


RESEARCH

Open Access



Quantitative urine proteomics in pregnant women for the identification of predictive biomarkers for preeclampsia

Sakari Joenväärä^{1†}, Matilda Holm^{1*†} , Mayank Saraswat^{1,2}, Rahul Agarwal³, Tiialotta Tohmola¹, Eero Kajantie^{4,5,6}, Katri Räikkönen⁷, Hannele Laivuori^{8,9,10,11}, Pia M. Villa^{8,12,13}, Esa Hämäläinen^{14,15} and Risto Renkonen¹

Abstract

Background: Preeclampsia (PE) is a life-threatening disease characterized by elevated blood pressure and proteinuria. Predictive biomarkers of PE are needed, especially those predicting PE in early pregnancy. The aim of this pilot study was to identify urine proteins that could be candidates for new non-invasive markers for PE.

Methods: Urine samples at three time points of pregnancy (12–14, 18–20 and 26–28 weeks of gestation) were prospectively collected from high-risk women who subsequently developed PE ($n = 7$), high-risk women who did not develop PE ($n = 6$), and women without known risk factors for PE ($n = 4$). The samples were analyzed using mass spectrometry and we subsequently quantified 361 proteins used for further analysis. Rigorous statistical analysis with multiple methods was performed to identify biomarker candidates.

Results: Of the clinical risk factors analyzed, pre-pregnancy body mass index (BMIBP) was found to be the most important predictor of PE. We identified multiple proteins that correlated with BMIBP and could improve the prediction of PE in combination with BMIBP. Other statistical analyses identified six proteins that each could differentiate women who subsequently developed PE from those who did not at all three time points.

Conclusions: We identified multiple urine proteins that could be used to predict PE in combination with BMIBP. We also identified six proteins that are strong candidates for predicting PE already in early pregnancy.

Keywords: Preeclampsia, Prediction, Biomarkers, Proteomics, Mass spectrometry, Urine

Background

Preeclampsia (PE) is a potentially life-threatening, multi-organ disease characterized by elevated blood pressure (> 140 mmHg systolic or > 90 mmHg diastolic) and proteinuria (≥ 0.3 g/day) after 20 weeks of gestation. PE affects 3–5% of all pregnancies and is a major cause of maternal mortality and morbidity. PE also leads to an increased risk of neonatal death and morbidity [1,

2]. Risk factors for PE include age, conditions such as chronic hypertension and diabetes, obesity, and a prior history or family history of PE [3]. Studies have shown that treatment with low-dose aspirin, started before 16 weeks of gestation, reduces the risk and incidence of PE [4–6]. Currently available biomarkers, such as serum angiogenic and antiangiogenic vasoactive agents including placental growth factor (PlGF) and soluble fms-like tyrosine kinase 1 (sFlt-1) are not sensitive enough for the early prediction of PE [7]. Clinical risk factors, including chronic hypertension, diabetes, obesity, primiparity and a previous history of PE can be used in combination with known biomarkers to identify women at increased risk of

*Correspondence: matilda.holm@helsinki.fi

[†]Sakari Joenväärä and Matilda Holm contributed equally to this work.

¹ Transplantation laboratory, Faculty of Medicine, University of Helsinki

and Helsinki University Hospital, Helsinki, Finland

Full list of author information is available at the end of the article



developing PE [8]. However, there is still a need for new predictive biomarkers for PE.

The pathological changes of PE are initiated during early gestation, while symptoms manifest later [9]. Tools to predict PE before the clinical manifestation of the disease are needed. To be able to influence the pathological process, at-risk women should be identified in early pregnancy. As PE is usually characterized by proteinuria and urine contains a variety of proteins including plasma proteins that have passed through the glomerular filtration barrier, urine is ideal for proteomic analyses and can be collected noninvasively.

Previous urine proteomics studies have identified several proteins that could be of value for the diagnosis of PE and for distinguishing PE from other hypertensive disorders [10]. A study using SELDI-TOF mass spectrometry to analyze urine samples identified a proteomic signature characterized by fragments of SERPINA1 and albumin that could predict PE and distinguish it from other disorders [11]. A study using micro-time-of-flight mass spectrometry identified a model of 50 biomarkers that was associated with subsequent PE, although this model did not reliably predict PE at time points earlier than 28 weeks of gestation [12]. Another study identified 35 peptides that could distinguish severe or mild PE from controls [13]. These authors also identified a panel consisting of 22 peptides that could differentiate PE from chronic and gestational hypertension, as well as normotensive controls [14]. However, there is still a lack of studies that have identified individual proteins that could be able to reliably predict PE in early gestation.

In this study, we have analyzed the urine proteome of prospectively collected serial urine samples from seventeen pregnant women. We included women with clinical risk factors for PE, some of whom later developed PE, and women without known risk factors. The aim of this study was to identify candidates for new non-invasive biomarkers to predict PE in early gestation. Here, we present multiple urine proteins that are strong candidates for new predictive markers for PE.

Methods

Patient cohort and study design

The present nested case-control study is a part of a multidisciplinary "Prediction and prevention of pre-eclampsia" (PREDO) study [15, 16] approved by the Ethics committee of Obstetrics and Gynecology, Hospital District of Helsinki and Uusimaa (Dnro 3/E8/05). The study was carried out between September 2005 and December 2009 in ten participating hospital maternity clinics in Finland. The project has been previously described in detail [16]. Briefly, 1082 women were recruited, 972 with clinical risk factors for PE and 110 women without known risk

factors as a control group. The participants were enrolled in this study at 12–14 weeks of gestation on the basis of existing clinical risk factors for PE (with the exception of those women without clinical risk factors who were enrolled as a control group) and their pregnancies were followed normally. Randomly selected women gave fasting blood samples and overnight urine samples at three time points (12–14, 18–20 and 26–28 weeks of gestation) during pregnancy. Written informed consent was obtained from all participants before joining the study. Pregnancy data were collected from the medical records of maternity clinics and hospitals. Pregnancy outcomes were ascertained by a jury of two medical doctors and a midwife who met face-to-face and agreed upon the diagnosis of each participant. PE was diagnosed as new-onset hypertension after 20 weeks of gestation with proteinuria >0.3 g per day. This study was carried out in accordance with the Declaration of Helsinki.

In this study, we included seven women with clinical risk factors for PE who eventually developed PE before 34+0 weeks of gestation, defined as early-onset PE (PE group). However, none of these seven women had clinical disease at any of the time points when samples were obtained. Six women with clinical risk factors who did not develop PE (risk controls, RC) and four women without risk factors for PE (healthy controls, HC) were randomly selected as controls. The clinical risk factors present in the women are given in Table 1. In this sub-study, we chose to include women from the larger PREDO cohort from whom urine samples had been obtained at all three time points during pregnancy. A total of 51 urine samples were obtained from these seventeen women. Enrolment in the study did not impede PE diagnosis in these women.

The participants in this study were Caucasian of Finnish origin. Family history was asked about in a questionnaire but not obtained for all participants. For those women from whom information was obtained (PE group: 2 women, RC group: 4 women, HC group: 4 women), none were born from a pre-eclamptic pregnancy. The age of the women ranged from 23 to 42 years, with the median being 34 years. The median weight was 75 kg and the median pre-pregnancy body mass index (BMIBP) was 25.9 kg/m², increasing to 32.0 kg/m² at the end of the pregnancy. Detailed clinical parameters and measurements are given in Additional file 2.

Sample collection

Overnight urine samples were collected without preservative from each subject during three time points of pregnancy; at 12⁺⁰ to 14⁺⁰, 18⁺⁰ to 20⁺⁰, and 26⁺⁰ to 28⁺⁰ weeks and days of gestation. The overnight morning urine, which was collected in commercial plastic

Table 1 The clinical risk factors for PE the women in this study had when enrolled in the prospective PREDO cohort

Inclusion criteria (clinical risk factors for PE) at enrolment in the PREDO cohort (at 12–14 weeks of gestation)	At-risk women who developed PE (PE group, <i>n</i> = 7)	At-risk women who did not develop PE (RC group, <i>n</i> = 6)
PE during a previous pregnancy	2	3
Small for gestational age newborn in a previous pregnancy	2	2
Chronic hypertension	0	1
Gestational diabetes in a previous pregnancy	0	0
BMI over 30	2	2
Age under 20	0	0
Age over 40	1	0
Other: systemic lupus erythematosus, Sjögren's syndrome, type 1 diabetes mellitus, fetal demise in a previous pregnancy	0	0

containers and kept refrigerated, was carefully mixed and at least two portions of 10 ml of urine from each collection were stored in deep-freeze tubes at -80°C until analysis.

Sample preparation

Urine samples were thawed and 5 mL of urine was reduced (10 mM DTT) and alkylated (40 mM iodoacetamide) for 1 h each inside the Amicon Filter (3 kDa MWCO, Millipore). Between consecutive steps, the previous solution was removed by centrifugation. To deplete albumin, 400 μL of CaptureSelect anti-Human serum albumin resin was added to the filter and incubated for 10 min with rotation to mix the sample. Resin was washed with 5 mL MQ water two times and resuspended in 100 μL MQ water before being transferred to centrifuge tubes. After centrifuging at low speed, the supernatant was collected. Protein concentration was determined using the BCA assay and equal amounts of protein were digested with sequencing grade trypsin (1:50, Promega) overnight at 37°C . The next day, the total peptides were cleaned with C18 spin columns (ThermoFisher) and dried with a vacuum dryer (Savant, ThermoFisher). The cleaned peptides were dissolved in 0.1% formic acid at a concentration of 1.4 μg peptide in 4 μL (1 injection) and analyzed by ultra performance liquid chromatography-ultra definition mass spectrometry (UPLC-UDMS^E).

UPLC-UDMS^E

TRIZAIC nanoTile 85 $\mu\text{m} \times 100 \text{ mm}$ HSS-T3u wTRAP was used as a separation device. Buffers were made in the following way: Buffer A: 0.1% formic acid in water and Buffer B: 0.1% formic acid in acetonitrile. Samples were loaded, trapped and washed (8 $\mu\text{L}/\text{min}$ of 1% buffer B) before the start of the analytical gradient. The analytical gradient used was as follows: 0–1 min 1% B, at 2 min 5% B, at 65 min 30% B, at 78 min 50% B, at 80 min 85% B, at

83 min 85% B, at 84 min 1% B, and at 90 min 1% B with 450 nL/min. Buffer A was 0.1% formic acid in water while buffer B was 0.1% formic acid in acetonitrile.

Data-independent acquisition was used with UDMS^E mode in Synapt G2-S HDMS (Waters Corporation, MA, USA). Calibration was performed with Glu1-Fibrinopeptide B MS2 fragments and the precursor ion was used as a lock mass. Data was collected in the range of 50–2000 m/z with 1 s scan time. Ion-mobility spectroscopy (IMS) wave velocity was 650 m/s. All the samples were run as triplicates before being analyzed by Progenesis QI for Proteomics (Nonlinear Dynamics, Newcastle, UK). The coefficient of variation (%CV) for the dataset was 9.83%.

Data analysis

Data analysis was performed as previously described [17]. Briefly, after importing the raw files to Progenesis QI for Proteomics, lock mass correction was applied using doubly charged Glu1-Fibrinopeptide B (785.8426 m/z). Automatic run alignment and peak picking was performed with the default parameters of the software. Label-free quantitation was performed according to Silva et al. [18]. Protein Lynx Global Server facilitated peptide identification. Every injection of peptides was spiked with 50 fmole of Hi3 peptides mix (Waters), which enabled normalization and quantitation. Uniprot human FASTA sequences (UniprotKB Release 2015_09, 20,205 sequence entries) appended with a ClpB protein sequence (Hi3 mix, CLPB_ECOLI (P63285) was used as a database for searching. Fixed modification at cysteine (carbamidomethyl) and variable at methionine (oxidation) and trypsin as a cleaving agent (two miscleavages allowed) was specified. Default peptide and fragment error tolerances were used and the false discovery rate (FDR) was set to maximally 1%, after which 361 proteins with two or more unique peptides were identified. The software stops searching

the database when the FDR reaches the specified 1%. The original number of proteins identified was 642. Progenesis QI for Proteomics applies the parsimony principle for protein grouping.

Regression modeling and correlation analysis

Feature selection in terms of which clinical covariates were important for modeling and prediction was done using two methods: relative importance and stepwise regression. For regression modeling, no PE vs. PE was selected as the response for feature selection. Continuous variables such as age, height, weight at the beginning of pregnancy, BMIBP and BMI at the end of pregnancy were used as features for the regression modeling. The R package *relaimpo* was used for both methods. For the correlation analysis, t-tests were used to identify proteins with significantly different ($p < 0.05$) abundances between the groups. The significantly different proteins as identified by t-test were then correlated to BMIBP by calculating the Pearson correlation coefficient. In total, three sets of correlations were performed. In the first set, the significantly different proteins as identified by t-test from the PE vs. HC comparison were selected and their abundances in HC samples were correlated to BMIBP. In the second set, the significantly different proteins from the PE vs. RC comparison were selected and their abundances in RC samples were correlated to BMIBP. In the third set, proteins that displayed significantly different abundances in either comparison (PE vs. HC or PE vs. RC) were chosen and their abundances in PE samples were correlated to BMIBP. Only samples from the 12–14 weeks group were used for this analysis. A Pearson correlation coefficient of 0.65 or above for a positive correlation and a coefficient of -0.65 or below for a negative correlation with BMIBP was considered significant. Further details can be found in Supplementary Methods (found in Additional file 1).

Further analysis for the selection of candidate biomarkers

Orthogonal projections to latent structures-discriminant analysis (OPLS-DA) was performed with EZInfo 3.0 (Umetrics, Sweden) and variable influence on projection (VIP) values for each comparison were calculated. The R Package *ropls* was used for calculating VIP values in partial least squares-discriminant analysis (PLS-DA) modeling for the first time point (12–14 weeks) comparisons. Nearest shrunken centroid (NSC) was used to find important proteins for the first time point comparison using the R package *pamr*. Training was done on data from each of the three sampling time points using the “*pamr.train*” function from *pamr*. The importance of each protein was obtained using 10-fold cross-validation. OPLS-DA modeling was performed and VIP plots

produced for every pairwise comparison. Three pairwise comparisons were done for every sampling time point (12–14, 18–20, and 26–28 weeks of gestation), giving a total of nine comparisons. The three comparisons for every time point were PE vs. HC, PE vs. RC, and RC vs. HC. In the coefficients vs. VIP plot, proteins with VIP values of 1 or more were considered to be significantly different between the groups. Receiver operating characteristic (ROC) curve analysis was performed using MetaboAnalyst 4.0 (<https://www.metaboanalyst.ca>). The area under the curve (AUC) was calculated for individual proteins using the univariate option in MetaboAnalyst 4.0. For the combination of the six proteins subsequently chosen, the multivariate option was used. Column scatter plots were drawn using the R packages *ggplot2* and *ggpubr*. Further details are given in Supplementary Methods (found in Additional file 1).

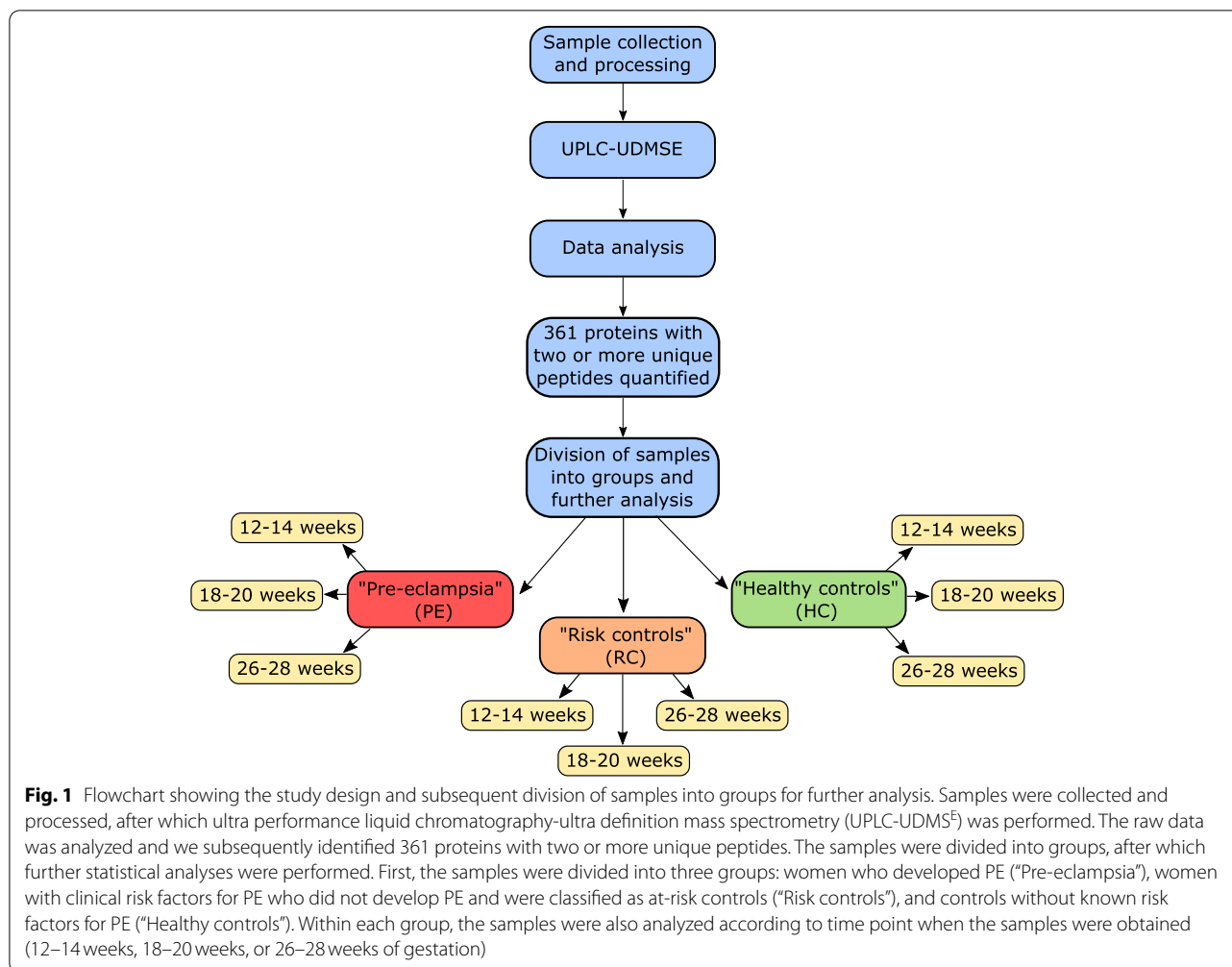
Results

Study design and protein identification

The samples in this study were divided into three groups: women who developed PE (PE, seven women, 21 samples), women with clinical risk factors for PE who did not develop the disease and were therefore classified as at-risk controls (RC, six women, 18 samples), and controls without known risk factors for PE (so-called healthy controls, HC, four women, 12 samples). At the time of sampling, none of the subjects had PE. We quantified a total of 361 proteins with two or more unique peptides (found in Additional file 2). The samples within each group were also divided according to sampling time points (12–14 weeks, 18–20 weeks, and 26–28 weeks of gestation), leading to a total of nine different comparisons. The study design and division of samples into groups for further analysis are presented in Fig. 1.

Regression modeling and correlation

Regression modeling was performed and we subsequently identified BMIBP as the clinical parameter most important for the prediction of the endpoint outcome (incidence of PE). The final Akaike information criterion of the modeling was -22.1 and the misclassification error was 0.35, indicating that 65% of all PE cases could be successfully predicted with our model using BMIBP. BMIBP was then correlated to the expression of those proteins with t-test p -values of less than 0.05 (found in Additional file 2). The results of this analysis are given in Supplementary Table 3C in Additional file 2. As can be seen, a total of 37 proteins displayed a negative correlation with BMIBP and 13 proteins displayed a positive correlation with BMIBP in the first set of correlations (see the section “Regression modeling and correlation analysis” in Methods for details). These proteins



could therefore be used in combination with BMIBP to predict PE in the HC group. In the second set of correlations, between PE and RC, four proteins were identified that negatively correlated with BMIBP and a total of 67 proteins were identified that positively correlated with BMIBP. These proteins could be used in combination with BMIBP to predict PE in the RC group. In the third set of correlations, six proteins were found to negatively correlate with BMIBP and 20 proteins were found to positively correlate with BMIBP. These proteins could be used together with BMIBP to predict PE in both the HC and RC group. The top three proteins that positively correlated with BMIBP in the third set of correlations were keratin, type I cytoskeletal 24 (KRT24), matrix extracellular phosphoglycoprotein (MEPE), and CD320 antigen (CD320).

Identification of candidate biomarkers

Three different methods were used for finding candidates for biomarkers that could classify the groups PE vs. HC

and RC. OPLS-DA, coefficient vs. VIP plots, PLS-DA, and NSC were used for the selection of appropriate biomarker candidates. The coefficients vs. VIP plots comparing PE vs. HC and PE vs. RC when only the 12–14 weeks of gestation group was considered are shown in Fig. 2 (for VIP values for the 12–14 weeks of gestation group, see S4 Table). The coefficients vs. VIP plot comparing HC vs. RC for the 12–14 weeks of gestation group can be found in Additional file 1.

Proteins able to classify the groups HC vs. RC, PE vs. HC, and PE vs. RC were compared to each other via Venn diagram. As seen in the Venn diagrams in Fig. 3, the proteins (VIP > 1) in common between the comparisons PE vs. HC and PE vs. RC can differentiate those women who develop PE from HC and RC already in early pregnancy. These 35 proteins are presented as part of S4 Table. The comparisons when only the 18–20 weeks of gestation group was considered identified 18 proteins (part of S5 Table), while the comparisons when only the 26–28 weeks of gestation group was considered identified

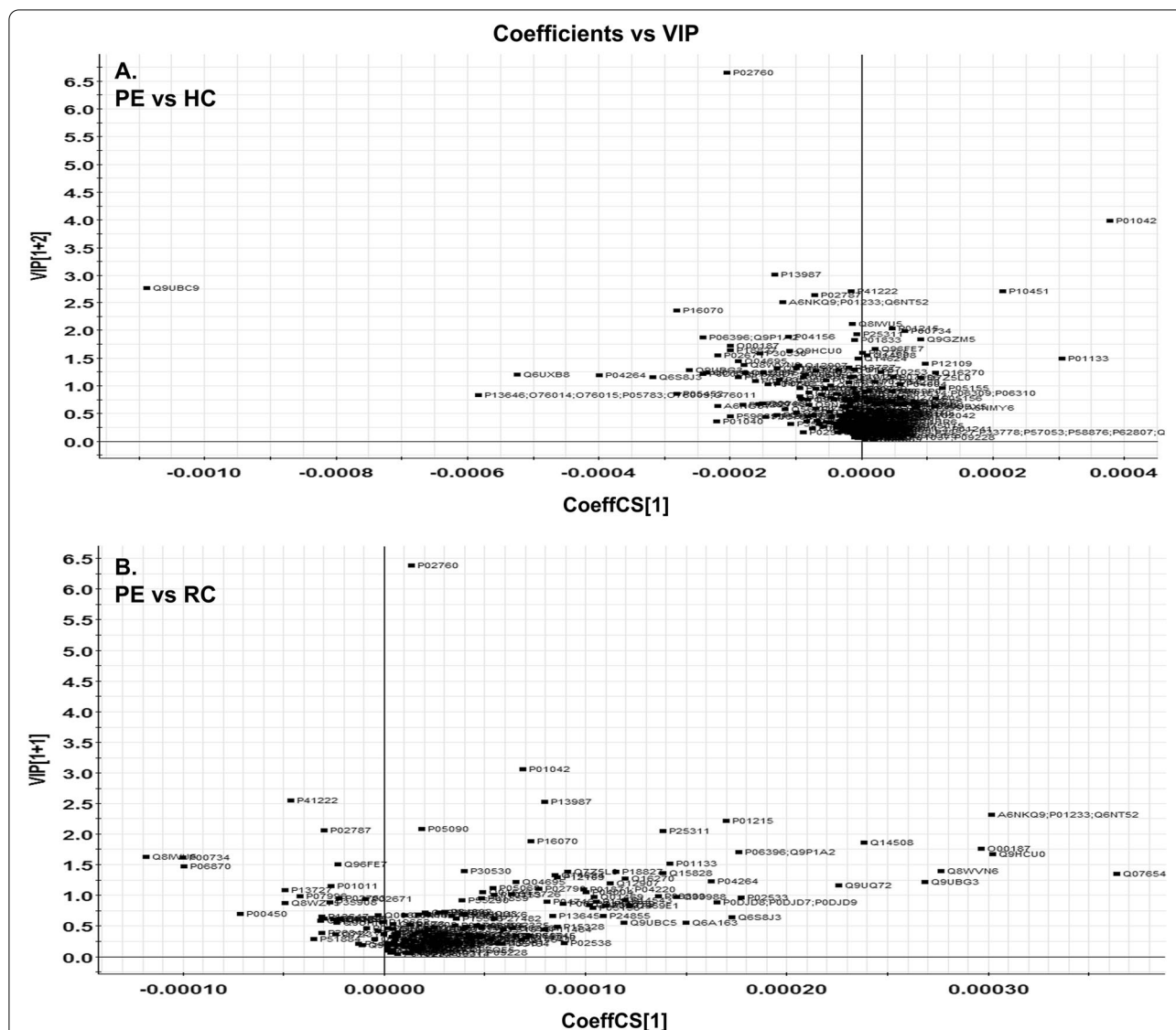


Fig. 2 The coefficients vs. VIP plots obtained from OPLS-DA modeling. A) PE vs. HC. B) PE vs. RC. OPLS-DA modeling was performed with EZInfo 3.0 software. The x-axis shows the coefficients and the y-axis the VIP. Proteins passing the cutoff of VIP values ≥ 1 were considered significantly different. VIP = variable influence on projections, OPLS-DA = orthogonal projections to latent structures-discriminant analysis, PE = preeclampsia, HC = healthy controls, RC = risk controls

14 proteins (part of S6 Table) that were able to classify the different groups. Six proteins, given in Table 2, were common to all comparisons and were therefore able to classify PE vs. HC and PE vs. RC at all three time points.

The VIP values obtained from PLS-DA of all significant proteins in the first comparison are given in S7 Table. In NSC-based variable selection, the NSC model had ~10% classification error calculated by 10-fold cross-validation in the PE vs. HC comparison and ~15% classification error in the PE vs. RC comparison. The proteins glyco-protein hormone alpha chain and prothrombin were found in the PE vs. HC comparison, while the other four

were found in the PE vs. RC comparison. The complete list of variables provided by NSC is given in S8 Table.

The normalized abundances of the six proteins identified that could classify PE vs. HC and PE vs. RC at all three time points (12–14, 18–20, and 26–28 weeks of gestation) are presented in column scatter plots in Fig. 4. The abundances of these six proteins are significantly higher in women who developed PE compared to women with and without risk factors for PE that did not develop the disease.

ROC curve analysis was performed for these six proteins by taking their normalized abundance values

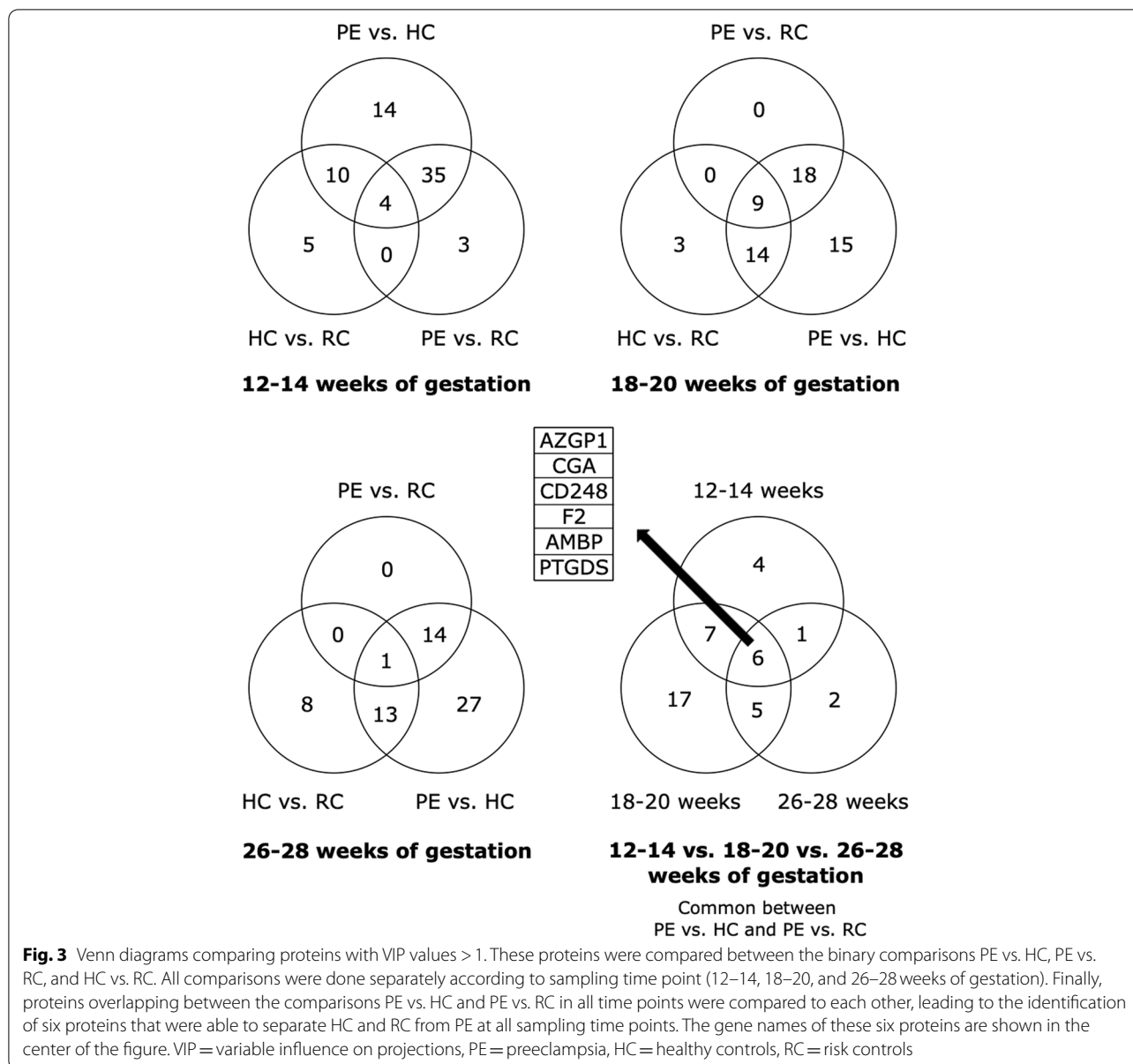


Table 2 The six proteins identified that could classify PE vs. HC and PE vs. RC at all time points

Uniprot accession	All peptides	Unique peptides	Confidence score	Protein name	Gene name	CoeffCS[1]	VIP[1]
P02760	174	151	584.24	Protein AMBP	AMBP	-0.00020422	6.65
P41222	51	45	229.30	Prostaglandin-H2 D-isomerase	PTGDS	-1.51E-05	2.70
P01215	14	11	54.34	Glycoprotein hormones alpha chain	CGA	4.81E-05	2.03
P00734	35	24	317.42	Prothrombin	F2	6.68E-05	1.98
P25311	67	53	425.96	Zinc-alpha-2-glycoprotein	AZGP1	-5.64E-06	1.92
Q9HCU0	25	17	186.26	Endosialin	CD248	-0.00010824	1.62

The VIP values for the comparison PE vs. HC are given in the last column. PE Preeclampsia, RC Risk controls, HC Healthy controls

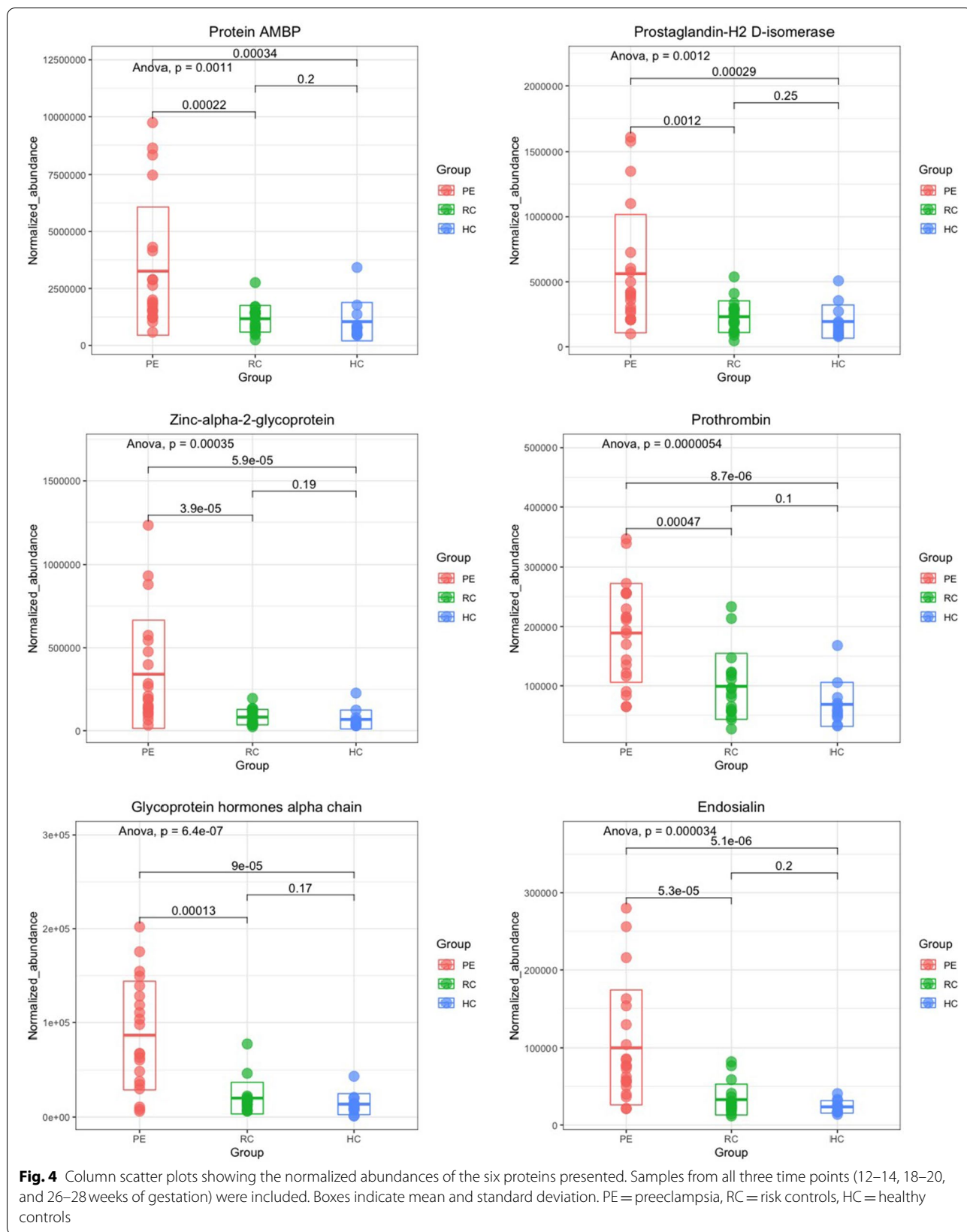


Fig. 4 Column scatter plots showing the normalized abundances of the six proteins presented. Samples from all three time points (12–14, 18–20, and 26–28 weeks of gestation) were included. Boxes indicate mean and standard deviation. PE = preeclampsia, RC = risk controls, HC = healthy controls

separately from the PE vs. RC and PE vs. HC comparisons from the first group (12–14 weeks). The area under the curve (AUC) values for this group are given in Table 3 for both comparisons. We also tested if it was possible to make a panel comprised of these six proteins that could classify PE vs. HC and RC. We calculated the combined AUC values of various combinations of these proteins and the highest AUC value that could be reached with a combination of all six proteins was 0.81 for distinguishing between PE vs. HC (see Additional file 1). For PE vs. RC, the highest AUC value for a combination of these six proteins was 0.847 (see Additional file 1). The AUC values for individual proteins were therefore higher, indicating that proteins including endosialin and protein AMBP alone are able to distinguish between PE vs. HC and RC, respectively.

Discussion

In this pilot study, we used UPLC-UDMS^E to comprehensively analyze the proteome of prospectively collected urine samples from pregnant women divided into three main groups: PE, RC, and HC. The samples were also divided according to three sampling time points (12–14, 18–20, and 26–28 weeks of gestation), enabling comparisons according to both risk of PE and gestational weeks. We quantified a total of 361 proteins with two or more unique peptides and used multiple methods of statistical analysis to identify candidates for early predictive markers of PE.

Using logistic regression modeling, we were able to identify BMIBP as the strongest predictor of PE among the clinical parameters analyzed. A high body mass index (BMI) is a risk factor for PE [19], although fairly weak as a stand-alone predictor [20]. Here, we identified multiple urine proteins whose abundances significantly correlated with BMIBP that are candidates for new predictive markers of PE. In the third set of correlations in this study, proteins that displayed significantly different abundances in either of the first two comparisons (PE vs. HC or PE vs. RC) were chosen and their abundances in PE samples correlated to BMIBP. The proteins that were subsequently found to have a significant positive correlation with

BMIBP, including KRT24, MEPE, and CD320, are likely of particular value as predictive biomarkers, as they may be able to predict PE in both healthy women and women at risk of PE. While these proteins may not be able to classify PE vs. RC and/or HC on their own, they may be of value as predictive markers when used in combination with BMIBP to predict PE. This is something that would be of interest to investigate in future studies. The correlation of urine proteins identified by proteomic analysis with BMIBP for the prediction of PE has, to the best of our knowledge, not been performed before. Only samples from 12 to 14 weeks of gestation were used for the correlation analysis, as proteins whose levels change later during pregnancy may not be as valuable for predicting PE at an early stage. The urine proteins identified here are worth investigating further, as they could be used in combination with BMIBP to predict PE at an early stage.

Using OPLS-DA, PLS-DA, and NSC modeling, we subsequently identified six urine proteins (Table 1) that are strong candidates for new predictive biomarkers of PE. These six proteins were found to have significantly different levels between the PE group and the HC and RC groups using all three methods. These proteins were able to differentiate between women who subsequently develop PE from those who do not at all three time points. The AUC values obtained from ROC analysis ranged between 0.833 and 0.964, indicating that these six proteins are excellent classifiers for distinguishing between PE vs. HC and RC already at 12–14 weeks of gestation. Endosialin displayed an AUC value of 0.964 for distinguishing between PE and HC, while protein AMBP displayed an AUC value of 0.929 for distinguishing between PE and RC, indicating their potential as predictive markers of PE. These six proteins were able to classify PE vs. HC and PE vs. RC already at 12–14 weeks of gestation (as well as at 18–20 weeks and 26–28 weeks of gestation) and are therefore of great value for predicting PE during early pregnancy (as well as during mid- and late pregnancy). As can be seen in Fig. 4, the normalized abundances of these six proteins are significantly higher in women who developed PE compared to women with and without risk factors for PE that did not develop the

Table 3 The AUC values for the six proteins that were able to classify PE vs. HC and RC

Uniprot accession	Protein name	Gene name	AUC (PE vs. HC)	AUC (PE vs. RC)
Q9HCU0	Endosialin	CD248	0.964	0.881
P00734	Prothrombin	F2	0.929	0.857
P25311	Zinc-alpha-2-glycoprotein	AZGP1	0.893	0.881
P02760	Protein AMBP	AMBP	0.893	0.929
P41222	Prostaglandin-H2 D-isomerase	PTGDS	0.893	0.905
P01215	Glycoprotein hormones alpha chain	CGA	0.857	0.833

AUC Area under the curve, PE Preeclampsia, RC Risk controls, HC Healthy controls

disease when samples from all time points were analyzed. This indicates their importance as new predictive markers for PE, although further validation of these proteins is still necessary.

One of the six proteins identified here, namely protein AMBP, has been previously identified as a potential biomarker of PE [21–23]. Prothrombin, another of the six proteins identified, is involved in the clotting process. Coagulation parameters have been shown to be altered in women who develop PE [24], indicating a possible role for prothrombin in PE and further suggesting that it could be of value as a predictive marker of PE. The identification of proteins previously linked to PE serves as a form of validation for our pilot study and strengthens our findings.

A recent proteomic study compared urine samples from women with PE and controls and subsequently identified several proteins that were differentially expressed between the groups. However, this study did not determine if changes in the levels of these proteins could be seen early during pregnancy [25]. Another recent study analyzed serum samples from women at 10–12 weeks of gestation, of which half later developed PE. The authors identified nine serum proteins with significantly different levels between the groups that could be of value as markers for early PE [26]. However, serum samples are more invasive to obtain than urine samples. The findings of our study are of significant value as we have identified six urine proteins that are independently able to classify women who developed early-onset PE from both at-risk women and controls without risk factors for PE during early, mid- and late pregnancy.

The inclusion of women who developed early-onset PE in this sub-study was due to the importance of identifying predictive markers for early-onset disease. In this study, six of the seven women in the PE group eventually developed severe PE. While this may indicate that severe PE is overrepresented here, it also shows that for most women who develop early-onset PE, the disease is severe. Further, there are several criteria for the diagnosis of severe PE, including oliguria, seizures, and persistent severe central nervous system symptoms. The presence of any one of these will suffice for the diagnosis of severe PE [27]. As PE is a multi-organ progressive disorder, many patients will therefore eventually meet the criteria for severe PE. While this study was limited by the small sample size, it was strengthened by the multiple statistical methods used for the identification of candidate biomarkers, which increases the confidence of the findings. It was also strengthened by the use of non-invasive urine samples, which are easy to obtain. Due to the inclusion of women who subsequently developed early-onset PE

in this study, it is possible that our findings may not be generalized to late-onset PE as well. However, it is important to find ways to predict early-onset PE, as the consequences are often severe.

Conclusions

In this study, we have used mass spectrometry and rigorous statistical analyses to analyze the urine proteome of pregnant women at varying risk of PE. We have identified multiple proteins that could be of value as predictive markers of PE in combination with BMIBP, as well as six proteins that were identified using OPLS-DA, PLS-DA, and NSC modeling that are strong candidates for predictive markers for PE. While validation of these candidates is still needed, our pilot study indicates that comparative urine proteomics is valuable for the identification of new, non-invasive markers for the early prediction of PE.

Abbreviations

AUC: Area under the curve; BMI: Body mass index; BMIBP: Pre-pregnancy BMI; HC: Healthy controls; NSC: Nearest shrunken centroid; OPLS-DA: Orthogonal projections to latent structures-discriminant analysis; PE: Preeclampsia; PLS-DA: Partial least squares-discriminant analysis; PREDO: Prediction and prevention of pre-eclampsia; RC: Risk controls; ROC: Receiver operating characteristic; UPLC: Ultra performance liquid chromatography; UDMSE: Ultra definition mass spectrometry; VIP: Variable influence on projection.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41231-022-00108-6>.

Additional file 1: Supplementary Figures 1–3. Supplementary methods and figures. Detailed descriptions are given in the file.

Additional file 2: Supplementary Tables 1–8. Patient data, the identified and quantified proteins, as well as proteins selected by various statistical techniques are presented. Detailed descriptions are given in the file.

Acknowledgements

We would like to thank the participants of the PREDO project as well as doctors and assisting personnel for their efforts in the project.

Authors' contributions

HL, PMV, EH, and RR conceived and designed the study. HL, PMV, EH, and EK collected the samples. HL, PMV, EH, and EK curated the clinical data. MS, SJ, and TT acquired the data. MS, SJ, MH, and RA analyzed and interpreted the data. MS, SJ, and MH drafted and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the University of Helsinki (HL, PV), Finska Läkarsällskapet (HL), EVO research funding (a special Finnish state subsidy for health science research [HL, EH]), the Academy of Finland (EK, KR, HL), the Signe and Ane Gyllenberg Foundation (EK, KR), the Sigrid Juselius Foundation (EK), the Finnish Medical Foundation (HL), the Foundation for Pediatric Research (Lastentautien tutkimussäätiö, EK), the Novo Nordisk Foundation (EK), the Jane and Aatos Erko Foundation (HL), and the Päivikki and Sakari Sohlberg Foundation (HL, PV). The funders had no role in the design of the study, data collection and analysis, interpretation of the data, or writing the manuscript.

Availability of data and materials

The mass spectrometry proteomics data supporting the conclusions of this article are available at the to the ProteomeXchange Consortium via the PRIDE partner repository [28] with the dataset identifier PXD010276. Additional data supporting the conclusions of this article are included within the article and its additional files.

Declarations

Ethics approval and consent to participate

Written informed consent was obtained from all participants before their inclusion in the study. This study was approved by the Ethics committee of Obstetrics and Gynecology, Hospital District of Helsinki and Uusimaa (Dnro 3/ E8/05), and carried out in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Transplantation laboratory, Faculty of Medicine, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. ²Present address: Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. ³Department of Reproductive Biology, All India Institute of Medical Sciences, New Delhi, India. ⁴Finnish Institute for Health and Welfare, Helsinki, Finland. ⁵Children's Hospital, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. ⁶PEDEGO Research Unit, MRC Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland. ⁷Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. ⁸Department of Obstetrics and Gynecology, Tampere University Hospital, Tampere, Finland. ⁹Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. ¹⁰Department of Medical and Clinical Genetics, Faculty of Medicine, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. ¹¹Institute for Molecular Medicine Finland, Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland. ¹²Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. ¹³Hyvinkää Hospital, Helsinki and Uusimaa Hospital District, Hyvinkää, Finland. ¹⁴Department of Clinical Chemistry, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. ¹⁵Department of Clinical Chemistry, Institute of Clinical Medicine, University of Eastern Finland, Kuopio University Hospital, Kuopio, Finland.

Received: 14 June 2021 Accepted: 10 January 2022

Published online: 19 January 2022

References

- Lisonkova S, Joseph KS. Incidence of preeclampsia: risk factors and outcomes associated with early- versus late-onset disease. *Am J Obstet Gynecol*. 2013;209(6):544 e1 - e12.
- Report of the National High Blood Pressure Education Program Working Group on high blood pressure in pregnancy. *Am J Obstet Gynecol*. 2000;183(1):S1-S22.
- Steegers EA, von Dadelszen P, Duvekot JJ, Pijnenborg R. Pre-eclampsia. *Lancet*. 2010;376(9741):631-44.
- Rolnik DL, Wright D, Poon LC, O'Gorman N, Syngelaki A, de Paco MC, et al. Aspirin versus placebo in pregnancies at high risk for preterm preeclampsia. *N Engl J Med*. 2017;377(7):613-22.
- Askie LM, Duley L, Henderson-Smart DJ, Stewart LA, Group PC. Antiplatelet agents for prevention of pre-eclampsia: a meta-analysis of individual patient data. *Lancet*. 2007;369(9575):1791-8.
- Roberge S, Giguere Y, Villa P, Nicolaidis K, Vainio M, Forest JC, et al. Early administration of low-dose aspirin for the prevention of severe and mild preeclampsia: a systematic review and meta-analysis. *Am J Perinatol*. 2012;29(7):551-6.
- Villa PM, Hämäläinen E, Mäki A, Räikkönen K, Pesonen A-K, Taipale P, et al. Vasoactive agents for the prediction of early- and late-onset preeclampsia in a high-risk cohort. *BMC Pregnancy Childbirth*. 2013;13(1):110.
- O'Gorman N, Wright D, Syngelaki A, Akolekar R, Wright A, Poon LC, et al. Competing risks model in screening for preeclampsia by maternal factors and biomarkers at 11-13 weeks gestation. *Am J Obstet Gynecol*. 2016;214(1):103 e1 - e12.
- 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 Gynaecol Obstet*. 2019;145(Suppl 1):1-33.
- Kolialexi A, Mavreli D, Tounta G, Mavrou A, Papantoniou N. Urine proteomic studies in preeclampsia. *Proteomics Clin Appl*. 2015;9(5-6):501-6.
- Buhimschi IA, Zhao G, Funai EF, Harris N, Sasson IE, Bernstein IM, et al. Proteomic profiling of urine identifies specific fragments of SERPINA1 and albumin as biomarkers of preeclampsia. *Am J Obstet Gynecol*. 2008;199(5):551 e1-16.
- Carty DM, Siwy J, Brennand JE, Zurbig P, Mullen W, Franke J, et al. Urinary proteomics for prediction of preeclampsia. *Hypertension*. 2011;57(3):561-9.
- Kononikhin AS, Starodubtseva NL, Bugrova AE, Shirokova VA, Chagovets VV, Indeykina MI, et al. An untargeted approach for the analysis of the urine peptidome of women with preeclampsia. *J Proteome*. 2016;149:38-43.
- Kononikhin AS, Zakharova NV, Sergeeva VA, et al. Differential diagnosis of preeclampsia based on urine Peptidome features revealed by high resolution mass spectrometry. *Diagnostics (Basel)*. 2020;10(12):1039.
- Villa PM, Kajantie E, Raikkonen K, Pesonen AK, Hamalainen E, Vainio M, et al. Aspirin in the prevention of pre-eclampsia in high-risk women: a randomised placebo-controlled PREDO trial and a meta-analysis of randomised trials. *BJOG*. 2013;120(1):64-74.
- Girchenko P, Lahti M, Tuovinen S, Savolainen K, Lahti J, Binder EB, et al. Cohort profile: prediction and prevention of preeclampsia and intrauterine growth restriction (PREDO) study. *Int J Epidemiol*. 2017;46(5):1380-1g.
- Saraswat M, Joenvaara S, Seppanen H, Mustonen H, Haglund C, Renkonen R. Comparative proteomic profiling of the serum differentiates pancreatic cancer from chronic pancreatitis. *Cancer Med*. 2017;6(7):1738-51.
- Silva JC, Gorenstein MV, Li GZ, Vissers JP, Geromanos SJ. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics*. 2006;5(1):144-56.
- Poorolajal J, Jenabi E. The association between body mass index and preeclampsia: a meta-analysis. *J Matern Fetal Neonatal Med*. 2016;29(22):3670-6.
- Cnossen JS, Leeflang MM, de Haan EE, Mol BW, van der Post JA, Khan KS, et al. Accuracy of body mass index in predicting pre-eclampsia: bivariate meta-analysis. *BJOG*. 2007;114(12):1477-85.
- van den Berg CB, Duvekot JJ, Guzel C, Hansson SR, de Leeuw TG, Steegers EA, et al. Elevated levels of protein AMBP in cerebrospinal fluid of women with preeclampsia compared to normotensive pregnant women. *Proteomics Clin Appl*. 2017;11(1-2).
- Kim SM, Cho BK, Kang MJ, Norwitz ER, Lee SM, Lee J, et al. Expression changes of proteins associated with the development of preeclampsia in maternal plasma: a case-control study. *Proteomics*. 2016;16(10):1581-9.
- Park J, Cha DH, Lee SJ, Kim YN, Kim YH, Kim KP. Discovery of the serum biomarker proteins in severe preeclampsia by proteomic analysis. *Exp Mol Med*. 2011;43(7):427-35.
- Chen Y, Lin L. Potential value of coagulation parameters for suggesting preeclampsia during the third trimester of pregnancy. *Am J Med Sci*. 2017;354(1):39-43.
- Ding W, Qiu B, Cram DS, Chen X, Li S, Zhou X, et al. Isobaric tag for relative and absolute quantitation based quantitative proteomics reveals unique urinary protein profiles in patients with preeclampsia. *J Cell Mol Med*. 2019;23(8):5822-6.
- Jacobo-Baca G, Salazar-Ybarra RA, Torres-de-la-Cruz V, Guzman-Lopez S, Elizondo-Omana RE, Guzman-Lopez A, et al. Proteomic profile of preeclampsia in the first trimester of pregnancy. *J Matern Fetal Neonatal Med*. 2020:1-7.

27. Sibai BM. Diagnosis and management of gestational hypertension and preeclampsia. *Obstet Gynecol.* 2003;102(1):181–92.
28. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* 2019;47(D1):D442–D50.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

