

Multomics Longitudinal Modeling of Preeclamptic Pregnancies

Ivana Maric (✉ ivanam@stanford.edu)

Stanford University School of Medicine <https://orcid.org/0000-0002-9441-521X>

Kévin Contrepois

Stanford University <https://orcid.org/0000-0001-9678-5161>

Mira Moufarrej

Stanford University School of Medicine

Ina Stelzer

Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine

<https://orcid.org/0000-0002-9974-4661>

Dorien Feyaerts

Stanford University

Xiaoyuan Han

Stanford University <https://orcid.org/0000-0001-6394-3055>

Andy Tang

Stanford University School of Medicine

Natalie Stanley

Stanford University

Ronald Wong

Stanford University School of Medicine <https://orcid.org/0000-0003-1205-6936>

Gavin Traber

Stanford University School of Medicine

Mathew Ellenberger

Stanford University School of Medicine

Alan Chang

Stanford University School of Medicine

Ramin Fallahzadeh

Department of Anesthesiology, Perioperative and Pain Medicine, School of Medicine, Stanford University, Stanford, CA

Huda Nassar

Stanford University School of Medicine

Martin Becker

Stanford University School of Medicine <https://orcid.org/0000-0003-4296-3481>

Maria Xenochristou

Stanford University School of Medicine

Camilo Espinosa

Stanford University

Davide De Francesco

Stanford University

Mohammad Sajjad Ghaemi

Stanford University

Elizabeth Costello

Stanford University <https://orcid.org/0000-0002-6441-2931>

Anthony Culos

Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine

<https://orcid.org/0000-0003-0083-6994>

Xuefeng Ling

Stanford University

Karl Sylvester

Stanford University School of Medicine

Gary Darmstadt

Stanford University School of Medicine

Virginia Winn

Stanford University <https://orcid.org/0000-0003-1136-2907>

Gary Shaw

Stanford University

David Relman

Stanford University <https://orcid.org/0000-0001-8331-1354>

Stephen Quake

Stanford University <https://orcid.org/0000-0002-1613-0809>

Martin Angst

Stanford University School of Medicine <https://orcid.org/0000-0002-1550-8136>

Michael Snyder

Stanford University School of Medicine <https://orcid.org/0000-0003-0784-7987>

David Stevenson

Stanford University

Brice Gaudilliere

Stanford University <https://orcid.org/0000-0002-3475-5706>

Nima Aghaeepour

Stanford University <https://orcid.org/0000-0002-6117-8764>

Keywords: preeclampsia, physiology, diagnosis

Posted Date: January 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-138586/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Title: Multiomics Longitudinal Modeling of Preeclamptic Pregnancies

Authors: Ivana Marić¹⁺, Kévin Contrepois²⁺, Mira Moufarrej⁵, Ina Stelzer³, Dorien Feyaerts³, Xiaoyuan Han⁸, Andy Tang³, Natalie Stanley³, Ronald J. Wong¹, Gavin M. Traber², Mathew Ellenberger², Alan Chang³, Ramin Fallahzadeh³, Huda Nassar³, Martin Becker³, Maria Xenochristou³, Camilo Espinosa³, Davide De Francesco³, Mohammad Sajjad Ghaemi^{3,10}, Elizabeth Costello⁹, Anthony Culos³, Xuefeng B. Ling⁷, Karl G. Sylvester⁷, Gary L. Darmstadt, Virginia D. Winn⁴, Gary M. Shaw¹, David A. Relman^{9,*}, Stephen R. Quake^{5,*}, Martin Angst^{3,*}, Michael P. Snyder^{2,*}, David K. Stevenson^{1,*}, Brice Gaudilliere^{1,3,*}, Nima Aghaeepour^{1,3,6,*}

¹Department of Pediatrics, Stanford University School of Medicine

²Department of Genetics, Stanford University School of Medicine

³Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine

⁴Department of Obstetrics and Gynecology, Stanford University School of Medicine

⁵Departments of Bioengineering and Applied Physics, Stanford University and Chan Zuckerberg Biohub

⁶Department of Biomedical Data Science, Stanford University

⁷Department of Surgery, Stanford University School of Medicine

⁸University of the Pacific, Arthur A. Dugoni School of Dentistry

⁹ Departments of Medicine, and of Microbiology & Immunology, Stanford University School of Medicine

¹⁰Digital Technologies Research Centre, National Research Council Canada, Toronto, Ontario, Canada

+Co-first authorship

*Co-senior authorship

CORRESPONDING AUTHOR:

Ivana Marić, PhD

Division of Neonatology and Developmental Medicine, Department of Pediatrics

Stanford University School of Medicine

Medical School Office Building, X1C55

1265 Welch Road

Stanford, CA 94305

Phone: 650-724-9571

Fax: 650-721-5751

Email: ivanam@stanford.edu

ABSTRACT

Preeclampsia is a complex disease of pregnancy whose physiopathology remains unclear and that poses a threat to both mothers and infants. Specific complex changes in women's physiology precede a diagnosis of preeclampsia. Understanding multiple aspects of such a complex changes at different levels of biology, can be enabled by simultaneous application of multiple assays. We developed prediction models for preeclampsia risk by analyzing six omics datasets from a longitudinal cohort of pregnant women. A machine learning-based multiomics model had high accuracy (area under the receiver operating characteristics curve (AUC) of 0.94, 95% confidence intervals (CI): [0.90, 0.99]). A prediction model using only ten urine metabolites provided an accuracy of the whole metabolomic dataset and was validated using an independent cohort of 16 women (AUC=0.87, 95% CI: [0.76, 0.99]). Integration with clinical variables further improved prediction accuracy of the urine metabolome model (AUC=0.90, 95% CI: [0.80, 0.99], urine metabolome, validated). We identified several biological pathways to be associated with preeclampsia. The findings derived from models were integrated with immune system cytometry data, confirming known physiological alterations associated with preeclampsia and suggesting novel associations between the immune and proteomic dynamics. While further validation in larger populations is necessary, these encouraging results will serve as a basis for a simple, early diagnostic test for preeclampsia.

INTRODUCTION

The World Health Organization estimates that more than 800 women worldwide die from pregnancy-related causes every day, with the highest rates of maternal mortality and morbidity in low-income countries¹. One of the main causes is the hypertensive disorder of pregnancy – preeclampsia - for which the only treatment is to deliver, often too early. Preeclampsia affects 3–5% of pregnancies in the United States and up to 8% of all pregnancies globally¹, and accounts for 10–15% of maternal deaths² and 15–20% of preterm births³.

The pathophysiology of preeclampsia is complex and is thought to be caused in part by abnormal placentation as well as a women's predisposition through genetic and immunological factors⁴. It is believed that the abnormal placentation leads to a maternal inflammatory response⁴. Placental ischemia, oxidative stress and the presence of a maternal angiogenic imbalance are all characteristics of preeclampsia^{5,6}, leading to endothelial and end-organ damage, and in some cases to stroke and even death.

Specific biological processes involved in the development of preeclampsia have remained understudied. Early prediction of preeclampsia has remained a clinical challenge, owing to incompletely understood causes, various risk factors and likely multiple pathogenic phenotypes of preeclampsia^{7,8}. The recent availability of high-throughput omics (including the genome, transcriptome, proteome and metabolome) assays, where each can be performed on small sample volumes, has enabled joint analyses of the high-dimensional multidomain or *multiomics* data measured from the same biological sample^{4,9,10}. The integrated analysis may capture complex dynamics involved in the pathogenesis of preeclampsia that could ultimately lead to

novel therapeutic interventions. Furthermore, applying machine learning methods capable of extracting the most predictive features from high-dimensional multiomics data, could lead to more accurate predictive models and better early detection of women at risk to develop preeclampsia.

In this unique study, we performed a multiomics analysis of the transcriptome, proteome, metabolome, lipidome, and microbiome from a coordinated set of biospecimen collected longitudinally from pregnant women; we then integrated immune system mass spectrometry features that were available for a subset of the patients; and we combined the multiomics data with the available clinical/demographics data and performed joint analysis. Our goals were to: 1) build an integrated multiomics predictive model of preeclampsia; 2) compare prediction capabilities of different omics sets; 3) develop a simple and interpretable predictive model based on a small number of biomarkers that can be used for a diagnostic test; 4) identify a specific signature of preeclampsia; and 5) gain insights into pathways involved in the pathogenesis of preeclampsia.

RESULTS

Multiomics Characterization of Normal and Preeclamptic Pregnancy Over Gestation

Thirty-three and sixteen women were included in the discovery and validation cohorts, respectively (Fig. 1A). Maternal characteristics, demographics, and gestational ages at delivery are shown in Table S1. In the discovery cohort, 17 women developed preeclampsia and 16 had a pregnancy unaffected with preeclampsia. Among the preeclampsia patients, severe vs. mild preeclampsia was observed in 10 and 7 women, respectively; early- vs. late-onset preeclampsia

was observed in 5 and 12 women, respectively (Table S2). An additional cohort of 16 women, out of which 12 had preeclampsia, was used to validate the metabolomics results. Blood samples were collected longitudinally at two or three time points during pregnancy (early, mid, and late – see Fig. 1). Plasma, urine, and vaginal swabs from each woman were used for measurements of cfRNA (plasma transcriptome), proteome (plasma), metabolome (plasma and urine), lipidome (plasma), and microbiome (vaginal swab). The number of measurements differed markedly among omics datasets, with transcriptome containing the highest number of measurements (Fig. S1A). In contrast, the number of principal components explaining 90% of variance, that quantifies the internal correlation of a dataset, exhibited smaller difference among datasets (Fig. S1B). Thus, although the amount of data varied several orders of magnitude among the dataset, their variability and thus the amount of information content was much more similar.

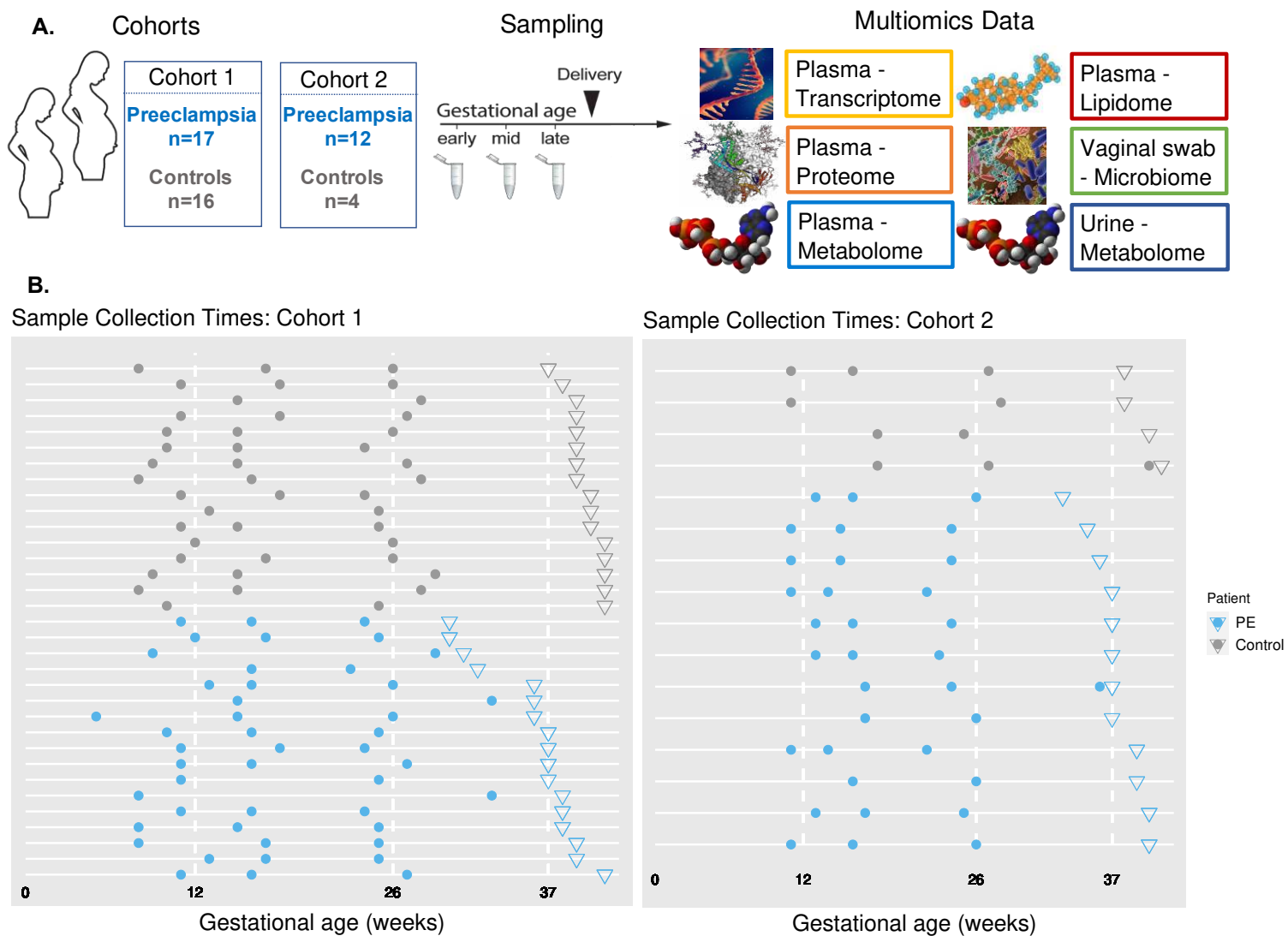


Figure 1. Overview of the study. A. Two independent cohorts were analyzed using 6 different assays. **B.** Sample collection timeline for plasma in discovery and validation cohorts. Circles indicate pre-delivery sample collection times and inverted triangles indicate delivery dates for individual women (one per horizontal line).

Machine Learning Modeling of Preeclampsia Over Gestation

Multivariate models of preeclampsia were built for each dataset using the Elastic Net (EN) algorithm (see Methods). Predictions from separate models were then integrated in a final model using stacked regression. The performance of all models was evaluated using the leave-

one-out cross-validation method). The integrated model exhibited high prediction accuracy (AUC=0.94, 95% confidence intervals (CI): [0.90, 0.99]) and outperformed predictions from each separate model in terms of the point estimate (Fig. 2). EN models generated from the proteome and urine metabolome exhibited high performance (AUC=0.89, 95% CI: [0.83, 0.96]; AUC = 0.87, 95% CI: [0.80, 0.94], respectively).

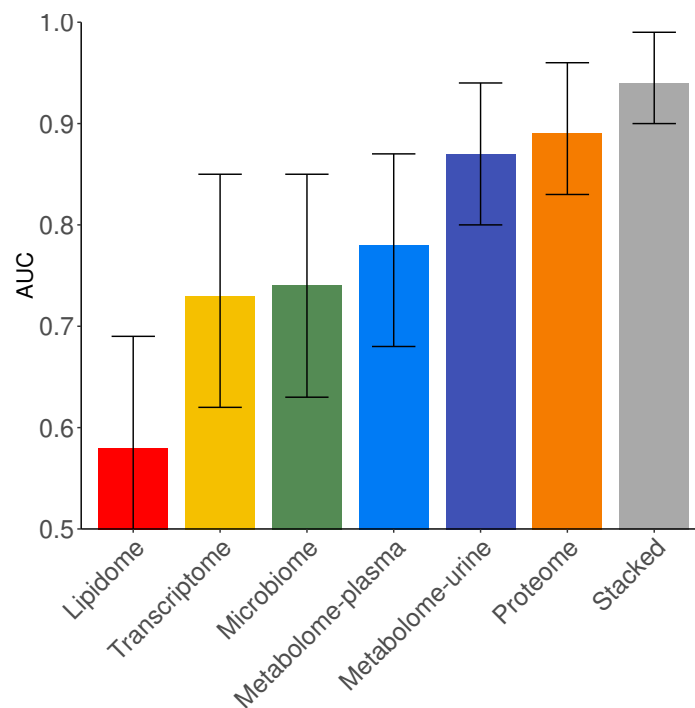


Figure 2. Single- and Multi-omics machine models for preeclampsia. Performance comparison of machine learning models shown on x-axis in terms of the area under the receiver operator curve (AUC) shown on y-axis. The integrated (stacked) model utilizing stacked regression exhibited the highest accuracy (AUC=0.94, 95% CI [0.9, 0.94]). Both proteome and metabolome (urine) had high prediction performance (AUC=0.89, 95% CI [0.83, 0.96] proteome; AUC = 0.87, 95% CI [0.80, 0.94] urine metabolome).

In order to test the applicability of these models to resource-limited settings, machine learning models were developed that use a small set of the most informative features from the plasma proteomic and urine metabolomic datasets – the two sets that had the best prediction accuracy – as follows. For each data set, in each cross-validation step, we used a subset of features selected by EN to refit a regression model. The performance of refitted models as a function of the number of features is shown in Fig. 3A. The best prediction model was obtained with 30 urine metabolites. A model using only ten metabolites provided the accuracy of the whole urine metabolomic dataset (AUC=0.88, 95% CI: [0.81, 0.95]) (Figs. 3A and 3B). Top urine metabolites included adenine, isovalerylglutamic acid, uric acid ribonucleoside, N-Acetyl-O-acetylneuraminic acid, 1,5-anhydroglucitol, dehydroepiandrosterone, sialyllactose, N-epsilon-acetyl-L-lysine, imidazolelactic acid, and nonanoylcarnitine. The levels of predictive metabolites differed greatly between control and preeclamptic women (Fig. 3C) and these differences were usually present during each trimester. As expected, EN models varied slightly due to variability of the chosen training set in each leave-one out cross-validation step¹¹ and therefore, the ten metabolites chosen by EN varied slightly across cross-validations. We recorded frequency of occurrence for every feature across all cross-validation steps (Fig. S2). Having high frequency of occurrence indicates that the feature is relevant for all or a majority of patients, that is, it is more stable¹¹. The model using ten urine metabolites was validated in an independent cohort, with an AUC of 0.87 (95% CI: [0.76, 0.99]) (Fig.4).

A model using the ten top-scoring plasma proteins achieved an AUC of 0.83 (95% CI: (0.73, 0.92]) (Figs 3A and 3B.) The most predictive plasma proteins selected by EN included leptin (LEP), vascular endothelial growth factor A (VEGFA), L-selectin (SELL), E-selectin (SELE),

interleukin-24 (IL-24), interleukin-22 (IL-22), and tyrosine-protein kinase transmembrane receptor (ROR1) (Fig. 3D).

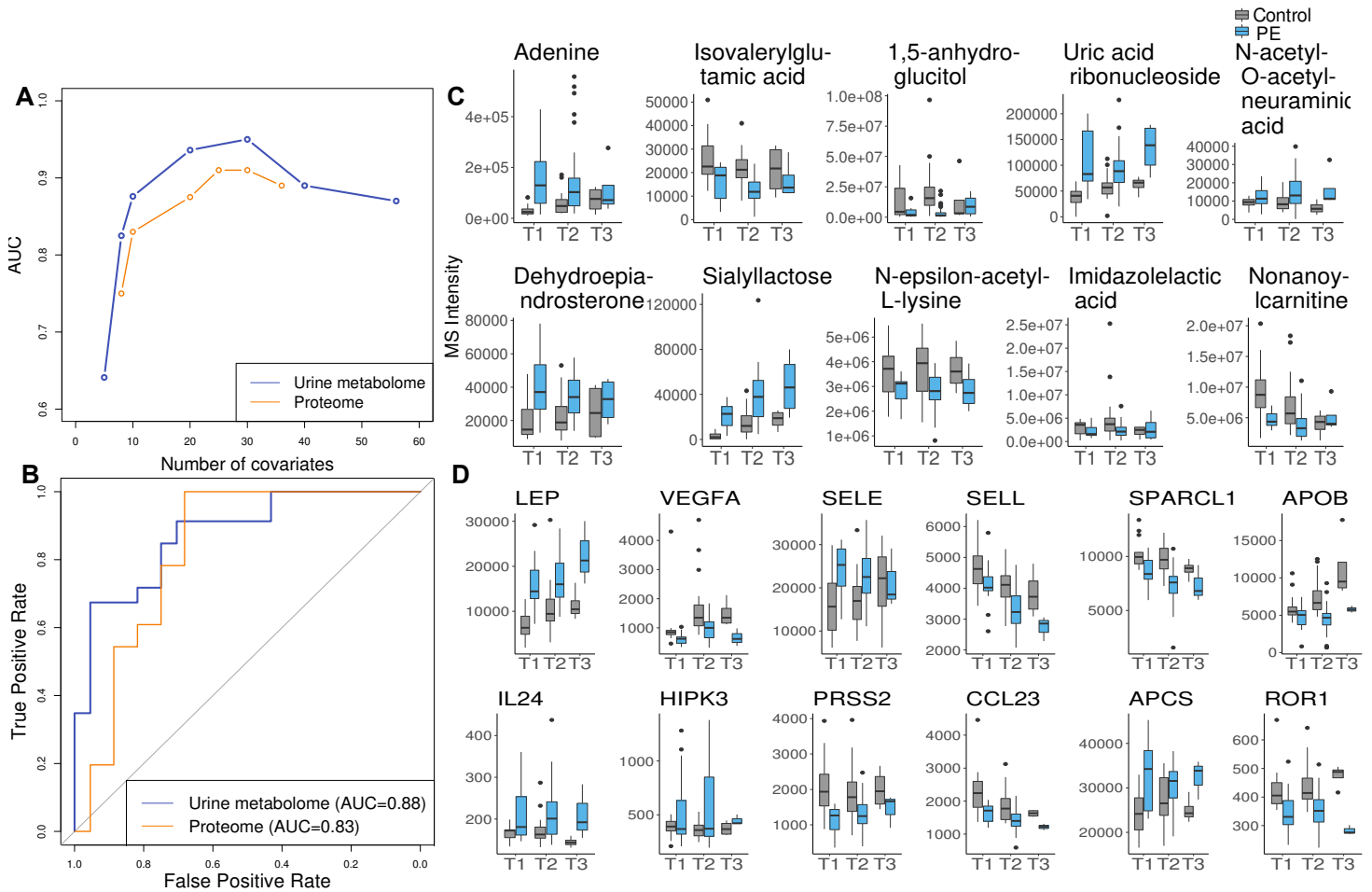


Figure 3. Refined model for prediction of preeclampsia in resource-limited settings: urine metabolomic and plasma proteomic data sets. **A.** Cross-validated performance of the refitted models as a function of the number of features (support) for urine metabolome (blue) and proteome (orange). The best performance is observed for a model with 30 features. A model using ten metabolites provided the accuracy of the whole urine metabolomic dataset (AUC=0.88, 95% CI: [0.81, 0.95]). **B.** Performance of cross-validated prediction model using ten metabolites (blue) or ten proteins (orange). For the urine metabolome, prediction from ten features had the same performance as the prediction from the full metabolomics set (AUC =

0.88, 95% CI [0.81, 0.95]). Prediction from ten proteins yielded performance of AUC = 0.83, 95% CI: [0.73, 0.92]. **C.** Most informative metabolites. Y-axis shows a metabolite value stratified by normal pregnancy (grey) and preeclamptic pregnancy (blue). **D.** Most informative proteins. Y-axis shows a protein value stratified by normal pregnancy (grey) and preeclamptic pregnancy (blue).

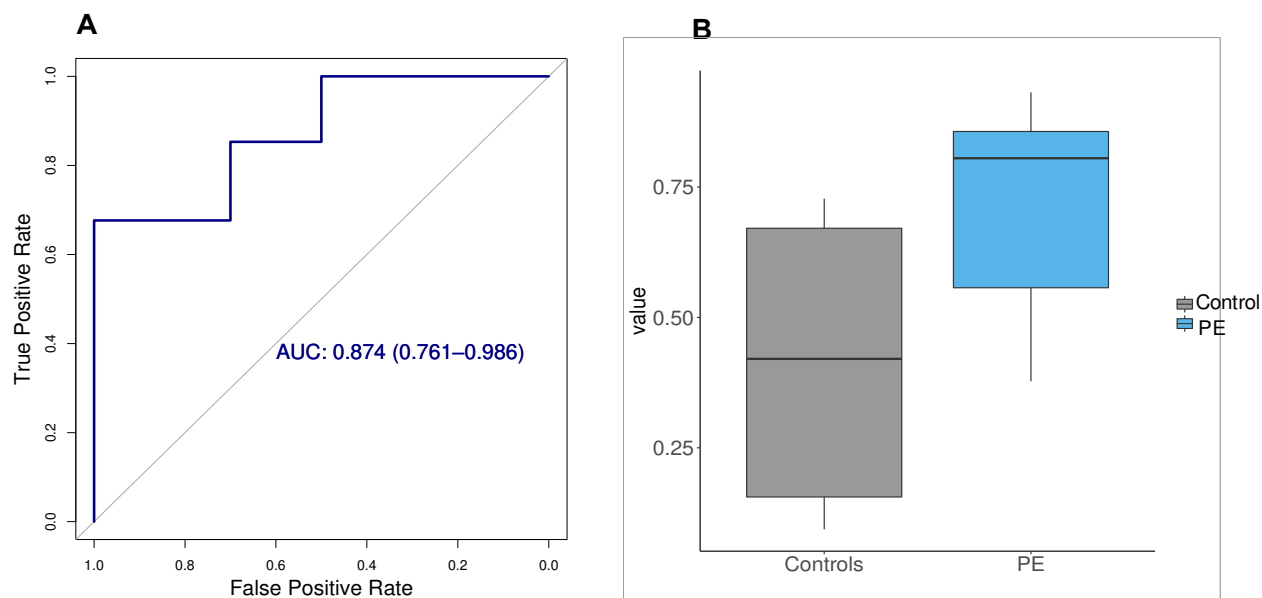


Figure 4. Validated model from urine metabolome using ten metabolites from the validation cohort. A. Area under the receiver operating curve, AUC = 0.874, 95% CI of [0.76, 0.99]. **B.** Prediction values (scores) for controls and preeclampsia (PE) groups.

In order to explore multiomics interactions of analytes associated with preeclampsia, correlation network for controls and preeclampsia patients were generated (Fig. 5). While some features retained their close correlations (e.g. LEP and urine metabolite N-Acetyl-O-acetylneuraminic acid) in preeclampsia patients (Fig. 5B) relative to controls (Fig. 5A), some interactions were changed (e.g., IL1RAP and adenine). This resulted in formation of different

clusters in the correlation network for the two groups of patients, potentially reflecting differential mechanisms in preeclampsia.

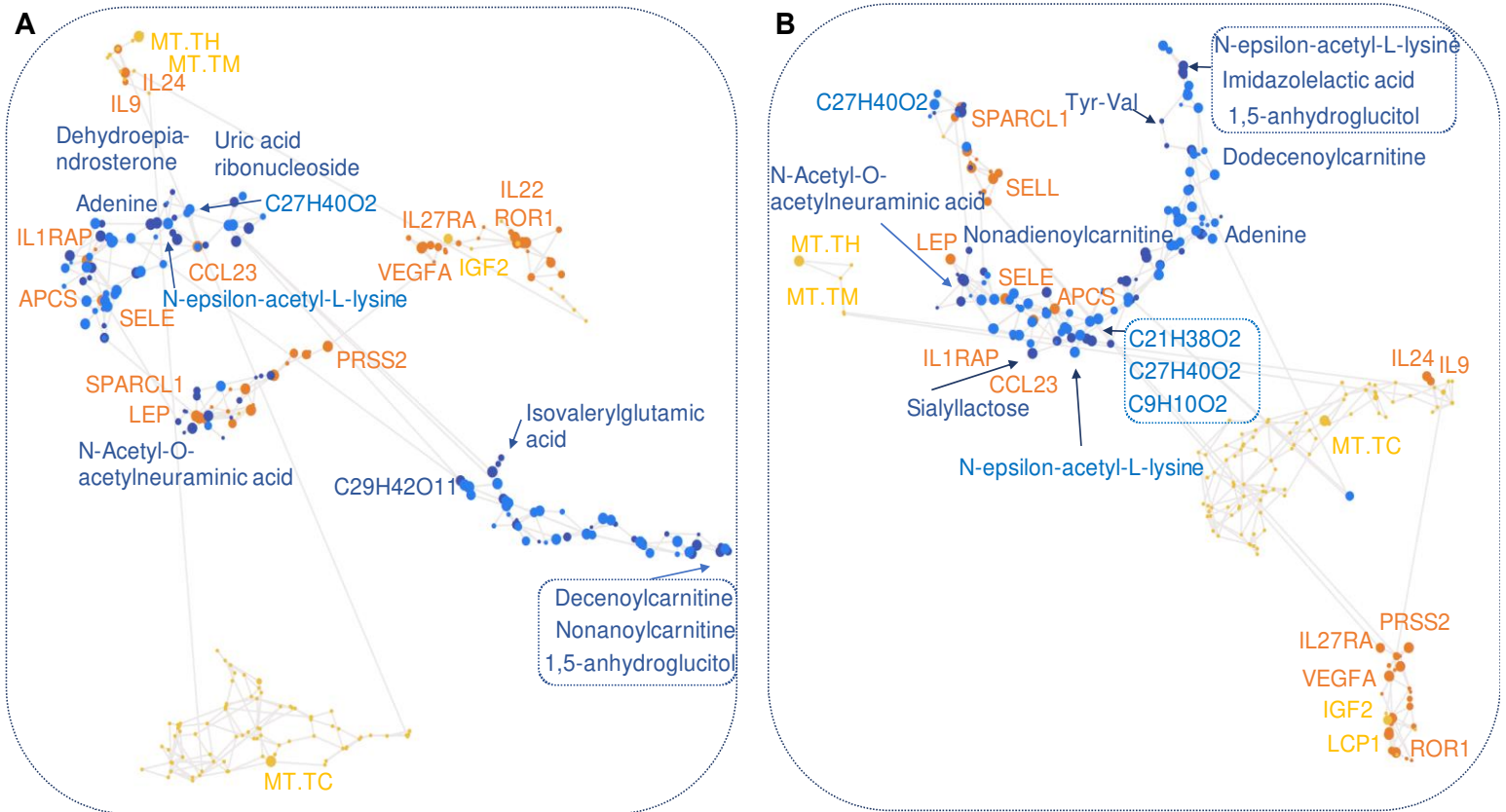


Figure 5. Visualization of predictive features of transcriptome (yellow), proteome (orange), urine metabolome (dark blue), and plasma metabolome (light blue). Size of each node is proportional to the frequency at which it was chosen in prediction models during cross-validation. High frequency of occurrence indicates that a feature is relevant for all or a majority of patients resulting in a more stable model. **A.** Controls. **B.** Preeclampsia patients. We observe different clusters for the two groups of patients.

Prediction of Preeclampsia in Early Pregnancy

From a clinical perspective, early prediction of preeclampsia (*i.e.*, within the first 16 weeks of gestation) is of critical importance as it would allow for early treatment of high-risk women (*e.g.*, with low-dose aspirin¹²). It would also enable closer monitoring of high-risk pregnancies and allow for the enrichment of preemptive interventional studies in women at risk for developing preeclampsia.¹³ Finally, the development of a simple and affordable diagnostic test would facilitate more ubiquitous testing in both high-income and low- and middle-income countries. For these reasons, an EN model using only biological data from samples collected only during first 16 weeks of pregnancy was derived. Plasma proteome and urine metabolome datasets generated predictive models with the highest performances (AUC=0.88, 95% CI: [0.75, 1]; AUC=0.83, 95% CI: [0.69, 0.98], respectively, Fig. S4A) similarly to what we observed when using all samples collected during pregnancy. The refitted model with 10 features for the plasma proteome and urine metabolome provided high accuracy (AUC=0.88, 95% CI: [0.75, 1]; AUC=0.87, 95% CI: [0.74, 1], respectively, Fig. 6A). The most predictive metabolites and proteins also had high frequency of occurrence in the leave-one-out validation (Fig. S3). The heatmap of rank values of features selected by EN in the refitted models from all omics sets shows that these features distinguish women with preeclampsia from controls with no misclassification error (Fig. 6B). The plasma levels of the top-ranking proteins (Fig. 6C), urine metabolites (Fig. S4B) and genes (Fig. S4C) selected by EN, measured early in pregnancy, markedly differed between controls and preeclamptic women. Fig. 7 summarizes the most predictive plasma protein and urine metabolites both in early pregnancy and during entire pregnancy. Plasma proteins that were predictive in both models included LEP, SELL, CCL23, HIPK3, APCS, GPNMB, and IL-24, while some were mostly predictive in early pregnancy (FGF19, TIMP2), and others

over entire gestation (VEGFA, SELE, SPARCL1, APOB, ROR1, and IL-22) (Fig. 7A). Urine metabolites that were predictive over entire gestation mostly differed from urine metabolites predictive in early pregnancy (Fig 7B) with the exception of adenine and nonadienoylcarnitine.

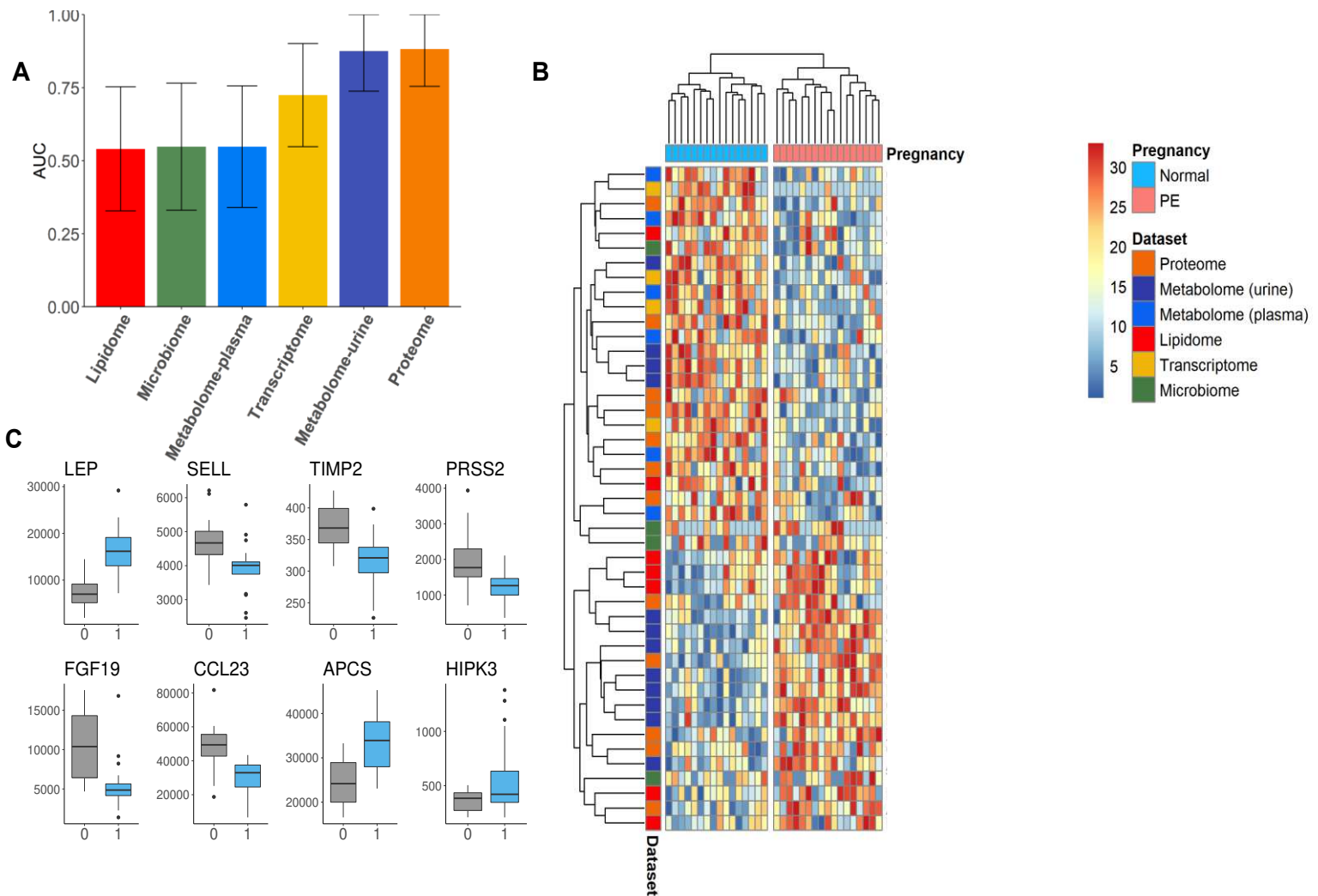


Figure 6. Prediction models in early pregnancy. **A.** Performance comparison of EN models derived from different omics data sets using the top-ranked ten features, in terms of the AUC. The plasma proteomic and the urine metabolomic models performed best (AUC=0.88, 95% CI of [0.75, 1] for proteome; AUC=0.87, 95% CI of [0.74, 1] for urine metabolome). **B.** Heatmap of ranked values of features identified by EN, perfectly distinguishing preeclamptic women from

controls. **C.** Top eight proteins. Y-axis shows the value in early pregnancy stratified by normal (grey) versus preeclamptic pregnancy (light-blue).

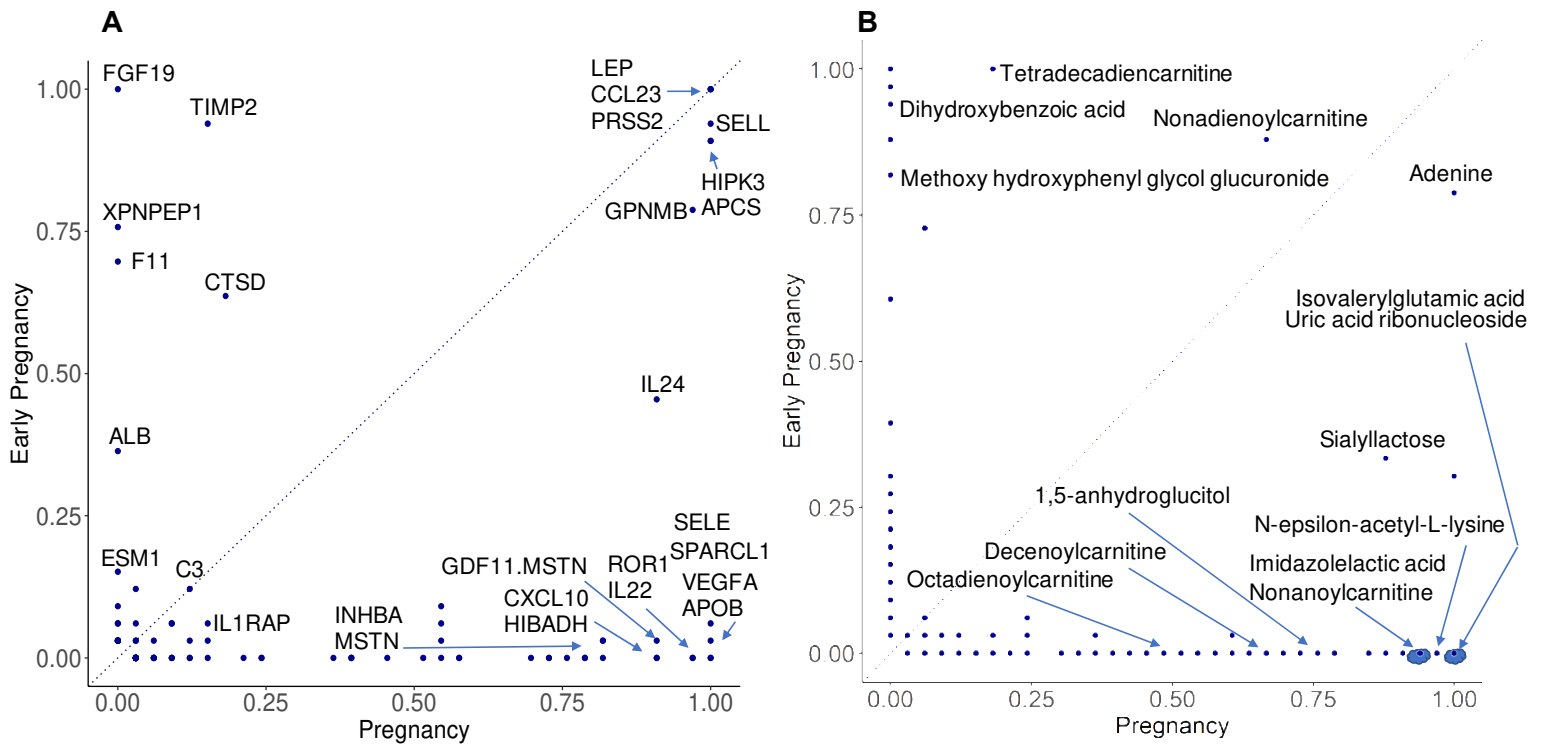


Figure 7. Biomarker comparison: entire pregnancy vs. early pregnancy. X-axis and Y-axis show the respective frequency of each biomarker in early pregnancy and over gestation. **A.** Most predictive proteins. **B.** Most predictive urine metabolites. Blue circles around dots imply the same position for more than one protein/urine metabolite.

Single Cell Characterization of the Immune System

Preeclampsia is strongly associated with inflammation and aberrant maternal immune system adaptations during pregnancy¹⁴. To assess immunity -- which is complementary to pathways covered by proteins and metabolites -- and connect differential abundances of plasma proteins and urine metabolites in preeclamptic pregnancies to biological changes, immune-system wide mass cytometry measurements of single cells obtained in a subset of the same patient cohort

were integrated with our plasma proteome and urine metabolome prediction models, as these two models had the best accuracy. Immune cell dynamics between 1st and 2nd trimester blood samples obtained from high-dimensional mass cytometry were previously used to develop a prediction model of preeclampsia¹⁵. We found that seven of the immune features reported by Han et al. correlated highly with the prediction based on our integrated algorithm (Spearman correlation $p < 0.05$) (Figure 8A, highlighted in orange), confirming the predictive value of both immune cell features as well as plasma proteins and urine metabolites. To investigate whether this correlation between predictive features was biologically meaningful, we focused on the correlations of feature behavior between the eight earlier reported predictive immune features (Fig. 8A) and the top twelve most informative plasma proteome features (Fig. 3E) across pregnancy (Fig. 8B). Leptin (LEP) and soluble L-Selectin (SELL) levels were particularly strongly correlated with the eight immune cell features (Fig. 8B). Interestingly, basal pSTAT5 signaling in Th1 cells (CD4+Tbet+), the top immune feature to distinguish control from preeclamptic pregnancies (Han et al. 2019), correlated with LEP levels in both control and preeclamptic patients. Uniquely in preeclamptic cases, LEP levels were correlated with basal pSTAT1 signaling in intMCs (Spearman correlation $p = 0.002$) and basal STAT5 signaling in mDCs (Spearman correlation $p = 0.01$). Moreover, SELL levels were uniquely correlated with immune features in preeclamptic pregnancies and not with controls, i.e. correlated with basal pNFkB and pSTAT1 signaling in cMCs, basal pSTAT5 signaling in Th1 cells and mDCs, and basal pMAPKAPK2 signaling in naive CD4 Tcells. Preeclamptic pregnancies were not characterized by - in other words, had potentially lost - concerted proteome/immune behavior which was prominently observed in healthy pregnancies, i.e. correlations of leptin with basal pP38

signaling in Treg and TCRgd cells. These correlations exemplify the biological connection between responsiveness of immune cells and its plasma environment.

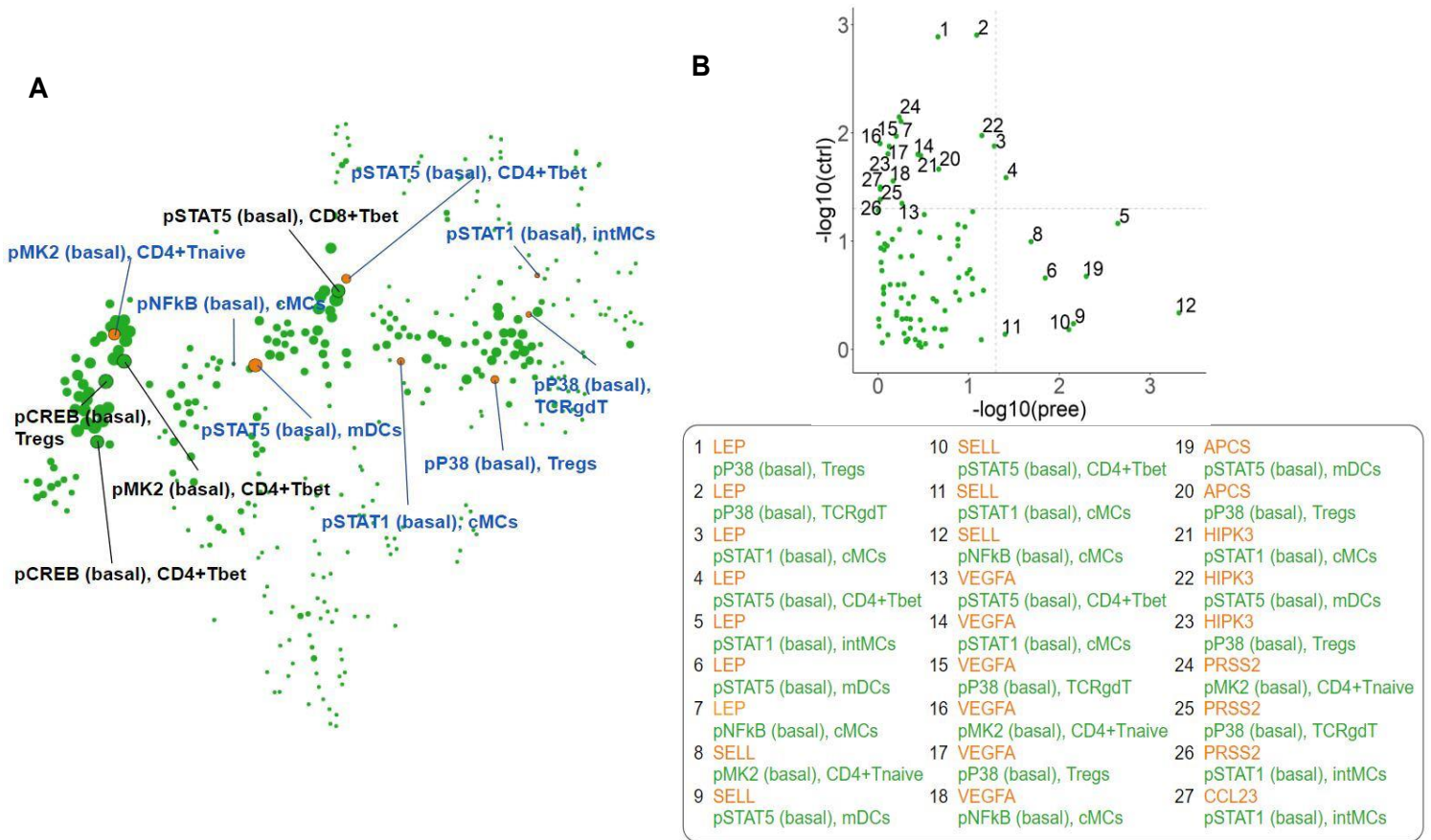


Figure 8. A. Visualization of immune features. Visualization indicates features most correlated with the prediction of the stacked model. Features shown in orange are the seven most predictive immune features that also highly correlate with the multiomics predictive model. Size of each node is proportional to the $-\log_{10}(\text{p-value})$ of spearman correlation. **B.** Comparison of p-value of correlation for the top immune and top proteome features. Each node is a pair comprising an immune and a proteome feature.

Relationship Between Clinical Data and Omics Measurements

Clinical and demographics data contains maternal characteristics known to be associated with the risk of preeclampsia, e.g., preexisting hypertension, race, BMI, height, gravida. We combined ten variables that were available in this data set (Table S1) with the most predictive sets: 1) plasma proteome and 2) urine metabolome models, to better understand their mutual relationship. The ten clinical variables were included together with the top ten omics features – all combined in the single cross-validation step. Inclusion of clinical and demographics data improved the performance both when combined with the plasma proteome and the urine metabolome (urine metabolome AUC=0.96, 95% CI: [0.92, 0.99] and; proteome AUC=0.91, 95% CI: [0.85, 0.97], respectively) (Fig. 9A). The most predictive clinical variables included maternal age, BMI, height, and preexisting hypertension. We observed several significant correlations (Spearman correlation $p < 0.05$) between clinical variables and plasma proteins/urine metabolites that were present only among preeclamptic women. These included: leptin with maternal BMI/weight, in agreement with existing literature¹⁶; CCL23 with height; SELL with gravida (Fig. 9B); maternal age with adenine – previously observed¹⁷ – and maternal age with isovalerylglutamic acid (Fig. 9C).

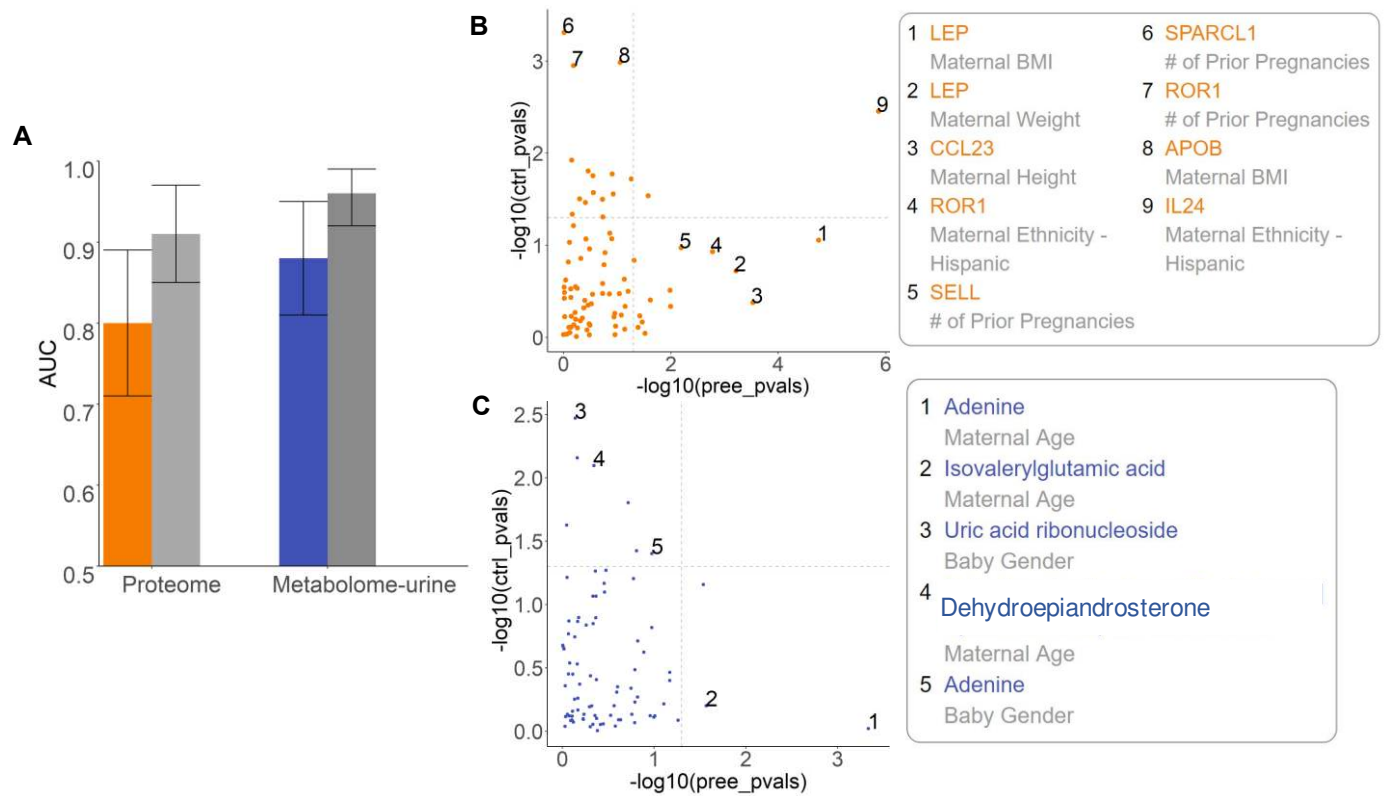


Figure 9. Relationship between urine metabolome and proteome with clinical features over gestation. **A.** Prediction accuracy of urine metabolome and plasma proteome. Dark blue (for urine metabolome) and orange (for proteome) bars show performance without clinical data (proteome: AUC = 0.83, 95% CI: [0.73, 0.92]; urine metabolome: AUC = 0.88, 95% CI [0.81, 0.95]). Grey bars show performance with clinical data (proteome AUC=0.91, 95% CI: [0.85, 0.97]; urine metabolome AUC=0.96, 95% CI: [0.92, 0.99]). **B.** Comparison of p-value of correlations of the top proteome and clinical features. Value of $-\log_{10} p$ for preeclamptic patients and controls is shown on x-axis and y-axis, respectively. Each node is a pair of a proteome and a clinical feature. **C.** Comparison of p-value of correlations of the top urine metabolites and EHR features. Each node is a pair of a proteome/urine metabolome and a clinical feature.

Exploration of Preeclampsia Pathogenesis from Multiomics Measurements

Over the Course of Pregnancy

Changes over gestation of 1215 metabolic features among 8718 were significantly associated with preeclampsia outcome (FDR < 0.05 LME Model with Benjamini-Hochberg procedure).

Pathways enrichment analysis using these urine metabolites identified the following pathways ($p < 0.05$) (Fig. 10A): 1) *tryptophan metabolism*; 2) *caffeine metabolism*; 3) *tyrosine metabolism*; 4) *steroid hormone biosynthesis*; 5) *pentose and glucuronate interconversions*; 6) *linoleic acid metabolism*. The steroid hormone biosynthesis pathway plays an important role in pregnancy progression¹⁸. Both the steroid hormone biosynthesis pathway and the caffeine metabolism pathways have previously been associated to pregnancy¹⁹, and tryptophan metabolism with preeclampsia²⁰. Metabolites in the steroid hormone biosynthesis pathway and in the caffeine metabolism pathway present in the data with high level of significance are respectively shown in Figs. S5C and S5D.

Among 1305 proteins, 437 had changes that were significantly associated with preeclampsia outcome over gestation (FDR < 0.05 LME Model with Benjamini-Hochberg procedure). The top 64 proteins at significance level $p < 5 \cdot 10^{-4}$ (LME Model) showing markedly different values between normal and preeclamptic women are depicted in Fig. S6. Top proteins included interleukin-1 receptor accessory protein (IL1RAP) and SELL – both known to play a role in the immune response²¹ (Fig. S5). Enriched pathways grouped into ten biological processes, the most prevalent being *positive regulation of cellular process* (including biological, cellular, protein metabolic, immune system, and apoptotic processes among others) (Fig. S7). In the cfRNA set, 306 features were significantly associated with preeclampsia outcome over gestation

(FDR < 0.05 LME Model with Benjamini-Hochberg procedure) resulting in several enriched pathways (Fig. S8). Top features included YOD1, BIRC2, CEP63, and LCP1. Top proteome, transcriptome and urine and plasma metabolome features formed 17 distinct communities (Fig. S9).

Early Pregnancy

In early pregnancy, 497 out of 8718 urine metabolic features had changes significantly associated with preeclampsia when compared to controls (FDR<0.05, Wilcoxon signed-rank test with Benjamini-Hochberg procedure). Pathways enrichment analysis on these urine metabolites identified the following pathways ($p < 0.05$) (Fig. 10B): 1) *tyrosine metabolism*; 2) *lysine degradation*; 3) *tryptophan metabolism*; 4) *phenylalanine metabolism*; 5) *steroid hormone biosynthesis*; 6) *arachidonic acid metabolism* which is a central regulator of the inflammatory response and has a known role in the pathogenesis of preeclampsia²²; 7) *phenylalanine, tyrosine and tryptophan biosynthesis* and; 8) *aminoacyl-tRNA biosynthesis*. Individual metabolite from these two pathways are shown in Figs. S5A and S5B.

In the proteome set containing 1305 proteins, three proteins -- LEP, CCL23 and FAM3D -- were significantly associated with preeclampsia outcome (FDR<0.05, Wilcoxon signed-rank test with Benjamini-Hochberg procedure) identifying one significantly enriched pathway, *negative regulation of glucagon secretion* (Fisher's exact test with Benjamini-Hochberg procedure, FDR< 0.05).

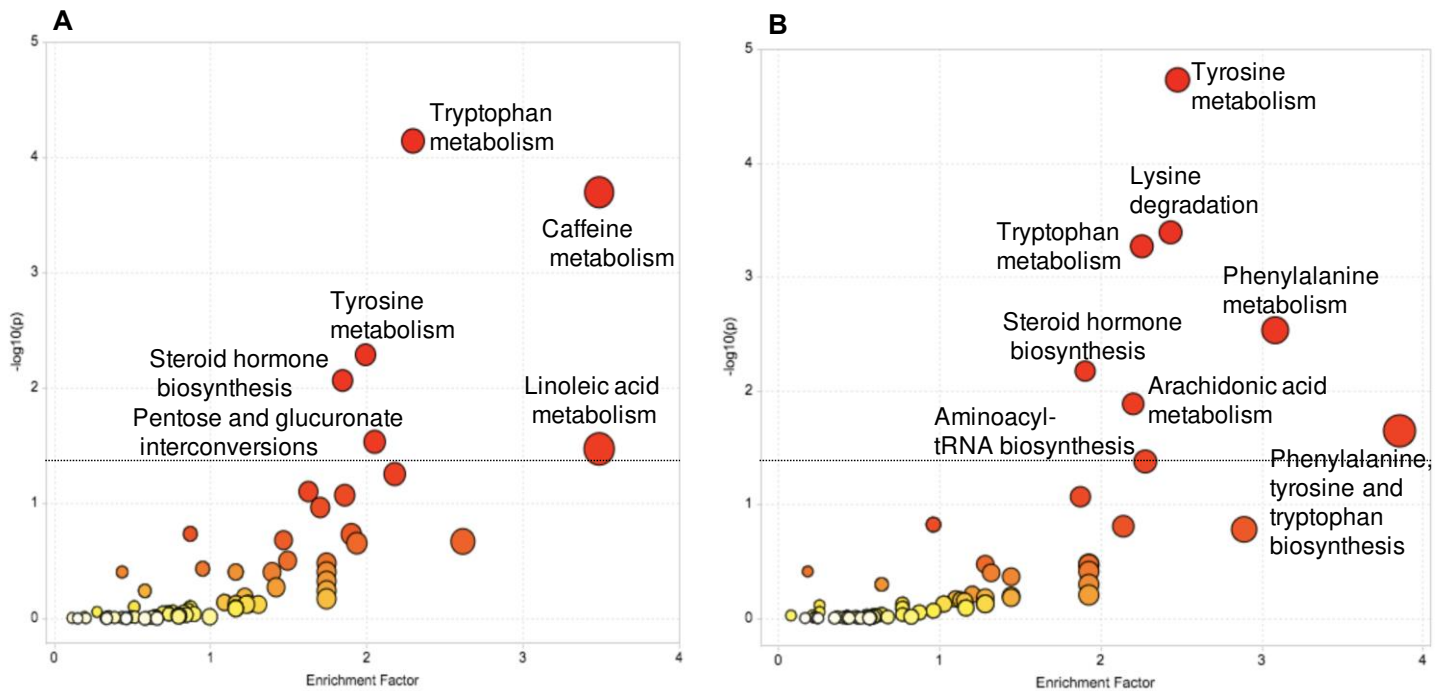


Figure 10. Identified enriched pathways from urine metabolome urine over gestation and in early pregnancy. A. Pathways enrichment analysis over gestation using metabolites from urine that were significant (FDR<0.05, Wilcoxon signed-rank test with Benjamini-Hochberg procedure). Pathways shown above the dotted line were significant ($p < 0.05$). **B.** Pathways enrichment analysis for early pregnancy using metabolites from urine that were significant (FDR<0.05, Linear Mixed Effects model with Benjamini-Hochberg procedure). The color and the size of a circle are proportional to the $-\log(p)$ and pathway impact value, respectively, where p denotes a p -value.

Outlier analysis

We observed that a few patients in our cohort were consistently misclassified by our prediction algorithm (Fig. S10). A few control patients resembled preeclamptic patients on a molecular level in some of the top predictive features, across omics sets. And vice versa, there were some preeclamptic patients whose top molecular features more closely resembled those of controls. Reexamination of the clinical charts revealed that one of the preeclampsia patients, while

clearly hypertensive, had proteinuria in the context of gross hematuria obscuring whether proteinuria related to preeclampsia. Therefore, the patient may have been misdiagnosed with preeclampsia but rather only had gestational hypertension. This highlights that the predictive model can pick up discrepancies within the clinical chart. For the other patients whose clinical diagnosis held, this implies that their phenotypical features that were classifying these patients in either control or preeclampsia group, did not match their *molecular phenotypes*. Of interest, one patient in the preeclamptic group, that prediction classified as control, developed HELLP very late in gestation at 41+3 weeks. So, if she had delivered closer to the due date, she would have been considered a control. Thus, if others in the control group have similar molecular phenotype this may represent a late-onset preeclampsia related to placental aging in the post-term period.

DISCUSSION

Recent omics studies of preeclampsia typically included up to two omics datasets^{10,23,24}. Ours is the first study to present the integrated analysis of six high-throughput omics datasets, containing more than 50,000 measurements per sample. This multiomics analysis enabled uniform comparison of omics sets, and revealed improved predictive ability for preeclampsia status relative to individual biological modalities, and indications of biological processes associated with the disease across multiple modalities.

One of the main strengths of our study is that, in our cohort, biological samples were not only collected longitudinally from each woman; but also, each individual sample was simultaneously measured for proteome, transcriptome, metabolome, lipidome, and vaginal swab for

microbiome, thereby providing a unique opportunity to systematically study changes due to preeclampsia over gestation, and compare the capability of each of these omics sets to predict and characterize preeclampsia. These analyses involved more than 50,000 measurements, which were used in the prediction algorithm to agnostically identify the best biomarkers of preeclampsia.

Among our six datasets, plasma proteomic and urine metabolomic datasets had the highest prediction accuracies, both over gestation and early in pregnancy. A prediction model using only ten urine metabolites provided high accuracy over gestation (AUC=0.88, cross-validated and AUC = 0.87 validated on an independent cohort) and early in pregnancy (AUC=0.875, cross-validated).

The EN prediction model with ten plasma proteins achieved AUC of 0.83 over gestation and of 0.88 in early pregnancy. Vascular endothelial growth factor A (VEGF-A) was among the most predictive proteins. Reduced levels of VEGF-A have previously been described in preeclamptic pregnancies due to increased levels of placental soluble fms-like tyrosine kinase-1 (sFLT-1) which validate our study²⁵⁻²⁷. Observed changes in several other proteins such as LEP, SELL, SELE, and ROR-1, were in agreement with existing literature^{21,28-30}. Other biomarkers including IL-24, IL-22, CCL23, and HIPK3 were also identified as highly predictive. In early pregnancy, FGF19 and TIMP2 were the most predictive. Univariate analysis also identified IL1RAP and IL6 - features known to play a role in immune response. Some of the other known biomarkers of preeclampsia – sFLT-1, PAPP-A, PIGF and ENG – were not significantly different between the controls and preeclamptic women neither over gestation, nor early in pregnancy (Fig. S11) and

were, consequentially, not identified by our prediction model. This is possibly due to a small size of our cohort. We did not have PP13 (Galectin13) measurements, another known biomarker of preeclampsia.

Preeclampsia is accompanied by a dysregulated maternal immune adaptation to pregnancy, which is already detectable in early pregnancy^{14,15}. This aberrant signature was previously identified in women who developed preeclampsia later on¹⁵. Here we report that the intricate functional capacities of immune cells are co-evolving with their environment throughout the course of pregnancy, showing that top informative immune feature levels are highly correlated with top informative plasma protein levels. This interconnectedness supports both prediction approaches, confirming their individual usefulness, while complementing each other's validity. Specifically, the results highlight known pathology of preeclampsia and suggest novel associations between immunological and proteomic dynamics. In preeclamptic pregnancies, immune responses were uniquely correlated with levels of leptin and soluble L-Selectin.

Leptin, known to be elevated in the plasma of pre-eclamptic women³¹, is an immune regulatory hormone produced by adipose tissue and by the placenta^{31,32}. Leptin activates the JAK/STAT and MAPK pathway, directly through binding to the leptin receptor expressed on leukocytes and thereby modulates both innate and adaptive immune responses^{33,34}, including skewing of CD4 T cells towards Th1 polarization³⁵ and inhibiting Treg proliferation³⁴. Accordingly, we observed that leptin levels in preeclamptic and control pregnancies correlated with STAT and MAPK pathway signaling both in innate and adaptive immune cells, suggesting that dysregulated leptin levels in preeclamptic pregnancies might contribute to the aberrant

immune signature, while, reciprocally, inflammation itself might enhance plasma leptin levels^{32,33}. Moreover, while in healthy pregnancies leptin levels correlated with pP38 signaling in Treg and TCRgd, this correlation was lost in preeclamptic pregnancies, suggesting that regulation of immune tolerance might be disrupted in preeclamptic pregnancies.

Further, we reported decreased L-Selectin levels in preeclamptic pregnancies that correlated with basal pSTAT, pNFkB and pMAPKAP2 signaling in innate (mDC and cMC) and adaptive immune cells (Th1 and naive CD4 T cells). L-Selectin is shed from leukocytes during activation and migration, and soluble L-Selectin can be used as a surrogate marker for inflammation³⁶. Notably, a drop in soluble L-Selectin levels is observed during sepsis³⁷. Previous studies reported conflicting results for circulating sL-Selectin levels in preeclampsia³⁸⁻⁴⁰, including low sL-Selectin levels at 20 weeks of gestation, prior to onset of preeclampsia³⁹. Pre-eclampsia-associated enhanced ectodomain shedding of cell adhesion molecules could be directly linked to changes in signaling responses in circulating immune cells by shedding-mediated activation of intracellular pathways³⁶. Alternatively, the correlation could reflect independent inflammatory mechanisms as decreased levels of circulating L-Selectin have been proposed to be due to its adsorption to luminal vascular ligands, which are upregulated by an activated endothelium, a feature of pre-eclampsia^{6,39,41}.

The model with urine metabolites was the most performant to predict PE condition.

Enrichment analysis identified discriminant biological pathways associated with PE when considering early and all time points. Steroid hormone biosynthesis pathway was significant ($p < 0.05$) in both models while arachidonic acid metabolism was significant in early

pregnancy. Arachidonic acid (AA) is a precursor to a myriad of bioactive lipids including prostaglandins (PG), prostacyclin, thromboxane, HPETE, leukotrienes, lipoxins, hypoxins, anandamide, and epoxyeicosatrienoic acids, that play key roles in inflammatory, vascular and coagulation processes⁴². As early as the 1960's the role of the eicosanoids in preeclampsia pathogenesis was proposed and by the 1970's evidence supported that an increase in thromboxane (TXA₂; produced by platelets) over prostacyclin (PGI₂; produced by endothelium) associated with preeclampsia⁴³. This is one of the biological underpinnings for the use of low-dose aspirin for the prevention of preeclampsia. Mills et al.⁴⁴ reported longitudinal measurements of the urinary metabolites of thromboxane and prostacyclin throughout gestation. Although they did not find a significant increase in the urinary concentrations of TXA₂, they did find a significant decrease in PGI₂ as early as 13–16 weeks of gestation and a significant elevation in the ratio of thromboxane to prostacyclin as early as 17–20 weeks of gestation in women destined to develop preeclampsia. While this PG imbalance is noted both prior to and at the time of clinical presentation (after 20 weeks), the fact that AA metabolism was only observed in early pregnancy may explain why clinical studies note that low-dose aspirin initiation prior to 16 weeks is needed for significant prevention of preeclampsia⁴⁵.

Tryptophan pathway was identified as highly associated with preeclampsia over gestation (Fig 10A). Indoleamine-2,3-dioxygenase (IDO) is the first and rate limiting enzyme in this pathway producing kynurenine which then is converted into a number of bioactive metabolites. IDO is an intracellular enzyme produced by many cell types and while not secreted, impacts neighboring cells by tryptophan depletion and production of bioactive metabolites. The role of IDO in both normal and abnormal pregnancies, including preeclampsia, has been recently

reviewed⁴⁶. IDO expression increases with pregnancy and tryptophan depletion in the placenta inhibits T- cell- mediated rejection of semiallogeneic fetal tissues⁴⁷. Kynurenine is an endogenous ligand that activates the aryl hydrocarbon receptor (AhR)⁴⁸. This activation skews the differentiation of T cells to immunosuppressive T regulatory cells rather than proinflammatory Th17 cells after exposure to TGF- β ^{49,50}. Notably, kynurenic acid and xanthurenic acid, two metabolites of kynurenine, can also activate AhR signaling and may participate in immune regulation^{51,52}. Therefore, deficiency of IDO impacts Treg development. Notably, IDO KO mice, when pregnant, develop a preeclampsia-like phenotype⁵³. The metabolic signal related to tryptophan metabolism in the model over gestation may be related to the immune signature of preeclampsia, highlighting the importance of the immune alteration occurring in the later stages of preeclampsia. Caffeine metabolism was also identified as highly associated with preeclampsia over gestation. This pathway has previously been associated to pregnancy progression¹⁹.

Models to predict preeclampsia early in pregnancy were previously based on maternal characteristics (demographics and medical history), followed by addition of uterine artery Doppler measurements and specific biomarkers⁵⁴⁻⁵⁹. Levels of angiogenic and/or anti-angiogenic proteins (PlGF, sFlt-1, and endoglin), or their ratios, have been established as biomarkers with high prediction accuracy later in pregnancy^{25,26,60}. More recently, analysis of omics datasets have been successfully applied to identify various biomarkers related to preeclampsia^{10,23,61}. Most of these studies were based on measurements from one or at most two omics datasets, and often from samples taken only at one time point during pregnancy.

Here we show that clinical and demographics characteristics (i.e., weight, height, race) were complementary to omics measurements and improved prediction models

Our study is limited by a small sample size and consideration of a cohort from a single hospital. Inherently to machine learning approach, developing a prediction model depends on the underlying sample distribution of the data which is used. Distribution shift, caused by differences among various cohorts, can impact the performance of a machine learning algorithm⁶². In this study, the mass cytometry data was not included in the multiomics prediction model because this data was not available for 14 out of 33 patients. However, integrative analysis on the restricted set of common samples revealed important connections between our model and key immune features.

While encouraging, our results need to be validated on a larger, more diverse set of patients. If the results prove generalizable, our findings demonstrating high predictive power from a small number of urine metabolites and proteins could lead to a simple prediction test based on a small number of urine metabolites suitable for use both in developed and developing parts of the world.

METHODS

Study Design

We performed a longitudinal, prospective study of a cohort of pregnant women receiving routine ante- and post-partum care at the Lucile Packard Children's Hospital at Stanford University, California, as previously described^{15,63}. Women were eligible for the study if they

were at least 18 years of age and were in their first trimester of pregnancy. The study was approved by the Institutional Review Board of Stanford University (#21956), and all participants signed an informed consent.

Peripheral blood samples (for mass cytometry analysis), plasma samples (for proteomic, cell-free transcriptomic (cfRNA), metabolomic, and lipidomic analyses), urine samples (for metabolomics analysis), and vaginal swabs (for microbiome analysis) were collected from each woman at two or three time points during pregnancy. Sample collection and their analyses were previously described⁹ and are presented in the Supplemental Materials. The validation cohort included 16 women from the same hospital, for which longitudinal samples with only metabolomic analyses were available. Metabolomic analyses were performed following the same methodology as for the discovery cohort.

Definition of Preeclampsia

Preeclampsia was defined using the American College of Obstetrics and Gynecology classification³ as follows: hypertension that develops after 20 weeks of gestation (systolic or diastolic blood pressure 140 mm Hg and/or 90 mm Hg, respectively, measured on at least two occasions, 4 hours to 1 week apart) and proteinuria (300 mg in a 24-hour urine collection, a protein/creatinine ratio of at least 0.3 (each measured as mg/dL) or if these were not readily available a random urine specimens containing 1+ protein by dipstick). In the absence of proteinuria preeclampsia was diagnosed if the presence of thrombocytopenia (platelet count less than 100,000/microliter), impaired liver function (elevated blood levels of liver transaminases to twice the normal concentration), the new development of renal insufficiency

(elevated serum creatinine greater than 1.1 mg/dL), pulmonary edema, or new-onset cerebral or visual disturbances. Early-onset and late-onset preeclampsia were distinguished based on whether diagnosis was before or after 34 weeks of gestation.

Machine learning analyses

A two-level cross-validation approach was used to build predictive models to estimate the risk of preeclampsia. At the first level, prediction models were developed for each omics set using an elastic net (EN) model⁶⁴. Given $N \times p$ matrix of predictors (measurements) $X = (x_1, \dots, x_p)$ and a vector of responses $y = (y_1, \dots, y_N)$, regression coefficients $\beta = (\beta_1, \dots, \beta_p)$ and an intercept term β_0 in the EN model are obtained by maximizing the likelihood, or equivalently minimizing the negative log-likelihood together with L_1 and L_2 penalty:

$$\left[\frac{1}{N} \sum_{i=1}^N L(\beta_0, \beta; y, X) + \lambda((1 - \alpha)\|\beta\|_2 + \alpha\|\beta\|) \right]$$

Logistic regression was used, for which the negative log-likelihood evaluates to:

$$L(\beta_0, \beta; y, X) = \sum_{i=1}^N y_i(\beta_0 + x_i^T \beta) - \log(1 + e^{\beta_0 + x_i^T \beta})$$

For the high-dimensional setting ($p \gg N$) considered here, EN, which performs both shrinkage and automatic selection of predictors, can provide both high accuracy and facilitate interpretability.

At the second level, predictions of EN models were integrated using stacked regression^{65–67}.

Specifically, in order to use EN models in the two-level approach, for each modality k , $k = 1, \dots, K$ and data $X^k = (x_1^k, \dots, x_{p_k}^k)$ a leave-one-out EN model, denoted $c_{-i}^k(x_i)$ was repeatedly

fitted and evaluated at patient i . At the second level, stacked regression with nonnegative coefficients¹¹ was used, so that the regression coefficients of the final model $(\gamma_1, \dots, \gamma_K)$ were determined by:

$$\min \sum_{i=1}^N \left(y_i - \sum_{k=1}^K \gamma_k c_{-i}^k(x_i^k) \right)^2 \text{ s. t. } \gamma_i \geq 0.$$

Note that the leave-one-out approach used in stacked regression has a purpose to form an unbiased linear combination of EN models⁶⁶. In contrast to the original stacking approach in which different prediction models fit on the same data are stacked, here, we use the same model (EN) but fit to different omics to obtain different estimators which are then stacked. A stacked regression model can be regarded as a special case of a two-layer neural network; its special construction provides for an easier interpretation.

One of our main goals was to identify a small subset of specific biomarkers that can predict preeclampsia with high accuracy and could thereby be used as a simple diagnostic test. For these reasons, performance of the refitted EN model for each omics set was evaluated by treating the EN model as a model-selection procedure and performing a refitting step on the selected support, in the same cross-validation step⁶⁸. It is known that L_1 -penalization used in EN performs excessive shrinkage of the large coefficients of the prediction model⁶⁹. Refitting can resolve this problem and obtain a model with a smaller number of features. Finally, to investigate a possible gain from integration of available clinical and demographics characteristics, a prediction model that takes omics (from a specific multiomics set), and clinical and demographics variables as an input to an EN model was fit and evaluated.

Performance was estimated using a leave-one-out cross-validation procedure, such that in each cross-validation step all measurements of one patient are left out from training set and are used for testing. In addition, urine metabolome prediction models, with and without clinical/demographics variables were validated on a separate validation cohort. The prediction accuracy of the model in terms of the area under receiver operating characteristics curve was evaluated. For the network visualization, a k-nearest neighbor graph (with $k = 2$), was constructed between features. The network layout was computed with the LargeVis algorithm⁷⁰. The analysis was performed using R software (version 3.6.1).

Pathway Enrichment Analysis

Univariate analysis was performed to identify features with significant associations between each feature and the pregnancy outcome, both in early pregnancy (Wilcoxon signed-rank test) and over gestation (Linear Mixed-effects Model). The Benjamini-Hochberg procedure was used to control the false discovery rate (FDR)⁷¹. Metabolome pathway enrichment analysis on identified metabolites was performed using MetaboAnalyst⁷². The hypergeometric test was used for overrepresentation analysis in MetaboAnalyst. Proteome pathway enrichment analysis was performed using GeneOntology^{73,74}. Circular Gene Ontology (CirGO) software for visualizing two-level hierarchically structured gene ontology terms⁷⁵, was used to visualize proteome and transcriptome pathway enrichment.

ACKNOWLEDGEMENTS:

This study was supported by the March of Dimes Prematurity Research Center at Stanford University School of Medicine, Stanford Maternal & Child Health Research Institute, the

Christopher Hess Research Fund, the National Institutes of Health, grants 1R01HL139844 and R35GM138353, Burroughs Wellcome Fund, and grants from the Bill & Melinda Gates Foundation OPP1112382 and OPP1113682. M.P.S. was supported by National Institutes of Health, grant 5RM1HG00773507. D.A.R. was supported by the Thomas C. and Joan M. Merigan Endowment at Stanford University and the Chan Zuckerberg Biohub Microbiome Initiative.

AUTHOR CONTRIBUTIONS:

I.M. and N.A. designed machine learning experiments and performed interpretation of results; I.M. performed computational analysis and wrote the paper; N.A. provided guidance and feedback and contributed to writing the paper; K.C. performed MS metabolomic analysis and contributed to metabolomic pathway analysis and writing of the paper; M.M. performed cell-free RNA transcriptome analysis; I.S., D.F., and X.H. performed immunome analysis and contributed to writing the paper; A.T. performed part of the computational analysis; N.S. assisted with the computational analysis; R.J.W. supervised the study data collection and edited the manuscript; G.M.T and M.E. performed MS metabolomic analysis; N.S., A.C., R.F., H.N., M.B., M.X., C.E., D. D. F., M.S.G., A.Cu. helped with editing the manuscript; E.C. performed microbiome analysis; X.B.L and K.G.S., G.L.D., D.A.R., S.R.Q., M.A. provided critical feedback on the study and edited the manuscript; D.A.R. was involved with the microbiome analysis; M.P.S. was involved in the MS metabolome analysis; V.D.W. wrote a part of the paper and contributed to interpretation of results; G.M.S. provided the study data, provided guidance and feedback, contributed to interpretation of results and edited the manuscript; B.G. designed the immunome part of the analysis, performed interpretation of the results and edited the

manuscript; D.K.S. provided the study data, provided guidance and supervised the research. All authors discussed results, and contributed to the final manuscript.

COMPETING INTERESTS: There are no competing interests.

Bibliography

1. WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division. Maternal mortality: Levels and trends 2000 to 2017. (2019).
2. Duley, L. The global impact of pre-eclampsia and eclampsia. *Semin Perinatol* **33**, 130–137 (2009).
3. Jeyabalan, A. Epidemiology of preeclampsia: impact of obesity. *Nutr. Rev.* **71 Suppl 1**, S18–25 (2013).
4. Than, N. G. *et al.* Integrated systems biology approach identifies novel maternal and placental pathways of preeclampsia. *Front. Immunol.* **9**, 1661 (2018).
5. Phipps, E. A., Thadhani, R., Benzinger, T. & Karumanchi, S. A. Pre-eclampsia: pathogenesis, novel diagnostics and therapies. *Nat. Rev. Nephrol.* **15**, 275–289 (2019).
6. Chaiworapongsa, T., Chaemsathong, P., Yeo, L. & Romero, R. Pre-eclampsia part 1: current understanding of its pathophysiology. *Nat. Rev. Nephrol.* **10**, 466–480 (2014).
7. Duckitt, K. & Harrington, D. Risk factors for pre-eclampsia at antenatal booking: systematic review of controlled studies. *BMJ* **330**, 565 (2005).
8. Tranquilli, A. L., Brown, M. A., Zeeman, G. G., Dekker, G. & Sibai, B. M. The definition of severe and early-onset preeclampsia. Statements from the International Society for the Study of Hypertension in Pregnancy (ISSHP). *Pregnancy hypertension* **3**, 44–47 (2013).

9. Ghaemi, M. S. *et al.* Multiomics modeling of the immunome, transcriptome, microbiome, proteome and metabolome adaptations during human pregnancy. *Bioinformatics* **35**, 95–103 (2019).
10. Benny, P. A., Alakwaa, F. M., Schlueter, R. J., Lassiter, C. B. & Garmire, L. X. A review of omics approaches to study preeclampsia. *Placenta* **92**, 17–27 (2020).
11. Meinshausen, N. & Bühlmann, P. Stability selection. *J Royal Statistical Soc B* **72**, 417–473 (2010).
12. Roberge, S. *et al.* The role of aspirin dose on the prevention of preeclampsia and fetal growth restriction: systematic review and meta-analysis. *Am. J. Obstet. Gynecol.* **216**, 110–120.e6 (2017).
13. American College of Obstetricians and Gynecologists; Task Force on Hypertension in Pregnancy. Hypertension in pregnancy. Report of the American College of Obstetricians and Gynecologists' Task Force on Hypertension in Pregnancy. *Obstet Gynecol.* (2013).
14. Redman, C. W. G. & Sargent, I. L. Immunology of pre-eclampsia. *Am J Reprod Immunol* **63**, 534–543 (2010).
15. Han, X. *et al.* Differential dynamics of the maternal immune system in healthy pregnancy and preeclampsia. *Front. Immunol.* **10**, 1305 (2019).
16. Samolis, S. *et al.* Relation between first trimester maternal serum leptin levels and body mass index in normotensive and pre-eclamptic pregnancies--role of leptin as a marker of pre-eclampsia: a prospective case-control study. *Gynecol Endocrinol* **26**, 338–343 (2010).
17. Hashimoto, M. *et al.* The effects of age on the release of adenine nucleosides and nucleotides from rat caudal artery. *J. Physiol. (Lond.)* **489** (Pt 3), 841–848 (1995).
18. Chatuphonprasert, W., Jarukamjorn, K. & Ellinger, I. Physiology and pathophysiology of steroid biosynthesis, transport and metabolism in the human placenta. *Front. Pharmacol.* **9**, 1027 (2018).

19. Liang, L. *et al.* Metabolic dynamics and prediction of gestational age and time to delivery in pregnant women. *Cell* **181**, 1680–1692.e15 (2020).
20. Nilsen, R. M. *et al.* Maternal tryptophan and kynurenine pathway metabolites and risk of preeclampsia. *Obstet. Gynecol.* **119**, 1243–1250 (2012).
21. Luppi, P. *et al.* Preeclampsia activates circulating immune cells with engagement of the NF-kappaB pathway. *Am J Reprod Immunol* **56**, 135–144 (2006).
22. Massobrio, M. *et al.* Arachidonic Acid Derivatives in the Pathophysiology of Pregnancy-Induced Hypertension. *Clinical and Experimental Hypertension. Part B: Hypertension in Pregnancy* **7**, 43–55 (1988).
23. Tarca, A. L. *et al.* The prediction of early preeclampsia: Results from a longitudinal proteomics study. *PLoS One* **14**, e0217273 (2019).
24. Austdal, M. *et al.* First trimester urine and serum metabolomics for prediction of preeclampsia and gestational hypertension: A prospective screening study. *Int. J. Mol. Sci.* **16**, 21520–21538 (2015).
25. Maynard, S. E. & Karumanchi, S. A. Angiogenic factors and preeclampsia. *Semin Nephrol* **31**, 33–46 (2011).
26. Maynard, S. E. *et al.* Excess placental soluble fms-like tyrosine kinase 1 (sFlt1) may contribute to endothelial dysfunction, hypertension, and proteinuria in preeclampsia. *J. Clin. Invest.* **111**, 649–658 (2003).
27. Rath, G. & Tripathi, R. Angiogenic balance and diagnosis of pre-eclampsia: selecting the right VEGF receptor. *J Hum Hypertens* **26**, 207–210 (2012).
28. Poston, L. Leptin and preeclampsia. *Semin Reprod Med* **20**, 131–138 (2002).
29. Chen, J. *et al.* Downregulation of receptor tyrosine kinase-like orphan receptor 1 in preeclampsia placenta inhibits human trophoblast cell proliferation, migration, and invasion by PI3K/AKT/mTOR pathway accommodation. *Placenta* **82**, 17–24 (2019).

30. Hao, S. *et al.* Changes in pregnancy-related serum biomarkers early in gestation are associated with later development of preeclampsia. *PLoS One* **15**, e0230000 (2020).
31. Taylor, B. D. *et al.* Serum leptin measured in early pregnancy is higher in women with preeclampsia compared with normotensive pregnant women. *Hypertension* **65**, 594–599 (2015).
32. Pérez-Pérez, A. *et al.* Leptin action in normal and pathological pregnancies. *J. Cell Mol. Med.* **22**, 716–727 (2018).
33. Naylor, C. & Petri, W. A. Leptin regulation of immune responses. *Trends Mol. Med.* **22**, 88–98 (2016).
34. Abella, V. *et al.* Leptin in the interplay of inflammation, metabolism and immune system disorders. *Nat. Rev. Rheumatol.* **13**, 100–109 (2017).
35. Martín-Romero, C., Santos-Alvarez, J., Goberna, R. & Sánchez-Margalet, V. Human leptin enhances activation and proliferation of human circulating T lymphocytes. *Cell Immunol.* **199**, 15–24 (2000).
36. Ivetic, A., Hoskins Green, H. L. & Hart, S. J. L-selectin: A Major Regulator of Leukocyte Adhesion, Migration and Signaling. *Front. Immunol.* **10**, 1068 (2019).
37. Seidelin, J. B., Nielsen, O. H. & Strøm, J. Soluble L-selectin levels predict survival in sepsis. *Intensive Care Med.* **28**, 1613–1618 (2002).
38. Docheva, N. *et al.* The profiles of soluble adhesion molecules in the “great obstetrical syndromes”. *J. Matern. Fetal Neonatal Med.* **32**, 2113–2136 (2019).
39. Chavarría, M. E., Lara-González, L., García-Paleta, Y., Vital-Reyes, V. S. & Reyes, A. Adhesion molecules changes at 20 gestation weeks in pregnancies complicated by preeclampsia. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **137**, 157–164 (2008).
40. Sabatier, F. *et al.* Neutrophil activation in preeclampsia and isolated intrauterine growth restriction. *Am. J. Obstet. Gynecol.* **183**, 1558–1563 (2000).

41. Rainer, T. H. L-selectin in health and disease. *Resuscitation* **52**, 127–141 (2002).
42. Sonnweber, T., Pizzini, A., Nairz, M., Weiss, G. & Tancevski, I. Arachidonic acid metabolites in cardiovascular and metabolic diseases. *Int. J. Mol. Sci.* **19**, (2018).
43. Walsh, S. W. Eicosanoids in preeclampsia. *Prostaglandins Leukot Essent Fatty Acids* **70**, 223–232 (2004).
44. Mills, J. L. *et al.* Prostacyclin and thromboxane changes predating clinical onset of preeclampsia: a multicenter prospective study. *JAMA* **282**, 356–362 (1999).
45. Cui, Y., Zhu, B. & Zheng, F. Low-dose aspirin at ≤ 16 weeks of gestation for preventing preeclampsia and its maternal and neonatal adverse outcomes: A systematic review and meta-analysis. *Exp Ther Med* **15**, 4361–4369 (2018).
46. Chang, R.-Q., Li, D.-J. & Li, M.-Q. The role of indoleamine-2,3-dioxygenase in normal and pathological pregnancies. *Am J Reprod Immunol* **79**, e12786 (2018).
47. Munn, D. H. *et al.* Prevention of allogeneic fetal rejection by tryptophan catabolism. *Science* **281**, 1191–1193 (1998).
48. DiNatale, B. C. *et al.* Kynurenic acid is a potent endogenous aryl hydrocarbon receptor ligand that synergistically induces interleukin-6 in the presence of inflammatory signaling. *Toxicol. Sci.* **115**, 89–97 (2010).
49. Nguyen, N. T. *et al.* Aryl hydrocarbon receptor negatively regulates dendritic cell immunogenicity via a kynurenine-dependent mechanism. *Proc. Natl. Acad. Sci. USA* **107**, 19961–19966 (2010).
50. Mezrich, J. D. *et al.* An interaction between kynurenine and the aryl hydrocarbon receptor can generate regulatory T cells. *J. Immunol.* **185**, 3190–3198 (2010).
51. Jaronen, M. & Quintana, F. J. Immunological relevance of the coevolution of IDO1 and AHR. *Front. Immunol.* **5**, 521 (2014).
52. Fazio, F. *et al.* Cinnabarinic acid and xanthurenic acid: Two kynurenine metabolites that

- interact with metabotropic glutamate receptors. *Neuropharmacology* **112**, 365–372 (2017).
53. Santillan, M. K. *et al.* Pregnant mice lacking indoleamine 2,3-dioxygenase exhibit preeclampsia phenotypes. *Physiol. Rep.* **3**, (2015).
 54. Wright, D., Syngelaki, A., Akolekar, R., Poon, L. C. & Nicolaides, K. H. Competing risks model in screening for preeclampsia by maternal characteristics and medical history. *Am. J. Obstet. Gynecol.* **213**, 62.e1-62.e10 (2015).
 55. Odibo, A. O. *et al.* First-trimester placental protein 13, PAPP-A, uterine artery Doppler and maternal characteristics in the prediction of pre-eclampsia. *Placenta* **32**, 598–602 (2011).
 56. Yu, C. K. H. *et al.* An integrated model for the prediction of preeclampsia using maternal factors and uterine artery Doppler velocimetry in unselected low-risk women. *Am. J. Obstet. Gynecol.* **193**, 429–436 (2005).
 57. Audibert, F. *et al.* Screening for preeclampsia using first-trimester serum markers and uterine artery Doppler in nulliparous women. *Am. J. Obstet. Gynecol.* **203**, 383.e1-8 (2010).
 58. Wright, D., Wright, A. & Nicolaides, K. H. The competing risk approach for prediction of preeclampsia. *Am. J. Obstet. Gynecol.* (2019). doi:10.1016/j.ajog.2019.11.1247
 59. North, R. A. *et al.* Clinical risk prediction for pre-eclampsia in nulliparous women: development of model in international prospective cohort. *BMJ* **342**, d1875 (2011).
 60. Parra-Cordero, M. *et al.* Prediction of early and late pre-eclampsia from maternal characteristics, uterine artery Doppler and markers of vasculogenesis during first trimester of pregnancy. *Ultrasound Obstet. Gynecol.* **41**, 538–544 (2013).
 61. Kelly, R. S. *et al.* Integration of metabolomic and transcriptomic networks in pregnant women reveals biological pathways and predictive signatures associated with preeclampsia. *Metabolomics* **13**, (2017).
 62. Recht, B., Roelofs, R., Schmidt, L. & Shankar, V. Do ImageNet Classifiers Generalize to ImageNet? *arXiv* (2019).

63. Aghaeepour, N. *et al.* An immune clock of human pregnancy. *Sci. Immunol.* **2**, (2017).
64. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J Royal Statistical Soc B* **67**, 301–320 (2005).
65. Breiman, L. Stacked regressions. *Mach Learn* **24**, 49–64 (1996).
66. Wolpert, D. H. Stacked generalization. *Neural Netw.* **5**, 241–259 (1992).
67. Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society: Series B (Methodological)* **36**, 111–133 (1974).
68. Chzhen, E., Hebiri, M. & Salmon, J. On Lasso refitting strategies. *Bernoulli* **25**, 3175–3200 (2019).
69. Hastie, T., Tibshirani, R. & Wainwright, M. in *Statistical Learning with Sparsity: The Lasso and Generalizations* 155–182 (Chapman and Hall/CRC, 2015). doi:10.1201/b18401-8
70. Tang, J., Liu, J., Zhang, M. & Mei, Q. Visualizing Large-scale and High-dimensional Data. in *Proceedings of the 25th International Conference on World Wide Web - WWW '16* 287–297 (ACM Press, 2016). doi:10.1145/2872427.2883041
71. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300 (1995).
72. Pang, Z., Chong, J., Li, S. & Xia, J. Metaboanalyst 3.0: toward an optimized workflow for global metabolomics. *Metabolites* **10**, (2020).
73. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
74. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
75. Kuznetsova, I., Lugmayr, A., Siira, S. J., Rackham, O. & Filipovska, A. CirGO: an alternative circular way of visualising gene ontology terms. *BMC Bioinformatics* **20**, 84

(2019).

Figures

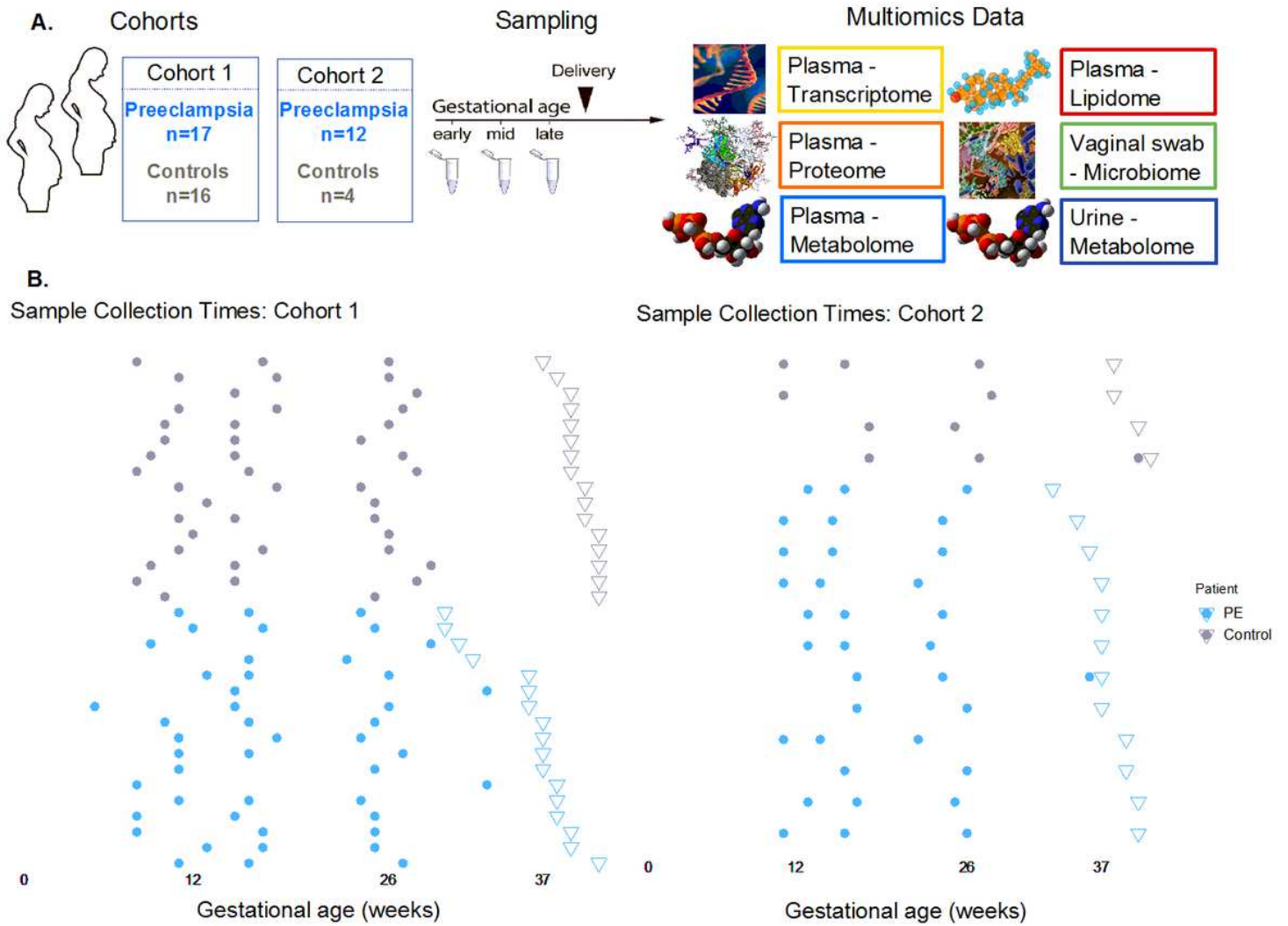


Figure 1

Overview of the study. A. Two independent cohorts were analyzed using 6 different assays. B. Sample collection timeline for plasma in discovery and validation cohorts. Circles indicate pre-delivery sample collection times and inverted triangles indicate delivery dates for individual women (one per horizontal line).

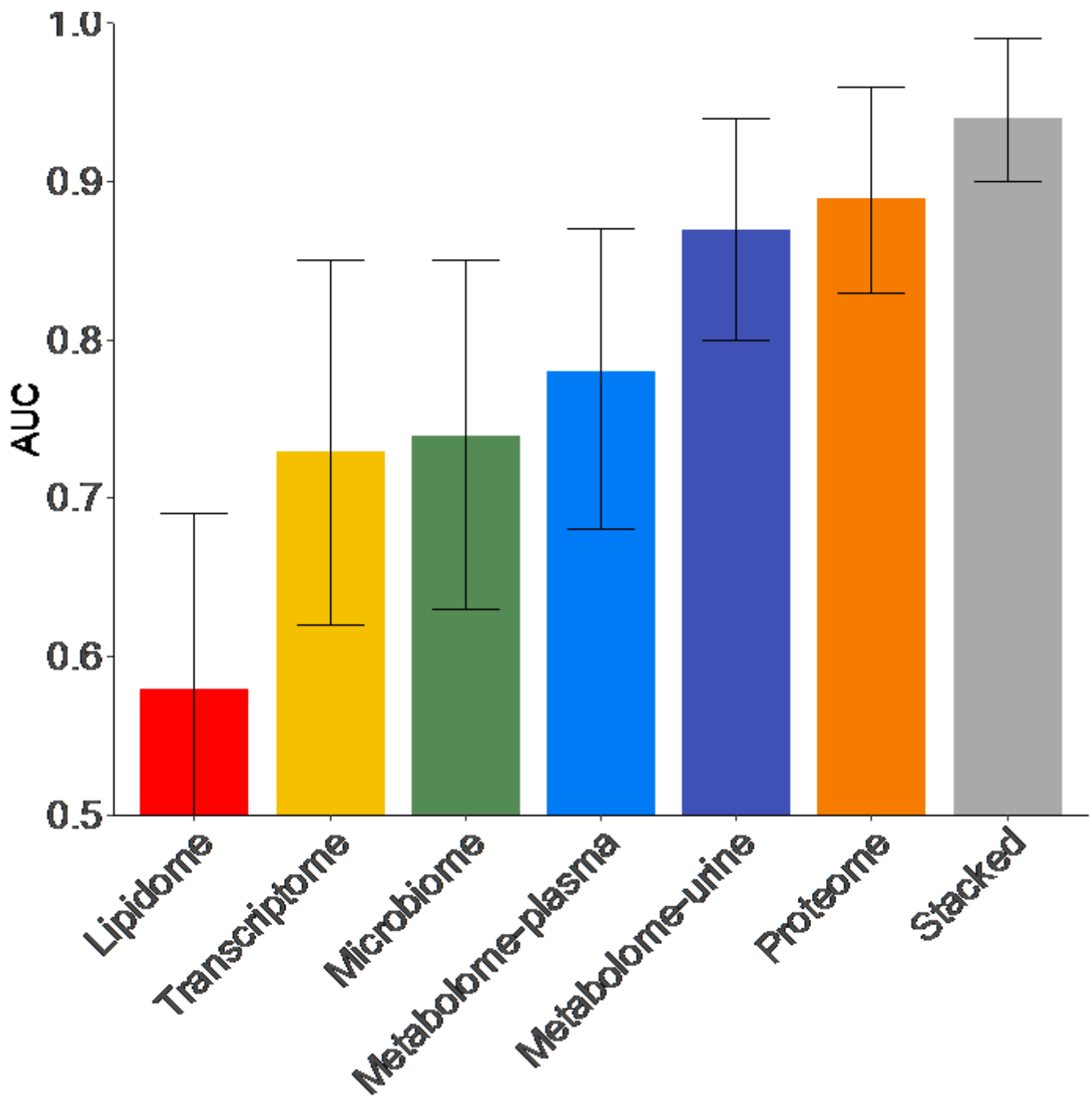


Figure 2

Single- and Multi-omics machine models for preeclampsia. Performance comparison of machine learning models shown on x-axis in terms of the area under the receiver operator curve (AUC) shown on y-axis. The integrated (stacked) model utilizing stacked regression exhibited the highest accuracy (AUC=0.94, 95% CI [0.9, 0.94]). Both proteome and metabolome (urine) had high prediction performance (AUC=0.89, 95% CI [0.83, 0.96] proteome; AUC = 0.87, 95% CI [0.80, 0.94] urine metabolome).

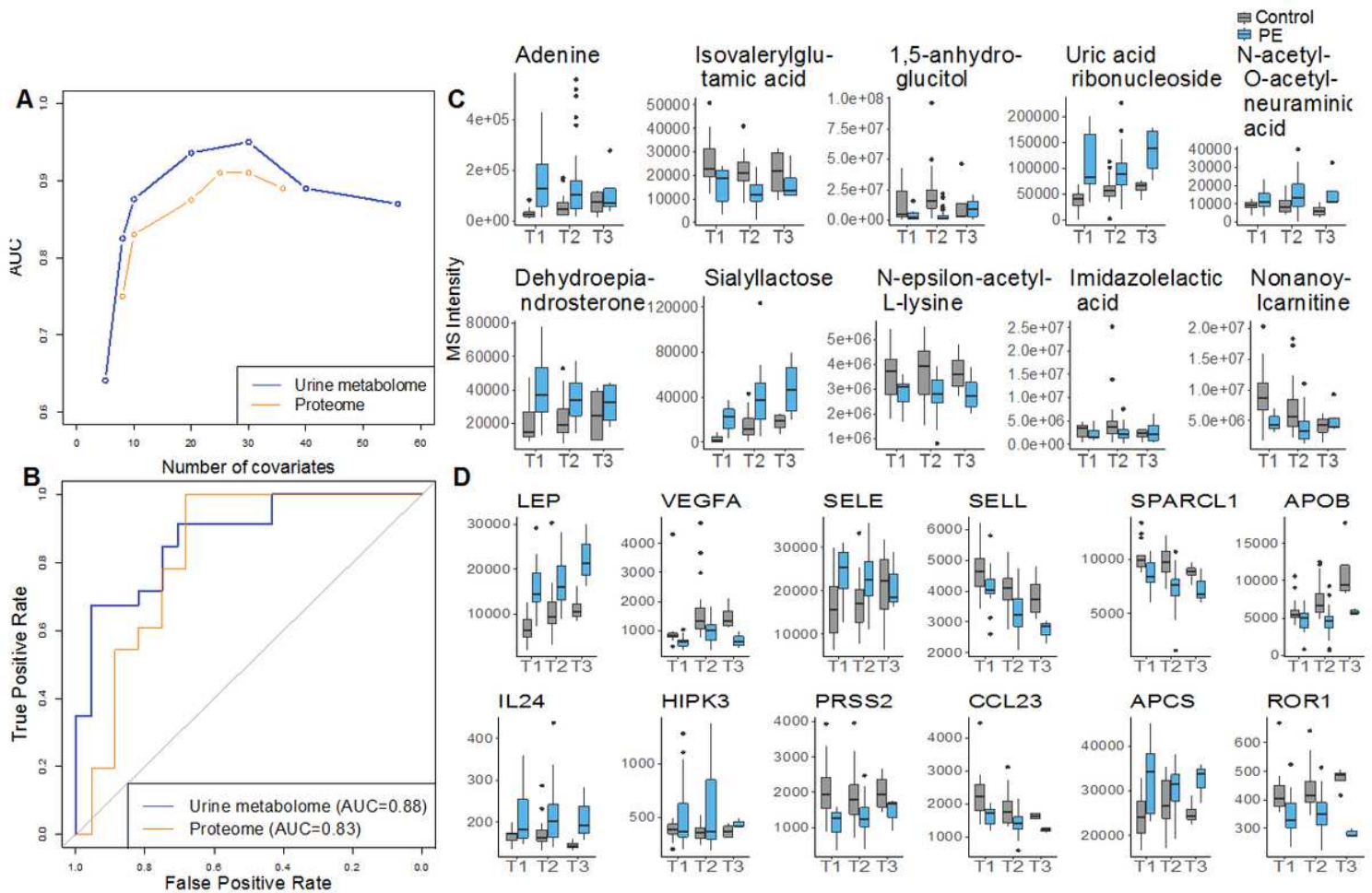


Figure 3

Refitted model for prediction of preeclampsia in resource-limited settings: urine metabolomic and plasma proteomic data sets. A. Cross-validated performance of the refitted models as a function of the number of features (support) for urine metabolome (blue) and proteome (orange). The best performance is observed for a model with 30 features. A model using ten metabolites provided the accuracy of the whole urine metabolomic dataset (AUC=0.88, 95% CI: [0.81, 0.95]). B. Performance of cross-validated prediction model using ten metabolites (blue) or ten proteins (orange). For the urine metabolome, prediction from ten features had the same performance as the prediction from the full metabolomics set (AUC = 0.88, 95% CI [0.81, 0.95]). Prediction from ten proteins yielded performance of AUC = 0.83, 95% CI: [0.73, 0.92]. C. Most informative metabolites. Y-axis shows a metabolite value stratified by normal pregnancy (grey) and preeclamptic pregnancy (blue). D. Most informative proteins. Y-axis shows a protein value stratified by normal pregnancy (grey) and preeclamptic pregnancy (blue).

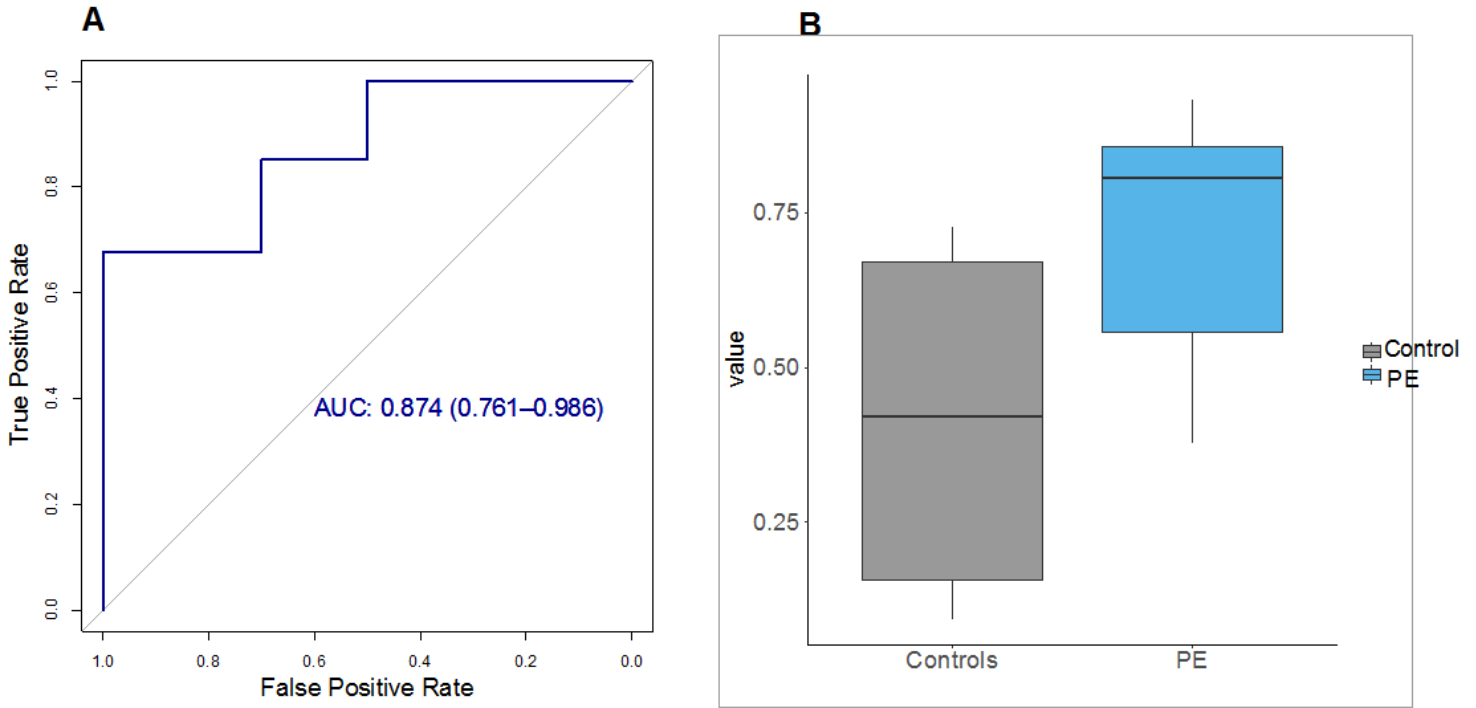


Figure 4

Validated model from urine metabolome using ten metabolites from the validation cohort. A. Area under the receiver operating curve, AUC = 0.874, 95% CI of [0.76, 0.99]. B. Prediction values (scores) for controls and preeclampsia (PE) groups.

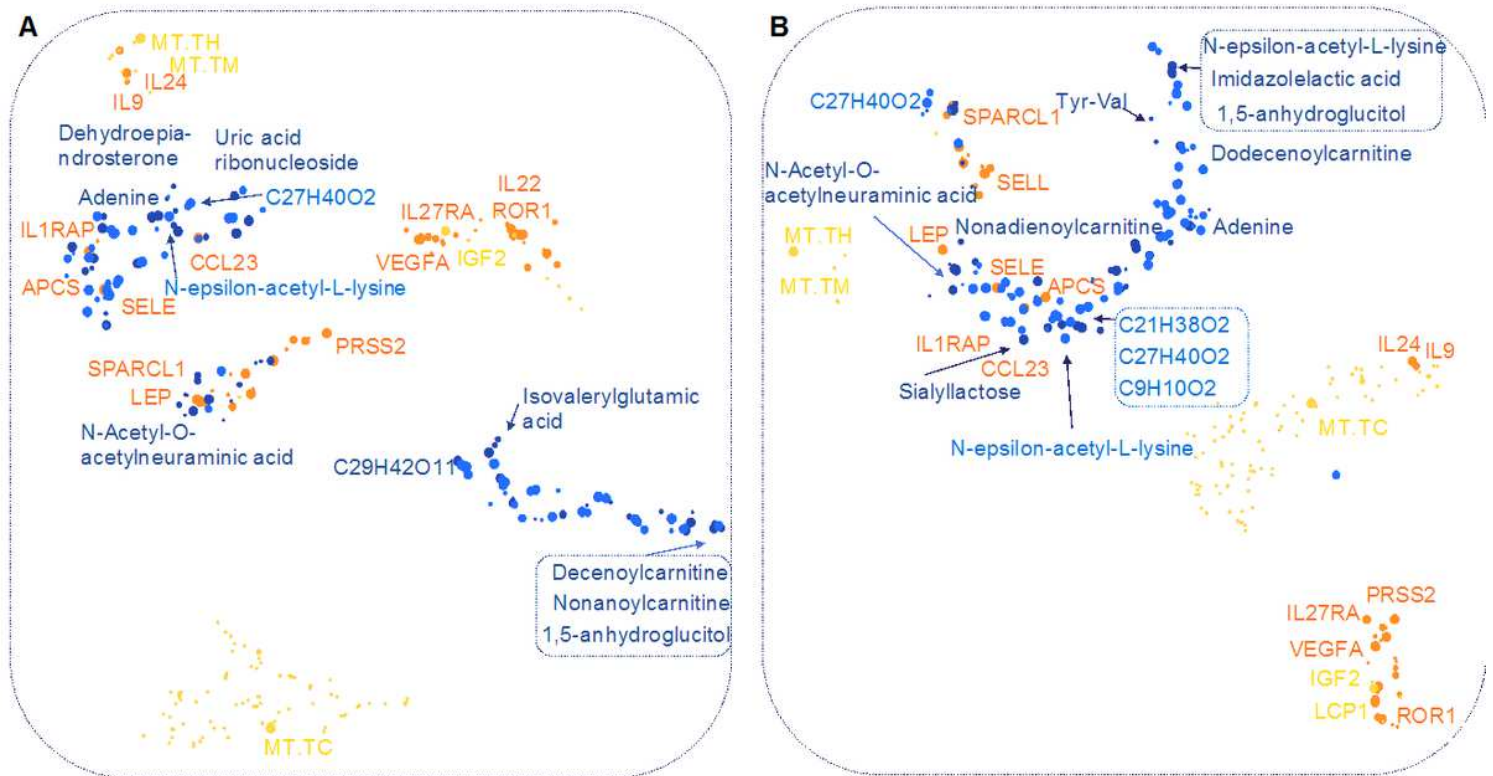


Figure 5

Visualization of predictive features of transcriptome (yellow), proteome (orange), urine metabolome (dark blue), and plasma metabolome (light blue). Size of each node is proportional to the frequency at which it was chosen in prediction models during cross-validation. High frequency of occurrence indicates that a feature is relevant for all or a majority of patients resulting in a more stable model. A. Controls. B. Preeclampsia patients. We observe different clusters for the two groups of patients.

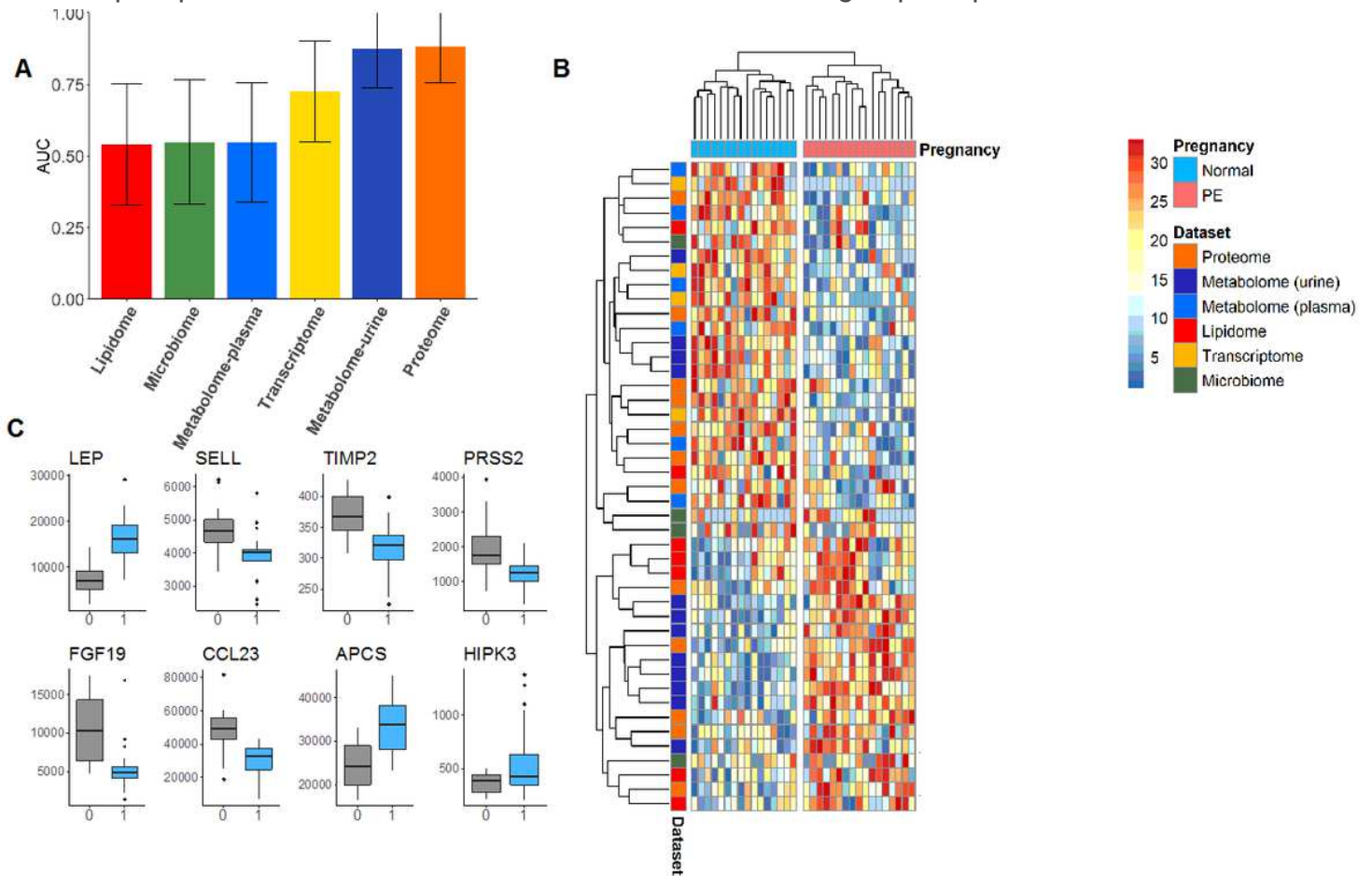


Figure 6

Prediction models in early pregnancy. A. Performance comparison of EN models derived from different omics data sets using the top-ranked ten features, in terms of the AUC. The plasma proteomic and the urine metabolomic models performed best (AUC=0.88, 95% CI of [0.75, 1] for proteome; AUC=0.87, 95% CI of [0.74, 1] for urine metabolome). B. Heatmap of ranked values of features identified by EN, perfectly distinguishing preeclamptic women from controls. C. Top eight proteins. Y-axis shows the value in early pregnancy stratified by normal (grey) versus preeclamptic pregnancy (light-blue).

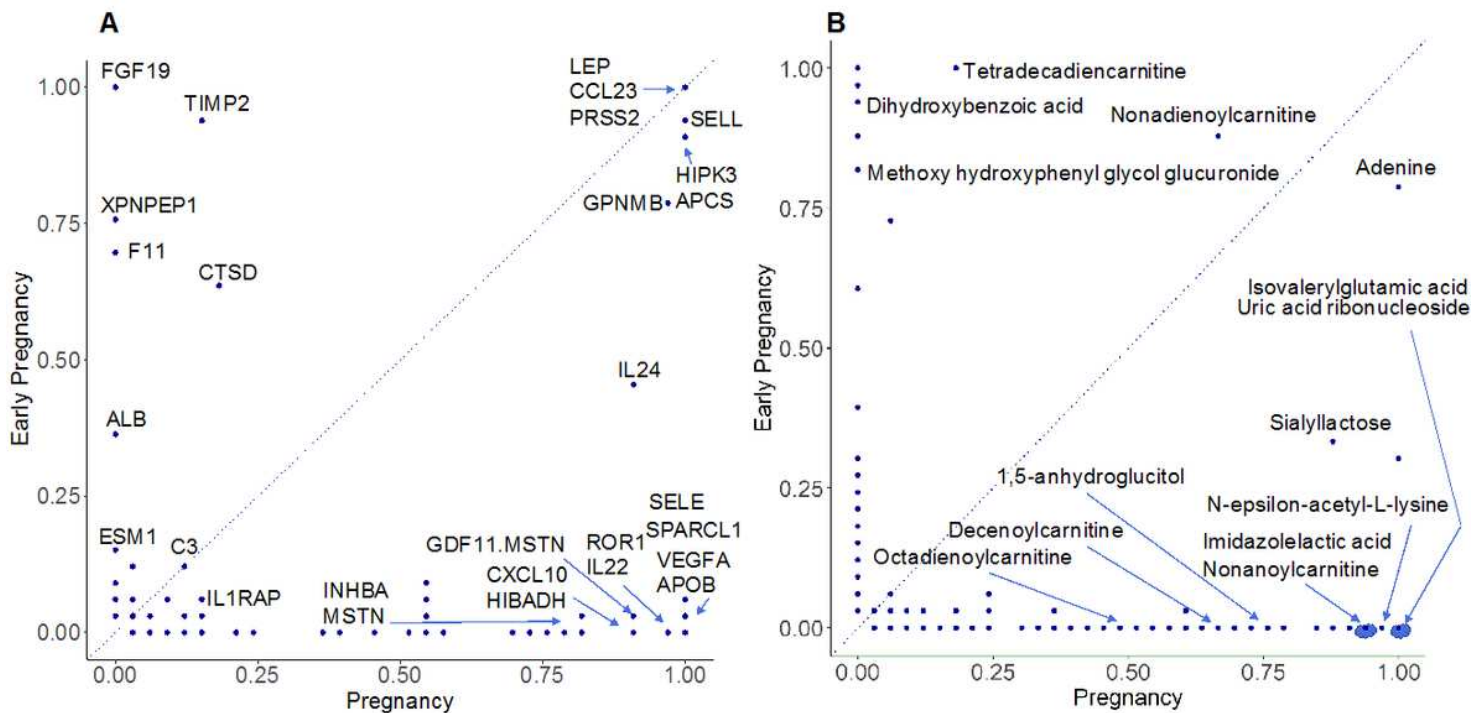


Figure 7

Biomarker comparison: entire pregnancy vs. early pregnancy. X-axis and Y-axis show the respective frequency of each biomarker in early pregnancy and over gestation. A. Most predictive proteins. B. Most predictive urine metabolites. Blue circles around dots imply the same position for more than one protein/urine metabolite.

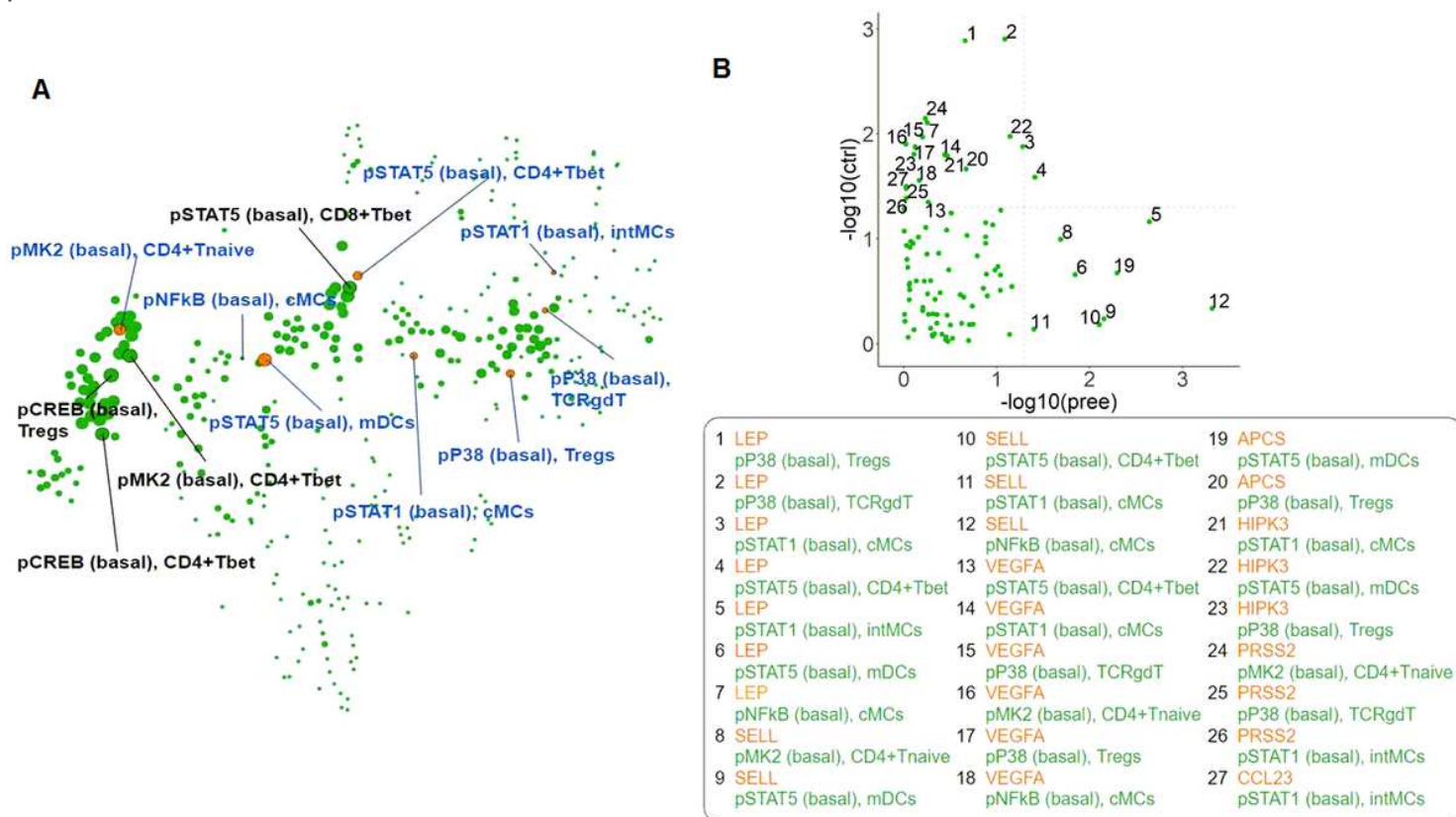


Figure 8

A. Visualization of immune features. Visualization indicates features most correlated with the prediction of the stacked model. Features shown in orange are the seven most predictive immunome features that also highly correlate with the multiomics predictive model. Size of each node is proportional to the $-\log_{10}(\text{p-value})$ of spearman correlation. B. Comparison of p-value of correlation for the top immune and top proteome features. Each node is a pair comprising an immune and a proteome feature.

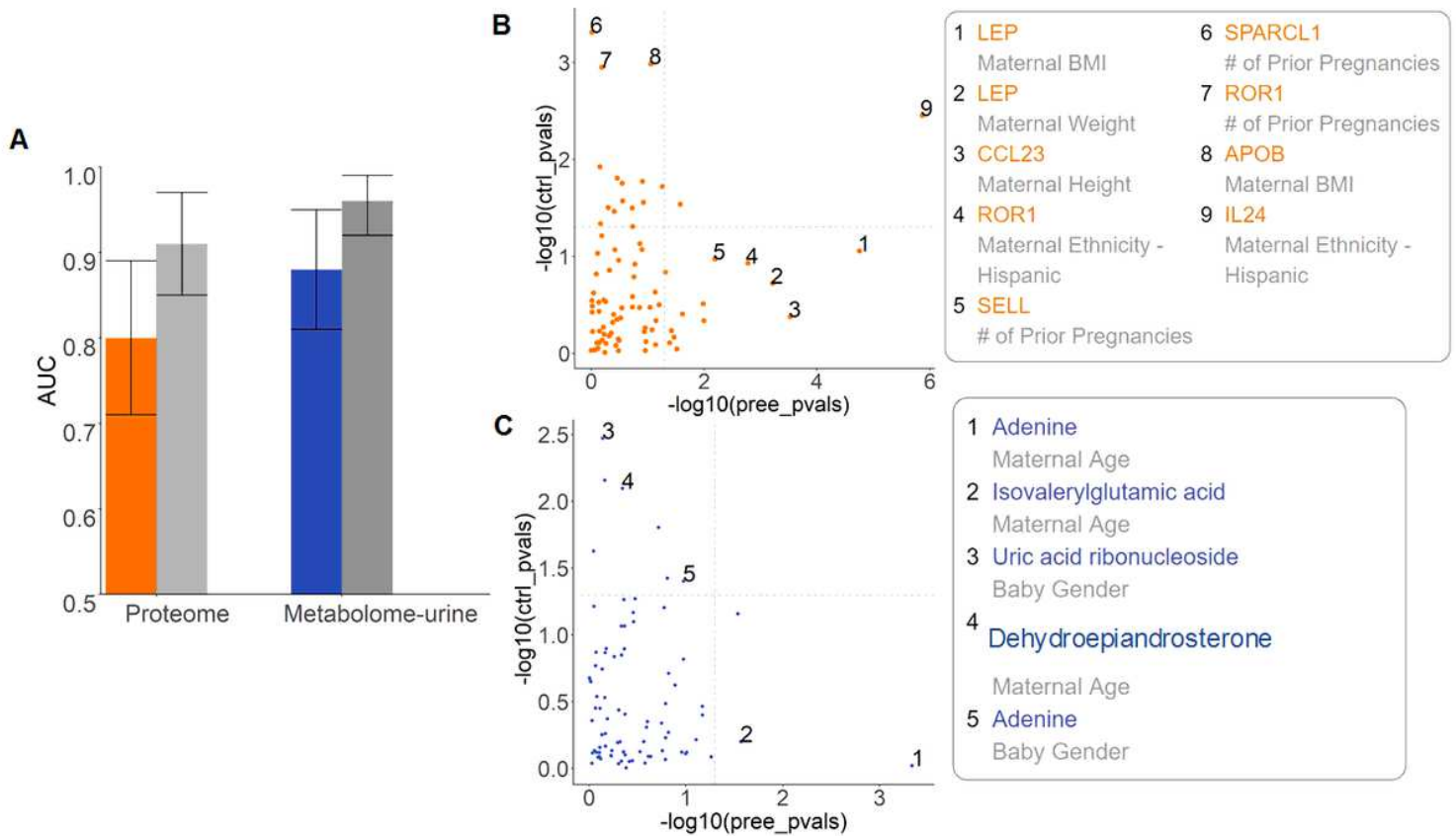


Figure 9

Relationship between urine metabolome and proteome with clinical features over gestation. A. Prediction accuracy of urine metabolome and plasma proteome. Dark blue (for urine metabolome) and orange (for proteome) bars show performance without clinical data (proteome: AUC = 0.83, 95% CI: [0.73, 0.92]; urine metabolome: AUC = 0.88, 95% CI [0.81, 0.95]). Grey bars show performance with clinical data (proteome AUC=0.91, 95% CI: [0.85, 0.97]; urine metabolome AUC=0.96, 95% CI: [0.92, 0.99]). B. Comparison of p-value of correlations of the top proteome and clinical features. Value of $-\log_{10}p$ for preeclamptic patients and controls is shown on x-axis and y-axis, respectively. Each node is a pair of a proteome and a clinical feature. C. Comparison of p-value of correlations of the top urine metabolites and EHR features. Each node is a pair of a proteome/urine metabolome and a clinical feature.

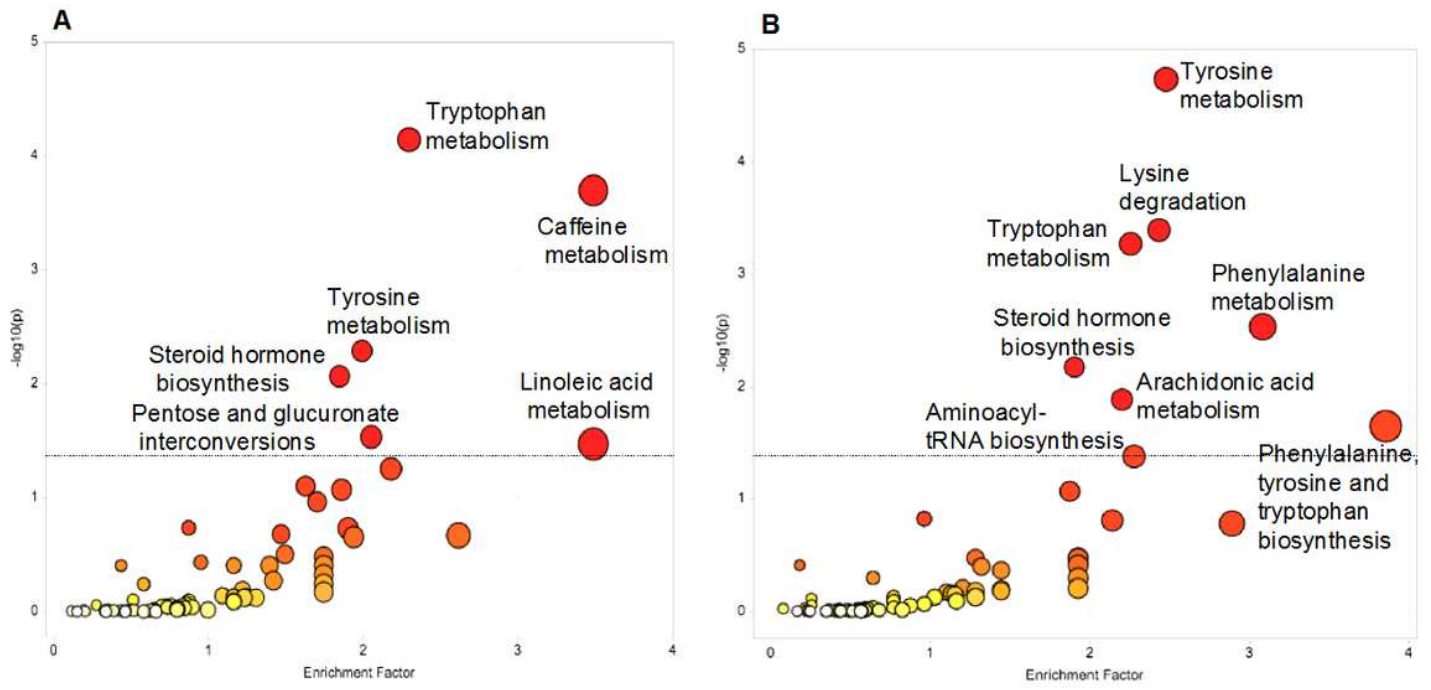


Figure 10

Identified enriched pathways from urine metabolome urine over gestation and in early pregnancy. A. Pathways enrichment analysis over gestation using metabolites from urine that were significant (FDR<0.05, Wilcoxon signed-rank test with Benjamini-Hochberg procedure). Pathways shown above the dotted line were significant ($p < 0.05$). B. Pathways enrichment analysis for early pregnancy using metabolites from urine that were significant (FDR<0.05, Linear Mixed Effects model with Benjamini-Hochberg procedure). The color and the size of a circle are proportional to the $-\log(p)$ and pathway impact value, respectively, where p denotes a p -value.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplement.docx](#)