Prameswari N, Irwinda R, Wibowo N, Saroyo YB. Maternal Amino Acid Status in Severe Preeclampsia: A Cross-Sectional Study. Nutrients. 2022;14(5).
32. Baylis C, Beinder E, Sütö T, August P. Recent insights into the roles of nitric oxide and renin-angiotensin in the pathophysiology of preeclamptic pregnancy. Semin Nephrol. 1998;18(2):208–30.
33. Cheng S, Rhee EP, Larson MG, Lewis GD, McCabe EL, Shen D, et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. Circulation. 2012;125(18):2222–31.
34. Kelly A, Stanley CA. Disorders of glutamate metabolism. Ment Retard Dev Disabil Res Rev. 2001;7(4):287–95.
35. Brosnan ME, Brosnan JT. Hepatic glutamate metabolism: a tale of 2 hepatocytes. Am J Clin Nutr. 2009;90(3):857S-861S.
36. Todros T, Paulesu L, Cardaropoli S, Rolfo A, Masturzo B, Ermini L, et al. Role of the macrophage migration inhibitory factor in the pathophysiology of pre-eclampsia. Int J Mol Sci. 2021;22(4):1–19.
37. Todros T, Bontempo S, Piccoli E, Ietta F, Romagnoli R, Biolcati M, et al. Increased levels of macrophage migration inhibitory factor (MIF) in preeclampsia. Eur J Obstet Gynecol Reprod Biol. 2005;123(2):162–6.
38. Mahmoud S, Nasri H, Nasr AM, Adam I. Maternal and umbilical cord blood level of macrophage migration inhibitory factor and insulin like growth factor in Sudanese women with preeclampsia. J Obstet Gynaecol J Inst Obstet Gynaecol. 2019;39(1):63–7.
39. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, et al. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 2004;32(Database issue):D258-61.
40. Hung TH, Charnock-Jones DS, Skepper JN, Burton GJ. Secretion of tumor necrosis factor-alpha from human placental tissues induced by hypoxia-reoxygenation causes endothelial cell activation in vitro: a potential mediator of the inflammatory response in preeclampsia. Am J Pathol. 2004;164(3):1049–61.
41. Jones K, Ballesteros A, Mentink-Kane M, Warren J, Rattila S, Malech H, et al. PSG9 Stimulates Increase in FoxP3+ Regulatory T-Cells through the TGF-β1 Pathway. PLoS ONE. 2016;11(7): e0158050.
42. Shanley DK, Kiely PA, Golla K, Allen S, Martin K, O'Riordan RT, et al. Pregnancy-specific glycoproteins bind integrin αIIbβ3 and inhibit the platelet-fibrinogen interaction. PLoS ONE. 2013;8(2): e57491.
43. Emini Veseli B, Perrotta P, De Meyer GRA, Roth L, Van der Donckt C, Martinet W, et al. Animal models of atherosclerosis. Eur J Pharmacol. 2017;816:3–13.
44. Chen H, Aneman I, Nikolic V, Karadzov Orlic N, Mikovic Z, Stefanovic M, et al. Maternal plasma proteome profiling of biomarkers and pathogenic mechanisms of early-onset and late-onset preeclampsia. Sci Rep. 2022;12(1):19099.
45. Myatt L, Clifton RG, Roberts JM, Spong CY, Hauth JC, Varner MW, et al. First-trimester prediction of preeclampsia in nulliparous women at low risk. Obstet Gynecol. 2012;119(6):1234–42.
46. Marić I, Contrepois K, Moufarrej MN, Stelzer IA, Feyaerts D, Han X, et al. Early prediction and longitudinal modeling of preeclampsia from multiomics. Patterns. 2022;3(12);100655.
47. Bahado-Singh R, Poon LC, Yilmaz A, Syngelaki A, Turkoglu O, Kumar P, et al. Integrated Proteomic and Metabolomic prediction of Term Preeclampsia. Sci Rep. 2017;7(1):1–10. Available from: https://doi.org/10.1038/s41598-017-15882-9
48. Wright D, Syngelaki A, Akolekar R, Poon LC, Nicolaides KH. Competing risks model in screening for preeclampsia by maternal characteristics and medical history. Am J Obstet Gynecol. 2015;213(1):62.e1-62.e10.
49. Odibo AO, Zhong Y, Goetzinger KR, Odibo L, Bick JL, Bower CR, et al. First-trimester placental protein 13, PAPP-A, uterine artery Doppler and maternal characteristics in the prediction of pre-eclampsia. Placenta. 2011;32(8):598–602.
50. Wright D, Wright A, Nicolaides KH. The competing risk approach for prediction of preeclampsia. Am J Obstet Gynecol. 2020Jul;223(1):12-23.e7.
51. Cheng Y, Leung TY, Law LW, Ting YH, Law KM, Sahota DS. First trimester screening for pre-eclampsia in Chinese pregnancies: case-control study. BJOG. 2018;125(4):442–9.
52. Hu J, Gao J, Liu J, Meng H, Hao N, Song Y, et al. Prospective evaluation of first-trimester screening strategy for preterm pre-eclampsia and its clinical applicability in China. Ultrasound Obs Gynecol. 2021;58(4):529–39.
53. Chaemsaithong P, Pooh RK, Zheng M, Ma R, Chaiyasit N, Tokunaka M, et al. Prospective evaluation of screening performance of first-trimester prediction models for preterm preeclampsia in an Asian population. Am J Obstet Gynecol. 2019;221(6):650.e1-650.e16.
54. Chaemsaithong P, Sahota DS, Poon LC. First trimester preeclampsia screening and prediction. Am J Obstet Gynecol. 2022;226(2):S1071-S1097.e2. Available from: https://doi.org/10.1016/j.ajog.2020.07.020

Publisher's Note