BMJ Open

# Maternal metabolic profiling to assess fetal gestational age and predict preterm delivery



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2020-040647 |
| Article Type: | Original research |
| Date Submitted by the Author: | 18-May-2020 |
| Complete List of Authors: | Sylvester, Karl; Stanford University<br>Hao, Shiying; Stanford University<br>You, Jin; Stanford University<br>Zheng, Le; Stanford University<br>Tian, Lu; Stanford University<br>Yao, Xiaoming; Hong Kong University<br>Mo, Lihong; University of California San Francisco Fresno Center for Medical Education and Research<br>Ladella, Subhashini; University of California San Francisco Fresno Center for Medical Education and Research<br>Wong, Ronald; Stanford University<br>Shaw, Gary M.; Stanford University<br>Stevenson, David; Stanford University<br>Cohen, Harvet; Stanford University<br>Whitin, John; Stanford University<br>McElhinney, Doff ; Stanford University<br>Ling, Xuefeng; Stanford University, |
| Keywords: | Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, OBSTETRICS |
| | |

*I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.*

*The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.*

*Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.*

# Maternal metabolic profiling to assess fetal gestational age and predict preterm delivery

Karl G. SYLVESTER, MD[1*†], Shiying HAO, PhD[2,3*], Jin YOU, PhD[1*], Le ZHENG, PhD[2,3], Lu TIAN[4], Xiaoming YAO, MD, PhD[5], Lihong MO, MD[6], Subhashini LADELLA, MD[6], Ronald J. WONG, BA[7], Gary M. SHAW, DrPh[7], David K. STEVENSON, MD[7], Harvey J. COHEN, MD, PhD[7], John C. WHITIN, PhD[7], Doff B. MCELHINNEY, MD[2,3], Xuefeng B. LING, PhD[1,3†]

[1]Department of Surgery, Stanford University School of Medicine, Stanford, CA, USA.
[2]Department of Cardiothoracic Surgery, Stanford University School of Medicine, Stanford, CA, USA.
[3]Clinical and Translational Research Program, Betty Irene Moore Children's Heart Center, Lucile Packard Children's Hospital, Palo Alto, CA, USA.
[4]Department of Health Research and Policy, Stanford University, Stanford, CA, USA.
[5]Translational Medicine Laboratory, Queen Mary Hospital, Hong Kong University, Hong Kong, China.
[6]Department of Obstetrics and Gynecology, University of California San Francisco-Fresno, Fresno, CA, USA.
[7]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA.

[†]Correspondence: Xuefeng B. Ling, Department of Surgery, Stanford University, bxling@stanford.edu; Karl G. Sylvester, Department of Surgery, Stanford University, karls@stanford.edu
[*]Co-first authors

**Word count:** 280 for Abstract; 2973 for Main text.

## ABSTRACT

**Objectives** The aim of this study was to develop a single blood test that could determine gestational age and estimate the risk of preterm birth by measuring serum metabolites. We hypothesized that serial metabolic modeling of serum analytes throughout pregnancy could be used to describe fetal gestational age and project preterm birth with a high degree of precision.

**Study design** A retrospective cohort study

**Setting** Two medical centers from US

**Participants** Thirty-six patients (20 full-term, 16 preterm) enrolled at Stanford University were used to develop gestational age and preterm birth risk algorithms, 22 patients (9 full-term, 13 preterm) enrolled at the University of Alabama were used to validate the algorithms.

**Outcome measures** Maternal blood was collected serially throughout pregnancy. Metabolic datasets were generated using mass spectrometry.

**Results** A model to determine gestational age was developed ($R^2 = 0.98$) and validated ($R^2 = 0.81$). 66.7% of the estimates fell within ± 1 week of ultrasound results during model validation. Significant disruptions from full-term pregnancy metabolic patterns were observed in preterm pregnancies ($R^2 = -0.68$). A separate algorithm to predict preterm birth was developed utilizing a set of 10 metabolic pathways that resulted in an area under the curve of 0.92 and a sensitivity of 0.86 during validation testing.

**Conclusions** In this study metabolic profiling was used to develop and test a model for determining gestational age during full-term pregnancy progression, and to determine

risk of preterm birth. With additional patient validation studies, these algorithms may be used to identify at-risk pregnancies prompting alterations in clinical care, and to gain biologic insights into the pathophysiology of preterm birth. Metabolic pathway-based pregnancy modeling is a novel modality for investigation and clinical application development.

**Keywords:** Metabolic, gestational age, preterm birth, pathway

## Strengths and limitations of this study

- This study demonstrates a new non-invasive methodology for monitoring pregnancy progression and identifying abnormal pregnancies at clinical settings.
- The insensitivity of the prediction model to gestational age (GA) window of sample collection increases its flexibility and opportunity for potential clinical use.
- This study is among the first to propose a pathway-based computational methodology to estimate GA and predict preterm birth.
- The overall cohort size is modest, and the distribution of sampling time are different between patients and cohorts.
- It is a retrospective study; a larger prospective cohort study is necessary before applying the estimates and prediction to a broader population for clinical utility.

## INTRODUCTION

Gestational age (GA) dating is a core element of standard prenatal care [1-4]. Prenatal ultrasound (US) is an established modality for estimating GA, monitoring fetal growth, and screening for fetal anomalies [5]. First trimester US imaging is the gold standard for GA determination, however there can be frequent discordance between US dating and a mother's last known menstrual period (LMP). In these cases, follow-up testing by US is utilized to more accurately estimate GA. US measurements are not currently used to determine risk of premature birth (PTB). The availability and expertise of US in disadvantaged areas is limited [6]. Therefore, there is a need to develop an alternative measure of fetal progression to estimate GA and pregnancy risk in a variety of settings and especially when US and LMP dates are unavailable or unreliable.

Compared with imaging methodologies, blood-based molecular testing may provide a more reproducible and precise modality in clinical applications for the frequent monitoring of health status and detection of early signs of disease. Genomic, gene expression, protein, and metabolite profiles measured in human blood have been increasingly utilized for the determination of disease risk and to gain disease specific pathophysiology insight. Attempts at estimating GA using molecular adaptations have included modeling of RNA, protein, or immune cell changes in maternal blood [7-10], but not metabolites. Similarly, risk prediction of PTB in clinical settings is currently primarily based on maternal history. Biomarkers have been suggested from genetic and proteomic analyses, but less effort has been focused on understanding metabolic signatures of pregnancy [11-16].

In this study, we hypothesized that longitudinal metabolic profiling of pregnancy reflects the temporal progression of fetal development with a high degree of precision. Moreover, we posited that if a normal pregnancy progression profile could be defined in metabolic terms, then aberrations from the normal profile may identify a pregnancy at risk for PTB. Herein, we have identified a panel of metabolic pathways measured in maternal serum that provides an estimation of GA over the course of a full-term pregnancy. A second and distinct set of metabolic pathways was also identified in maternal serum that could distinguish pregnancies ending with PTB (< 35 weeks) from full-term (≥ 37 weeks) with a high degree of precision. The models were developed and validated using two independent cohorts from two different institutions in order to test the robustness of the biologic features driving the classifications. Our findings suggest that composite metabolic panel modeling may serve as a reproducible and precision approach to GA dating of pregnancy and prediction of PTB.

## MATERIALS AND METHODS

### Definition

In this study, a full-term pregnancy was defined as a pregnancy ending with a delivery at ≥ 37 weeks. PTB was defined by delivery at < 35 weeks GA.

### Study design

The study was conducted in two phases: (1) modeling to devise a metabolite-based estimation of GA during full-term pregnancies; and (2) modeling to devise a metabolic panel predictive of PTB (Fig. 1). In this study, the 'gold' standard of GA was US measurement. Serum samples were collected in the 1st, 2nd, or 3rd trimester during

pregnancy for each individual woman. Each participant had 1 to 4 time-points collected prior to delivery. Samples were provided by Stanford Hospital and Clinics (SU) and the University of Alabama (UAB). Metabolic concentrations in each sample were measured by targeted and untargeted mass spectrometry (MS) analysis. Models that estimated GA or predicted PTB were developed using the SU cohort and validated using the UAB cohort. The study was approved by the Institutional Review Board of both sites. All samples were collected after informed consent was obtained. All statistical analyses were done in R software.

**Targeted and global MS analysis**

Samples of full-term and preterm patients as well as quality control (QC) samples were injected into the MS. Targeted MS analysis was done through flow injection methods by using Ultimate 3000 Ultra-High-Performance Liquid Chromatography (UHPLC) system and Quantiva Triple Quadrupole Mass Spectrometer. Global (i.e. untargeted) MS analysis was done by using a Vanquish UHPLC system coupled to a Q Exactive plus mass spectrometer and Q Exactive HF hybrid quadrupole-Orbitrap mass spectrometer.

**Data preprocessing and metabolic identification**

A data pre-processing procedure was conducted to convert the raw data generated by MS analysis into a matrix of relative concentrations of metabolites versus samples [17]. This procedure was done by R package. Metabolic values in each sample were then normalized by the median values measured with QC samples to reduce the batch effects.

Compounds detected by untargeted analyses were matched to metabolites in the Human Metabolome Database by putative identification [18]. Accurate mass was used for the

mapping. Metabolites were mapped to pathways using Kyoto Encyclopedia of Genes and Genomes (KEGG) and Human Metabolome Database (HMDB). Only endogenous pathways were considered.

**Metabolic compound selection, pathway computation, and model development**

Metabolites measured by targeted and untargeted MS were aggregated and filtered. The remaining metabolites were mapped to pathways. The value of each pathway was calculated as the weighted sum of the normalized concentrations of metabolites on the pathway divided by the number of metabolites. An XGBoost model was developed with the pathway values of samples from full-term patients to estimate the GA. R-squared ($R^2$; goodness-of-fit of the model), root-mean-square error (RMSE), and error distribution were calculated to evaluate the model performance. A second XGBoost model was developed to predict PTB. To evaluate the model performance, Mann–Whitney U tests were used to compare the distribution of final predictive estimates, i.e., XGBoost model values, on full-term and PTB samples. Results were compared with the insulin-like growth factor-binding protein 4 (IBP4)/sex hormone-binding globulin (SHBG) signature that is commercially available as a metabolic test for determining risk of PTB [12]. Additional details of model development were described in Text A.1.

**Patient and Public Involvement statement**

This retrospective research was done without patient involvement. Patients were not invited to comment on the study design and were not consulted to develop patient relevant outcomes or interpret the results. Patients were not invited to contribute to the writing or editing of this document for readability or accuracy.

## RESULTS

### Samples

As shown in Fig. 2, the SU cohort had 20 full-term pregnancies with 57 blood samples (17, 32, and 8 collected in the 1st, 2nd, and 3rd trimesters, respectively) and 16 preterm pregnancies with 32 blood samples (9, 19, and 4 collected in the 1st, 2nd, and 3rd trimesters, respectively). The UAB cohort had 9 full-term pregnancies with 13 blood samples (8 and 5 in the 2nd, and 3rd trimesters, respectively) and 13 preterm pregnancies with 22 blood samples (4 and 18 in the 1st and 2nd trimesters, respectively). In the SU cohort, 2 (12.5%) were extremely preterm (< 28 weeks), and 5 (31.3%) were very preterm (28–31 weeks). In the UAB cohort, 6 (46.2%) were extremely preterm, and 3 (23.1%) were very preterm. Demographics of the two cohorts are shown in Table 1.

**Table 1.** Maternal characteristics in SU and UAB cohorts

| Characteristic | SU | | | UAB | | |
|---|---|---|---|---|---|---|
| | **Full-term (n = 20)** | **Preterm (n = 16)** | ***P*** | **Full-term (n = 9)** | **Preterm (n = 13)** | ***P*** |
| Race, n (%) | | | **<0.001** | | | 0.5 |
| Asian | 0 | 2 (12.5) | | 0 | 0 | |
| White | 20 (100) | 5 (31.3) | | 0 | 2 (15.4) | |
| Black | 0 | 1 (6.3) | | 9 (100) | 10 (6.9) | |
| American Indian | 0 | 1 (6.3) | | 0 | 0 | |
| Pacific Islander | 0 | 1 (6.3) | | 0 | 0 | |
| Other/unknown | 0 | 6 (37.5) | | 0 | 1 (7.7) | |
| Hispanic, n (%) | 0 | 8 (50) | **<0.001** | 0 | 1 (7.7) | 0.9 |
| Maternal Age, year, mean | 31.9 (4.8) | 29.8 (7.5) | 0.3 | 25.6 (5.0) | 27.5 (4.5) | 0.4 |

| | | | | | | |
|---|---|---|---|---|---|---|
| (SD) | | | | | | |
| Gestational age at delivery, weeks, median (IQR) | 39.5 (39,41) | 32 (30,33) | **<0.001** | 38 (37,39) | 28 (26,32) | **<0.001** |
| Having previous pregnancy, n (%) | 9 (45) | 6 (37.5) | 0.7 | 9 (100) | 13 (100) | 0.4 |
| BMI, kg/m$^2$, median (IQR) | 22.3 (20.2,24.7) | 27.6 (23.4,33.9) | **0.003** | 30.4 (22.3,33.1) | 26.5 (22.6,36.5) | 0.8 |
| History of PTB, n (%) | 3 (15) | 8 (50) | **0.03** | 7 (77.8) | 13 (100) | 0.2 |

## LC-MS/MS metabolomics

The study targeted 315 metabolites by LC-MS/MS, including 13 categories: acyl-carnitine (11, 3.5%), amino acid (9, 2.9%), fatty acid (6, 1.9%), ceramide (12, 3.8%), ceramide 1-phosphate (8, 2.5%), galactosylceramide (5, 1.6%), phosphatidyl acid (15, 4.8%), phosphatidylethanolamine (52, 16.5%), phosphatidylglycerol (5, 1.6%), phosphatidylinositol (11, 3.5%), phophatidylcholine (130, 41.3%), cholesteryl ester (16, 5.1%), and sphingomyelin (35, 11.1%). The study also identified 1627 positively-and 295 negatively-charged compounds through untargeted analyses. Together these formed the initial set of 2237 compounds.

## Feature selection of GA estimation modeling

Of the 2237 compounds, 115 had an absolute Pearson correlation coefficient of > 0.35 with GA. The cutoff of ± 0.35 was selected based on the false discovery rate (FDR) values of the mapped pathways < 1% (Fig. A.1). The 115 compounds were mapped to 89 pathways, 33 of which were selected by the XGBoost model. The normalized value of

each pathway varied over the course of gestation (Fig. A.2). Univariate analysis of the 33 pathways is shown in Fig. A.3, and the top 10 pathways in the model is depicted in Fig. 3. The top 10 pathways included those associated in the metabolisms of: glycerophospholipid, arginine and proline, thiamine, purine, butanoate, galactose, sulfur, phenylalanine, and C5-branched dibasic acid.

**Performance of GA estimation**

The performance of GA estimates on full-term samples was similar in the development phase (SU cohort, $R^2 = 0.98$, RMSE = 1.09) and the validation phase (UAB cohort, $R^2 = 0.81$, RMSE = 2.36) (Fig. 4). In our validation testing, 66.7% of the estimates were within ± 1 week of the US results (Fig. A.4).

Intriguingly, model performance significantly deteriorated when applied to samples from PTB pregnancies ($R^2 = -0.68$ and RMSE = 6.6 in validation; see Fig. 4). It suggested that the relationships between metabolic parameters and full-term pregnancies were not maintained in PTB pregnancies. Furthermore, such disruptions were notable as early as 10 weeks' GA (Fig. 4) or early to mid-gestation. These findings prompted the development of a metabolic-based model of PTB estimation.

**Performance of PTB prediction**

Samples collected before 35 weeks' GA were used to develop a model that differentiated PTB pregnancies from those full-term. As before, the model was developed with the SU cohort that had 20 full-term (54 samples) and 16 preterm (32 samples) pregnancies, and was validated with the UAB cohort that had 9 full-term (13 samples) and 13 preterm (22 samples) pregnancies. In total, 148 metabolic compounds (with Mann-Whitney U test $P <$

0.05) were mapped to 66 pathways (FDR < 1.5%; see Fig. A.5). Further model development selected 10 pathways as strong predictors covering the metabolisms of glycerophospholipid, sphingolipid, taurine and hypotaurine, arachidonic acid, secondary bile acid biosynthesis, glycerolipid, cysteine and methionine, tryptophan, and arginine and proline (Fig. 5).

The level of prediction accuracy was maintained in the validation cohort ($P = 5x10^{-5}$, area under the curve [AUC] = 0.92; see Fig. 6). The prevalence-corrected positive predictive values (PPVs) across model values (*i.e.* scores) were plotted based on the national PTB prevalence in the United States (9.71% [12 19]; see Fig. A.6). A threshold value of 0.52 was selected as a high-risk threshold for PTB, which was associated with a PPV of 0.61, a relative risk (RR) of 6.3 compared to the United States population baseline (= 0.61/9.71%), a sensitivity of 0.86 (19 of 22), and a specificity of 0.92 (12 of 13; Fig. 7). The sensitivities and specificities with cutoff values are shown in Table A.1.

In the validation cohort, 12 of 13 full-term samples and 19 of 22 preterm samples were classified correctly. The misclassified full-term sample was from a mother that delivered at 37 weeks' GA. The 19 correctly classified PTB samples were from 13 PTB pregnancies. Of the 13 pregnancies, 9 were identified as high risk at or earlier than 16 weeks' GA. The median gap between the time of identification and the delivery was 11 weeks' GA (IQR: 8, 15.5).

To determine the performance of our metabolic model against existing models, a comparison between the metabolic PTB risk model and the commercially available IBP4/SHBG PTB test was performed and summarized in Text A.2.

**Metabolite-based model and pathway-based model: a comparison**

To determine the effectiveness of model performance based upon robustness of biologic features, we compared model performance using pathway or individual metabolite as selected features in estimating GA and predicting PTB. The performance of the pathway-based models were significantly better than the metabolite-based models, with a lower RMSE (Student's t-test $P = 4x10^{-3}$; Fig. A.7) and a larger AUC (DeLong test $P = 0.03$; Fig. A.8).

## DISCUSSION

### Principal Findings

In this study, metabolic modeling of maternal sera collected across gestation proved to be a robust method of determining GA during pregnancy progression of term deliveries (>37 weeks' GA), in that it was validated in a population of women from a different center. Intriguingly, PTB pregnancies do not demonstrate the same temporal relationship as term pregnancies upon metabolic modeling across gestation (Fig. 4). Indeed, PTB pregnancies (<35 weeks' GA) demonstrate a marked departure from the term metabolic profile (Fig. 4) that is not only dramatic ($R^2$= 0.98 train and 0.81 test for term model; compared to $R^2$= 0.50 train and -0.68 test for PTB pregnancy in term model), but is also recognizable as early as 10 weeks' GA as determined by the current standard of US dating. Recognizing the metabolic pathway aberration of PTB pregnancies, a second model was developed using metabolic pathway analyses to quantify the risk of PTB prior to 35 weeks' GA. Once again, metabolic profiling proved to be robust in identifying PTB pregnancies with a high degree of sensitivity (AUC 0.96 training; AUC 0.92 testing) and precision (training PPV 0.93 (0.78-0.99); testing PPV 0.95 (0.75-1). Taken together, this study

demonstrated a powerful new methodology for monitoring pregnancy progression and identifying abnormal pregnancies.

**Clinical and Research Implications**

The potential clinical utility of developing a test for pregnancy monitoring is appealing. There is a need to develop a more robust method than LMP and US that captures pregnancy progression, a complex relationship of fetal and placental growth, development, and function. To support these processes, there is a need for energy transfer between mother and fetus throughout gestation. We therefore reasoned that metabolic phenotyping would be ideally suited to capture this relationship. Despite a modest cohort size, the results of metabolic modeling demonstrate a high degree of concordance with clinical standard US dating performed by experts as reflected by 66.7% of model estimates falling within ± 1 week of US results (Fig. A.4). Moreover, unlike the deterioration experienced with US dating of pregnancy, metabolic modeling was shown to achieve near equivalent performance in the 1$^{st}$, 2$^{nd}$, and 3$^{rd}$ trimesters, indicating the potential for broad clinical applicability that might achieve independence of reliance on accuracy of LMP or concordance among modality testing. The result of PTB prediction is equally robust demonstrating a high degree of precision. Beyond relying on clinical histories or self-reported symptoms, the model proposed here provides a molecular classification that may be more accurate than current methods and further reflect a comprehensive measure of aberrant pregnancy based on metabolic changes. In practice, clinicians could use the PTB prediction model to differentiate high- from low-risk patients. Low risk patients would then be subject to GA estimation panel testing, all from the same blood draw.

A distinct advantage of the PTB risk prediction developed in this study is that it has a wide window of sampling. Samples were collected broadly before 35 weeks' GA, which is wider than the window of other well-established biomarkers such as fetal fibronectin (between 24 and 34 weeks' GA) [13], IBP4/SHBG (19 to 21 weeks) [12], and inter-alpha-trypsin inhibitor heavy chain 4 protein (24 and 28 weeks) [11]. Relatively stable AUC levels were maintained throughout the diagnostic window (Text A.2). The insensitivity of the prediction model to GA at testing increases its flexibility and opportunity for potential clinical use. An additional advantage of the model herein is the ability for early identification of high-risk women. Although there is no standardized guideline for early-gestation management of patients at risk of PTB delivery, metabolic modeling for PTB risk may provide a not previously possible opportunity for early gestation risk mitigation. Clinical trials have suggested that hormone treatment and maternal physical activity modifications applied between 16 to 37 weeks' GA reduced the PTB rate of women who were deemed at high risk due to a history of prior PTB delivery [20 21]. In many cases PTB can not be prevented, however any opportunity is deemed highly desirable for even a modest delay (1–2 weeks) in PTB or an enhanced ability to more accurately triage for delivery to centers with the capability to manage profoundly premature neonates [22-24].

This study is among the first to propose a pathway-based computational methodology to estimate GA and predict PTB. Metabolic pathways are linked to chemical functions, and the alteration or disruption of specific functions participate in disease phenotypes, facilitating the use of pathways to function as higher-level biomarkers of diseases [25]. The role of metabolic pathways in disease diagnosis has been explored in several preliminary clinical studies [26 27]. Pathway performance in differentiating patients with disease from

healthy controls has been found to be effective compared to using individual metabolites [27]. Similarly, we found the pathway-based models had less variability and higher sensitivity than metabolite-based models that were developed using the same population. One plausible explanation for this observation may be attributed to the calculation of pathway values, which represents the sum of individual metabolites and thus may amplify association to outcome relationships. This hypothesis is supported by the FDR comparison (Fig. A.7 and A.8): pathway-based analysis had lower FDR values than metabolite models. This study adds to the exploration of the feasibility of using pathways for health monitoring and prediction.

**Limitations**

This study has several limitations. First, the overall cohort size was modest. Second, blood samples were collected in a non-uniform manner with respect to GA timing and time of day. The time between two adjacent samples corresponding to the same patient varied. Third, the distribution of samples throughout pregnancy were different between patients and cohorts. In the SU cohort, none of the full-term patients had samples collected between 30 and 37 weeks. In the UAB cohort, none of the full-term patients had sampling in the 1st trimester, and none of the PTB patients had sampling in the 3rd trimester. Fourth, for methodologic reasons, not all serum analytes could be identified and mapped to known metabolites. Fifth, the study was retrospective, and the participants were solely from California and Alabama. A larger prospective cohort study is necessary before applying the estimates and prediction to a broader population for clinical utility.

**CONCLUSION**

The present study demonstrates that maternal serum based metabolic profiling is a highly sensitive and accurate method for determining GA and prediction of PTB. The pathway-based analysis supports the hypothesis of the orderly metabolic progression of pregnancy that can be reproducibly captured using metabolic profiling. The robustness of the modeling reinforces the potential appeal for further clinical development and as a platform to investigate the pathophysiology associated with aberrant fetal development and pregnancy progression. This study is the first to report a single blood test for metabolic pathway-based determination of GA dating, and early detection of PTB risk.

**Acknowledgments:** The authors thank colleagues at the Stanford University Pediatric Proteomics group and the March of Dimes Prematurity Research Center at Stanford University for critical discussions.

**Funding:** This work was supported in part by the March of the Dimes Prematurity Research Center at Stanford University, and Stanford Child Health Research Institute. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Conflict of Interest:** The authors report no conflict of interest.

**Author contributions:** XBL, KGS, and HJC contributed to concept development and design.

JY, RJW, and DKS contributed to the acquisition of data.

KGS, SH, LZ, XY, LT, LM, SL, RJW, GMS, DKS, JCW and DBM contributed to the analysis and interpretation of data.

KGS and SH drafted the manuscript.

JY, LZ, LT, XY, LM, SL, RJW, GMS, DKS, HJC, JCW, DBM, and XBL critically revised the manuscript.

All the authors gave final approval of the version to be submitted and agreed to be accountable for all aspects of the work.

**Data and materials availability:** The datasets used and/or analyzed in this study are available upon request to the corresponding author.

## References

1. Brownfoot FC, Gagliardi DI, Bain E, et al. Different corticosteroids and regimens for accelerating fetal lung maturation for women at risk of preterm birth. *Cochrane Database Syst Rev* 2013(8):CD006764. doi: 10.1002/14651858.CD006764.pub3
2. Raju TN, Mercer BM, Burchfield DJ, et al. Periviable birth: executive summary of a Joint Workshop by the Eunice Kennedy Shriver National Institute of Child Health and Human Development, Society for Maternal-Fetal Medicine, American Academy of Pediatrics, and American College of Obstetricians and Gynecologists. *Journal of perinatology : official journal of the California Perinatal Association* 2014;34(5):333-42. doi: 10.1038/jp.2014.70
3. Vohr B. Long-term outcomes of moderately preterm, late preterm, and early term infants. *Clin Perinatol* 2013;40(4):739-51. doi: 10.1016/j.clp.2013.07.006
4. Pereira AP, Dias MA, Bastos MH, et al. Determining gestational age for public health care users in Brazil: comparison of methods and algorithm creation. *BMC Res Notes* 2013;6:60. doi: 10.1186/1756-0500-6-60
5. Peek MJ, Devonald KJ, Beilby R, et al. The value of routine early pregnancy ultrasound in the antenatal booking clinic. *Aust N Z J Obstet Gynaecol* 1994;34(2):140-3.
6. Jehan I, Zaidi S, Rizvi S, et al. Dating gestational age by last menstrual period, symphysis-fundal height, and ultrasound in urban Pakistan. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics* 2010;110(3):231-4. doi: 10.1016/j.ijgo.2010.03.030
7. Knight AK, Craig JM, Theda C, et al. An epigenetic clock for gestational age at birth based on blood methylation data. *Genome biology* 2016;17(1):206. doi: 10.1186/s13059-016-1068-z
8. Ngo TTM, Moufarrej MN, Rasmussen MH, et al. Noninvasive blood tests for fetal development predict gestational age and preterm delivery. *Science* 2018;360(6393):1133-36. doi: 10.1126/science.aar3819
9. Aghaeepour N, Lehallier B, Baca Q, et al. A proteomic clock of human pregnancy. *American journal of obstetrics and gynecology* 2018;218(3):347 e1-47 e14. doi: 10.1016/j.ajog.2017.12.208
10. Aghaeepour N, Ganio EA, McIlwain D, et al. An immune clock of human pregnancy. *Sci Immunol* 2017;2(15) doi: 10.1126/sciimmunol.aan2946
11. Esplin MS, Merrell K, Goldenberg R, et al. Proteomic identification of serum peptides predicting subsequent spontaneous preterm birth. *American journal of obstetrics and gynecology* 2011;204(5):391 e1-8. doi: 10.1016/j.ajog.2010.09.021
12. Saade GR, Boggess KA, Sullivan SA, et al. Development and validation of a spontaneous preterm delivery predictor in asymptomatic women. *American journal of obstetrics and gynecology* 2016;214(5):633 e1-33 e24. doi: 10.1016/j.ajog.2016.02.001
13. Peaceman AM, Andrews WW, Thorp JM, et al. Fetal fibronectin as a predictor of preterm birth in patients with symptoms: a multicenter trial. *American journal of obstetrics and gynecology* 1997;177(1):13-8.

14. Strauss JF, 3rd, Romero R, Gomez-Lopez N, et al. Spontaneous preterm birth: advances toward the discovery of genetic predisposition. *American journal of obstetrics and gynecology* 2018;218(3):294-314 e2. doi: 10.1016/j.ajog.2017.12.009
15. Virgiliou C, Gika HG, Witting M, et al. Amniotic Fluid and Maternal Serum Metabolic Signatures in the Second Trimester Associated with Preterm Delivery. *Journal of proteome research* 2017;16(2):898-910. doi: 10.1021/acs.jproteome.6b00845
16. Li J, Lu YP, Reichetzeder C, et al. Maternal PCaaC38:6 is Associated With Preterm Birth - a Risk Factor for Early and Late Adverse Outcome of the Offspring. *Kidney Blood Press Res* 2016;41(3):250-7. doi: 10.1159/000443428
17. Dunn WB, Broadhurst D, Begley P, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc* 2011;6(7):1060-83. doi: 10.1038/nprot.2011.335
18. Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 2007;3(3):211-21. doi: 10.1007/s11306-007-0082-2
19. Martin JA, Hamilton BE, Osterman MJ, et al. Births: final data for 2013. *Natl Vital Stat Rep* 2015;64(1):1-65.
20. Meis PJ, Klebanoff M, Thom E, et al. Prevention of recurrent preterm delivery by 17 alpha-hydroxyprogesterone caproate. *The New England journal of medicine* 2003;348(24):2379-85. doi: 10.1056/NEJMoa035140
21. Evenson KR, Siega-Riz AM, Savitz DA, et al. Vigorous leisure activity and pregnancy outcome. *Epidemiology* 2002;13(6):653-9. doi: 10.1097/01.EDE.0000021463.45041.95
22. McIntire DD, Leveno KJ. Neonatal mortality and morbidity rates in late preterm births compared with births at term. *Obstetrics and gynecology* 2008;111(1):35-41. doi: 10.1097/01.AOG.0000297311.33046.73
23. Henderson-Smart DJ. The effect of gestational age on the incidence and duration of recurrent apnoea in newborn babies. *Aust Paediatr J* 1981;17(4):273-6.
24. Khashu M, Narayanan M, Bhargava S, et al. Perinatal outcomes associated with preterm birth at 33 to 36 weeks' gestation: a population-based cohort study. *Pediatrics* 2009;123(1):109-13. doi: 10.1542/peds.2007-3743
25. Lee DS, Park J, Kay KA, et al. The implications of human metabolic network topology for disease comorbidity. *Proceedings of the National Academy of Sciences of the United States of America* 2008;105(29):9880-5. doi: 10.1073/pnas.0802208105
26. Baumgartner C, Bohm C, Baumgartner D, et al. Supervised machine learning techniques for the classification of metabolic disorders in newborns. *Bioinformatics* 2004;20(17):2985-96. doi: 10.1093/bioinformatics/bth343
27. Huang S, Chong N, Lewis NE, et al. Novel personalized pathway-based metabolomics models reveal key metabolic pathways for breast cancer diagnosis. *Genome medicine* 2016;8(1):34. doi: 10.1186/s13073-016-0289-9

## Figure Legends

**Fig. 1.** Study design. Models were developed separately to estimate gestational age during full-term pregnancy, and to predict the risk of preterm birth. Both models were developed with the SU cohort and validated with the UAB cohort.

**Fig. 2.** Cohort construction. Each line represents an individual patient. Diamond and triangle markers indicate sample collection dates and delivery dates, respectively. The red dashed line represents 37 weeks' gestational age.

**Fig. 3.** The importance of the top 10 metabolic pathways in the gestational age estimation model. Pathways either positively or negatively correlated gestational age.

**Fig. 4.** Gestational age estimates of the gestational age model with the SU ($R^2$=0.98, RMSE=1.09 weeks) and UAB cohorts ($R^2$ = 0.81, RMSE = 2.36 weeks).

**Fig. 5.** (A) Univariate analysis of the 10 metabolic pathways in the preterm birth prediction model. Odds ratio of each pathway was calculated. **P*<0.05, ***P*<0.01, ****P*<0.005. (B) The importance of the metabolic pathways in the preterm birth prediction model. Pathways were either up- or down-regulated in relation to preterm birth.

**Fig. 6.** (A) Prediction of preterm birth risk grouped by full-term and preterm birth patients (top) and over the course of gestation (bottom). (B) AUC performance of the prediction in SU and UAB cohorts. *P* was calculated using Mann–Whitney U test. wks: weeks' gestational age.

**Fig. 7.** Performance of the preterm birth prediction model. (A) A contingency table showing the number of samples in each category. (B) Sensitivity, specificity, PPV, and NPV together with the 95% confidence intervals.

## Appendix Captions

**Fig. A.1** False discovery rate (FDR) analysis of the metabolic pathways significantly associated with the GA in full-term pregnancies. Pearson |r| was calculated as the correlation between metabolite serological abundance and GA. Only the metabolites with a Pearson |r| higher than the threshold would be selected as part of the significant pathways. FDR was estimated by a permutation-based method (permutation N=1000).

**Fig. A.2** Profile of the metabolic pathways in the GA estimation model over the course of gestation on SU cohort. All pathways are (A) positively or (B) negatively correlated to the GA (FDR<1%). Profile of each pathway was calculated as the weighted sum of the z-score normalized metabolite serological abundances divided by the number of metabolites. Means ± standard errors at each time point were plotted.

**Fig. A.3** Univariate analysis of the 33 metabolic pathways in the GA estimation model. Pearson correlation coefficient of each pathway to GA was calculated. **P*<0.05, ***P*<0.01, ****P*<0.005.

**Fig. A.4** Comparison of GA estimates using the model and US measurements. (A) Distributions of differences between GA measured by US and GA estimated by the model, in T2 (weeks 14–27), T3 (weeks 28–40), and T2+T3. n represents the number of full-term patients included. (B) Error distribution of GA estimation on a combination of SU and UAB cohorts in T2, T3, and T2+T3.

**Fig. A.5** False discovery rate (FDR) analysis of the metabolic pathways significantly associated with PTB. Mann-Whitney U test *P* measured the difference in metabolite serological abundances between full-term pregnancies and pregnancies ending in PTB.

Only metabolites with a Mann-Whitney U test *P* lower than the threshold were selected as part of the significant pathways. FDR was estimated by a permutation-based method (permutation N=1000).

**Fig. A.6** Stratification of patients by the classification model prediction on the UAB cohort. PPV was corrected by bootstrapping the full-term patients to reach the population PTB prevalence of 9.71% on singleton births. Two horizontal dashed lines represent the population mean of PTB risk that is 9.71% (black) and the PPV (= 0.61; red) at the high-risk cutoff. The grey dashed line indicates the high-risk cutoff value (= 0.52). The grey area represents the 95% confidence interval of the PPV. The box plot at the bottom shows the classification model value distribution stratified by the samples. GAB: GA at birth. wks: weeks of gestation.

**Fig. A.7** (A) False discovery rate (FDR) analysis of the metabolites and metabolic pathways significantly associated with GA in full-term pregnancies. Pearson |r| was calculated as the correlation between metabolite serological abundance and GA. Only the metabolites with a Pearson |r| higher than the threshold (=0.35) would be selected as part of the significant pathways. FDR was estimated by a permutation-based method (permutation N=1000). (B) A comparison of RMSE of the GA estimation model trained by pathways and the model trained by metabolites. All metabolites had a Pearson |r|>0.35. RMSE was measured with the full-term samples of the validation (UAB) cohort.

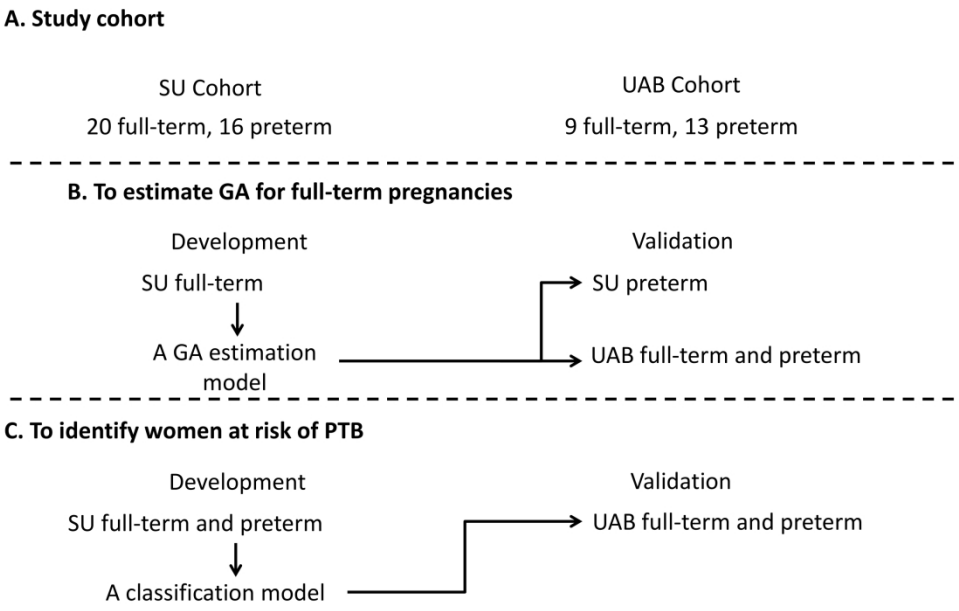
**Fig. A.8** (A) False discovery rate (FDR) analysis of the metabolites and metabolic pathways significantly associated with the PTB. Mann-Whitney U test *P* measured the difference in metabolite serological abundances between full-term pregnancies and pregnancies ending in PTB. Only the metabolites with a Mann-Whitney U test *P* lower

than the threshold (=0.05) would be selected as part of the significant pathways. FDR was estimated by a permutation-based method (permutation N=1000). (B) A comparison of the AUC of the PTB classification model utilizing pathways and the model utilizing metabolites. All the metabolites had a Mann-Whitney U test $P < 0.05$. AUC was measured with the samples of the validation (UAB) cohort.

**Table A.1** Sensitivity and specificity of the XGBoost model with respect to the cutoff point.

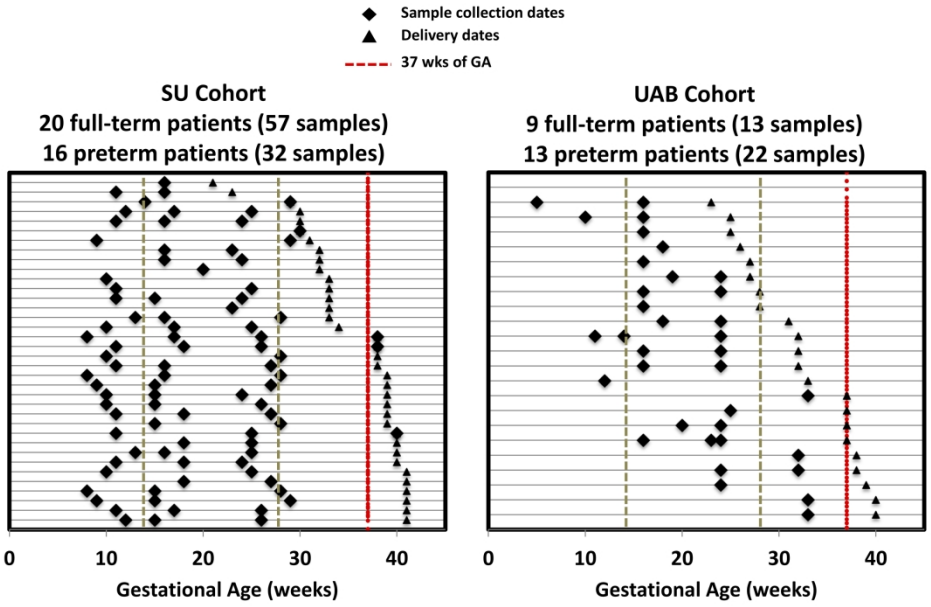
**Text A.1** Metabolic compound selection, pathway computation, and model development

**Text A.2** Metabolite model vs. IBP4/SHBG in predicting PTB



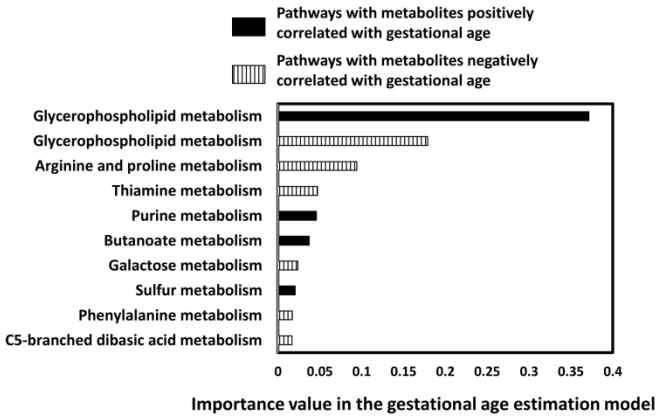

Study design. Models were developed separately to estimate gestational age during full-term pregnancy, and to predict the risk of preterm birth. Both models were developed with the SU cohort and validated with the UAB cohort.

254x190mm (300 x 300 DPI)



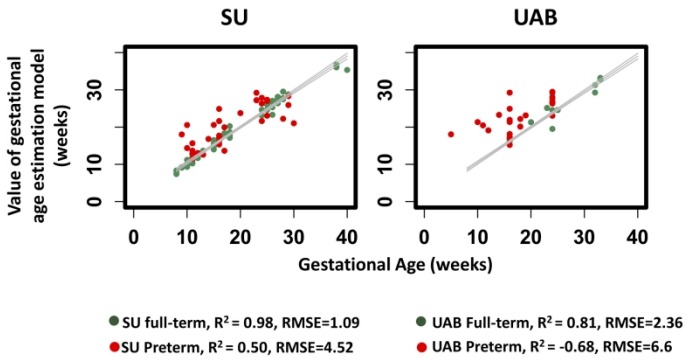

Cohort construction. Each line represents an individual patient. Diamond and triangle markers indicate sample collection dates and delivery dates, respectively. The red dashed line represents 37 weeks' gestational age.

254x190mm (300 x 300 DPI)



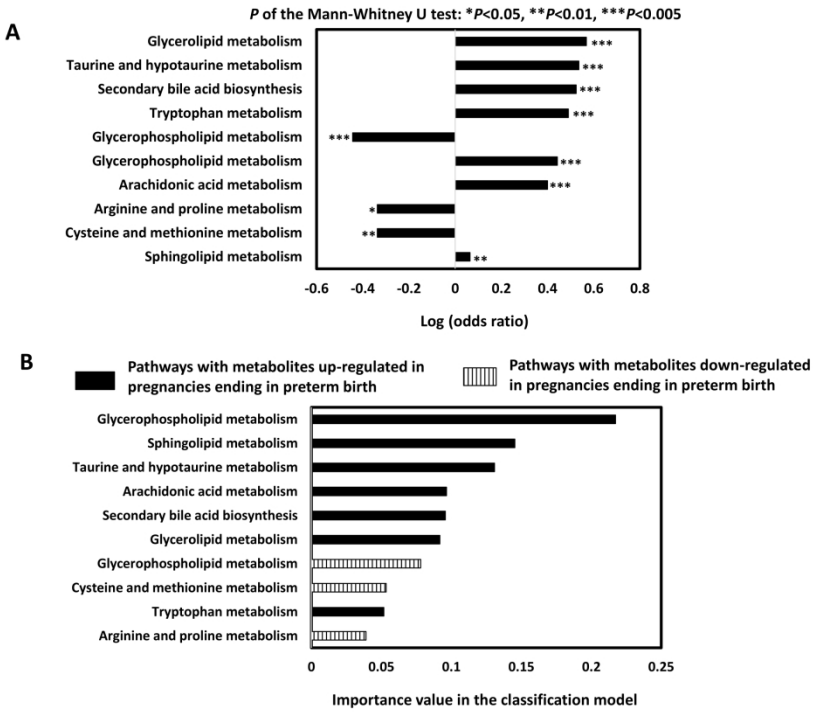

The importance of the top 10 metabolic pathways in the gestational age estimation model. Pathways either positively or negatively correlated gestational age.

254x190mm (300 x 300 DPI)



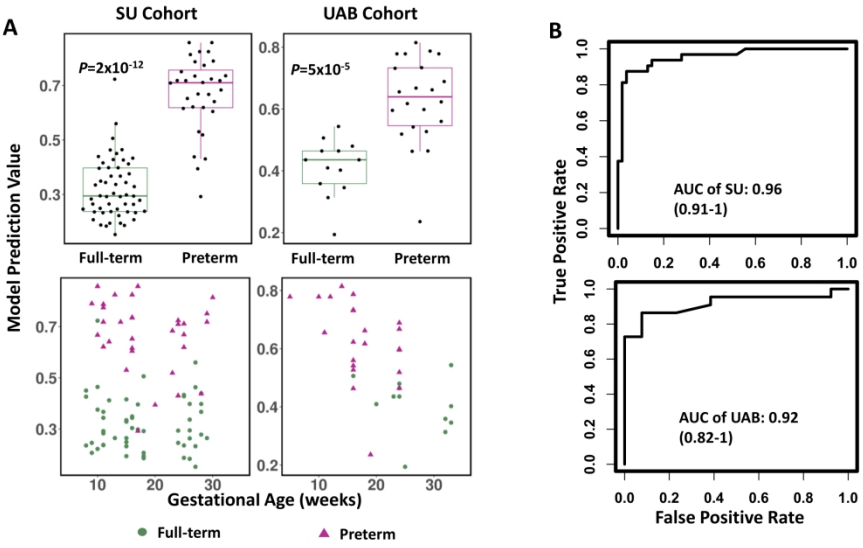

Gestational age estimates of the gestational age model with the SU (R2=0.98, RMSE=1.09 weeks) and UAB cohorts (R2 = 0.81, RMSE = 2.36 weeks).

254x190mm (300 x 300 DPI)




(A) Univariate analysis of the 10 metabolic pathways in the preterm birth prediction model. Odds ratio of each pathway was calculated. *P<0.05, **P<0.01, ***P<0.005. (B) The importance of the metabolic pathways in the preterm birth prediction model. Pathways were either up- or down-regulated in relation to preterm birth.

254x190mm (300 x 300 DPI)




(A) Prediction of preterm birth risk grouped by full-term and preterm birth patients (top) and over the course of gestation (bottom). (B) AUC performance of the prediction in SU and UAB cohorts. P was calculated using Mann–Whitney U test. wks: weeks' gestational age.

254x190mm (300 x 300 DPI)

A

**SU**

| | Preterm | Full-term |
|---|---|---|
| Classified as Preterm | 28 | 2 |
| Classified as Full-term | 4 | 52 |

**UAB**

| | Preterm | Full-term |
|---|---|---|
| Classified as Preterm | 19 | 1 |
| Classified as Full-term | 3 | 12 |

B

| Cohort | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) |
|---|---|---|---|---|
| **SU** | 0.88 (0.71-0.97) | 0.96 (0.87-1) | 0.93 (0.78-0.99) | 0.93 (0.83-0.98) |
| **UAB** | 0.86 (0.65-0.97) | 0.92 (0.64-1) | 0.95 (0.75-1) | 0.80 (0.52-0.96) |

Performance of the preterm birth prediction model. (A) A contingency table showing the number of samples in each category. (B) Sensitivity, specificity, PPV, and NPV together with the 95% confidence intervals.

254x190mm (300 x 300 DPI)

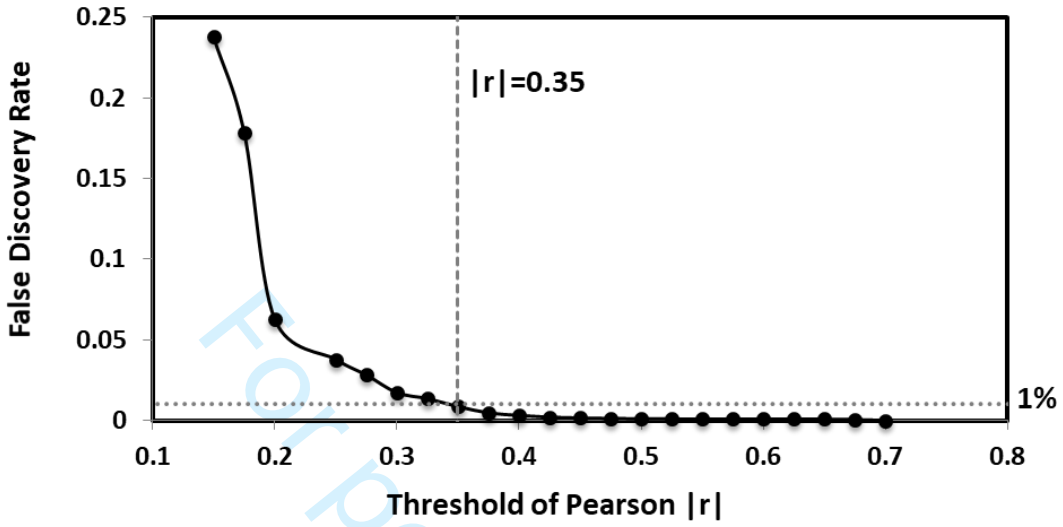



**Fig. A.1.** False discovery rate (FDR) analysis of the metabolic pathways significantly associated with the GA in full-term pregnancies. Pearson |r| was calculated as the correlation between metabolite serological abundance and GA. Only the metabolites with a Pearson |r| higher than the threshold would be selected as part of the significant pathways. FDR was estimated by a permutation-based method (permutation N=1000).

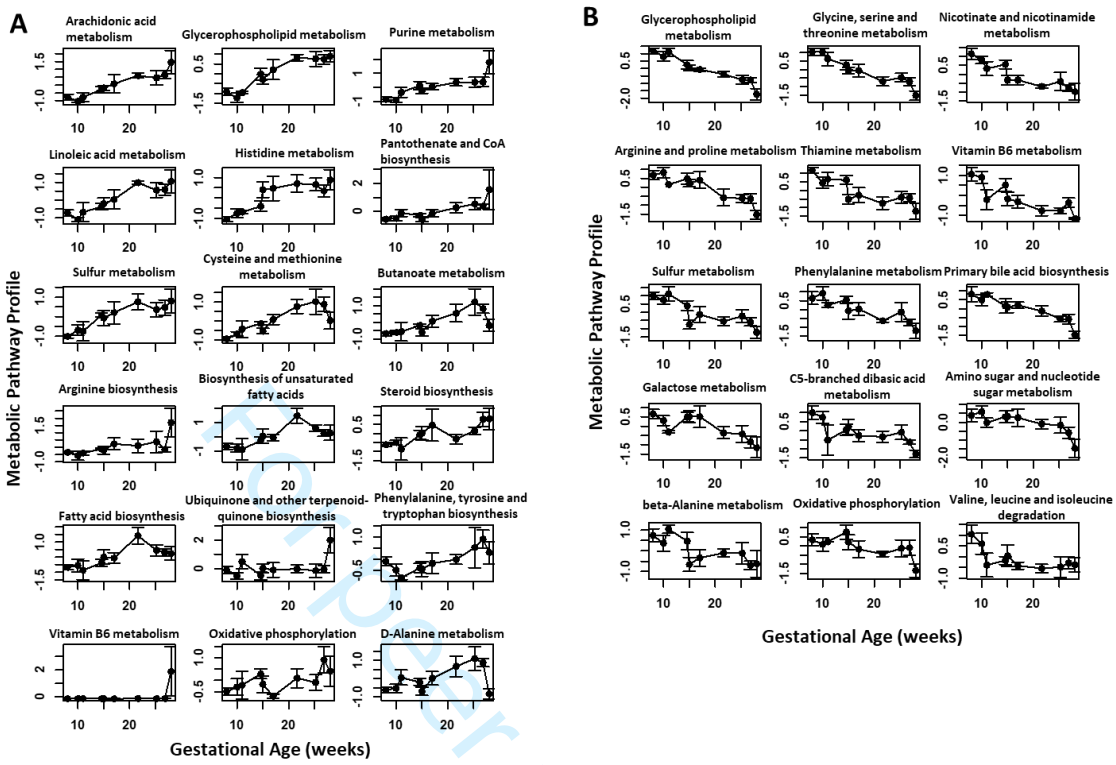



**Fig. A.2.** Profile of the metabolic pathways in the GA estimation model over the course of gestation on SU cohort. All pathways are (A) positively or (B) negatively correlated to the GA (FDR<1%). Profile of each pathway was calculated as the weighted sum of the z-score normalized metabolite serological abundances divided by the number of metabolites. Mean ± standard error of the mean at each time point was plotted.

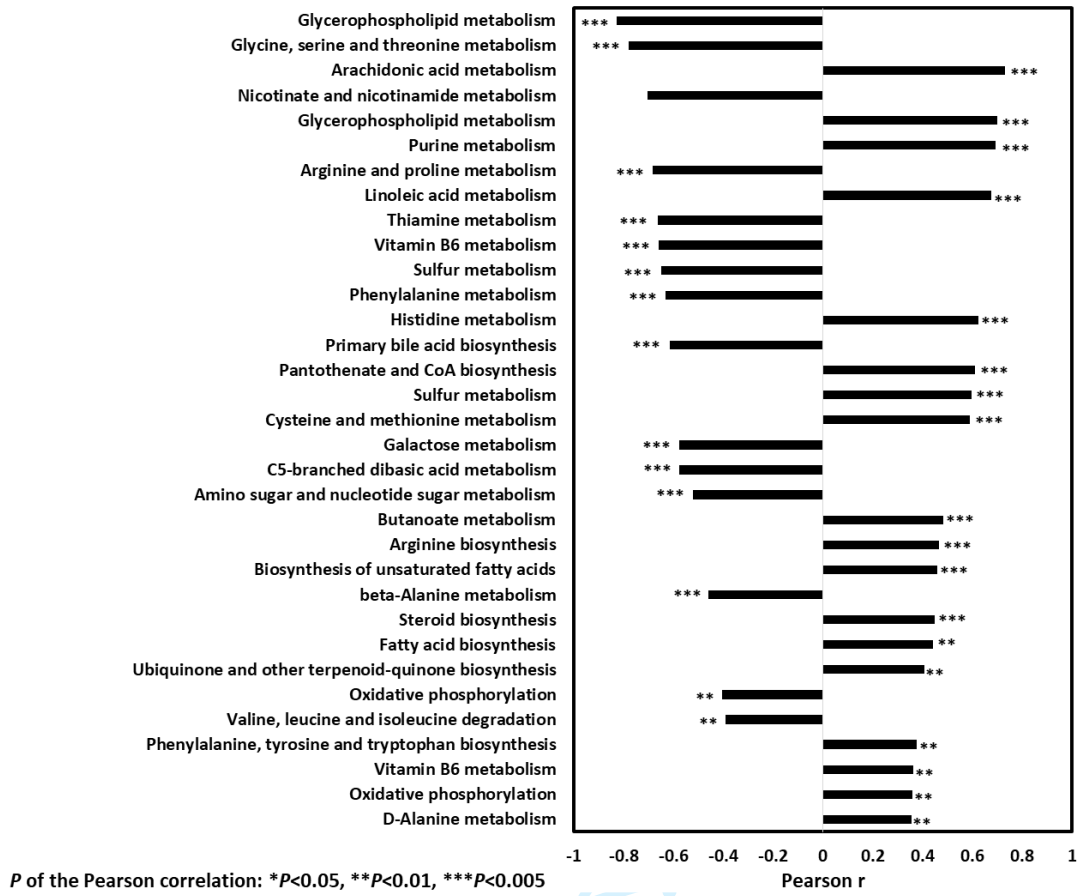



**Fig. A.3.** Univariate analysis of the 33 metabolic pathways in the GA estimation model. Pearson correlation coefficient r of each pathway to GA was calculated. *$P$<0.05, **$P$<0.01, ***$P$<0.005.

A

| Trimester and subject number | Δ [model estimation – ultrasound measurements (weeks)] (%) | | | | |
|---|---|---|---|---|---|
| | < -2 | -1 to -2 | ±1 | +1 to +2 | > +2 |
| SU (T2, n = 19) | 0 | 0 | 84.2 | 15.8 | 0 |
| SU (T3, n = 8) | 12.5 | 25 | 50 | 12.5 | 0 |
| SU (All, n = 20) | 0 | 5 | 85 | 10 | 0 |
| UAB (T2, n = 5) | 0 | 0 | 60 | 0 | 40 |
| UAB (T3, n = 5) | 20 | 0 | 80 | 0 | 0 |
| UAB (All, n = 9) | 11.1 | 0 | 66.7 | 11.1 | 11.1 |
| SU and UAB (T2, n = 24) | 0 | 0 | 79.2 | 12.5 | 8.3 |
| SU and UAB (T3, n = 13) | 15.4 | 15.4 | 61.5 | 7.7 | 0 |
| SU and UAB (All, n = 29) | 3.4 | 3.4 | 79.3 | 10.3 | 3.4 |

B

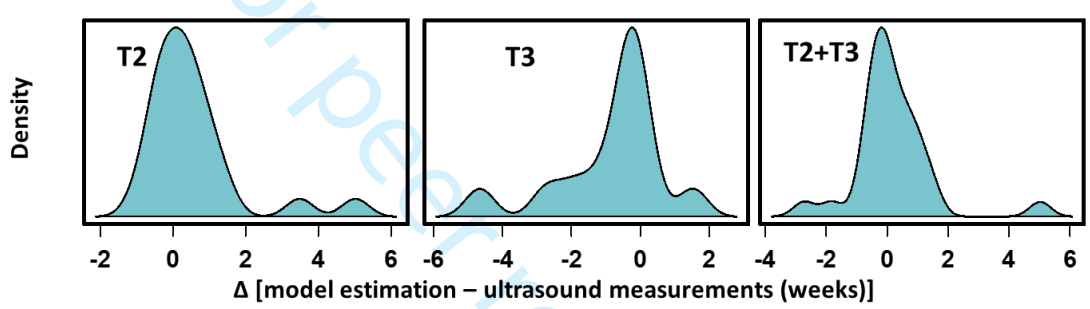



**Fig. A.4.** Comparison of GA estimates using the model and US measurements. (A) Distributions of differences between GA measured by US and GA estimated by the model, in T2 (weeks 14–27), T3 (weeks 28–40), and T2+T3. n represents the number of full-term patients included. (B) Error distribution of GA estimation on a combination of SU and UAB cohorts in T2, T3, and T2+T3.

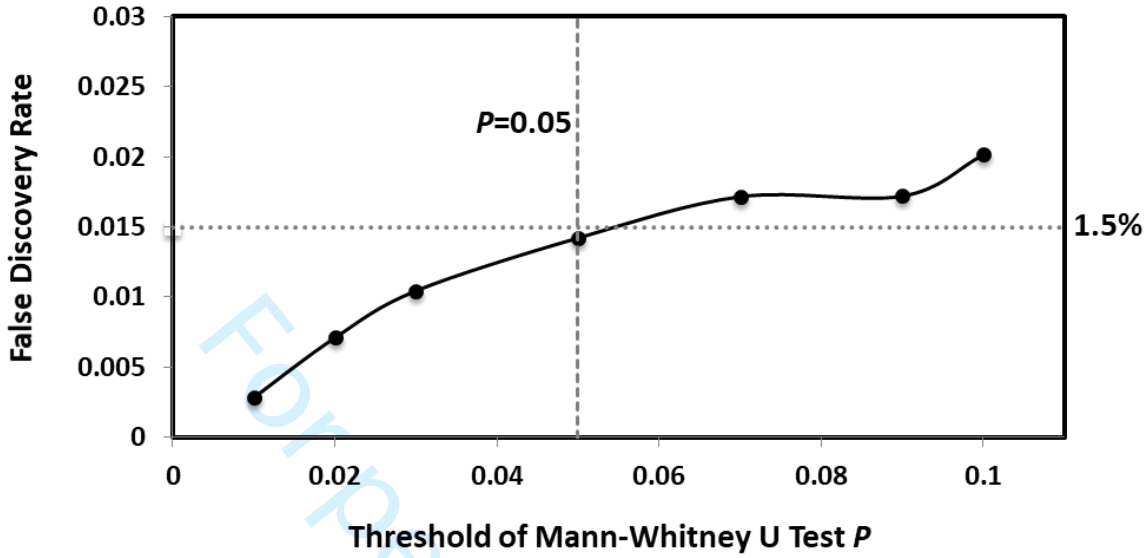



**Fig. A.5.** False discovery rate (FDR) analysis of the metabolic pathways significantly associated with PTB. Mann-Whitney U test *P* measured the difference in metabolite serological abundances between full-term pregnancies and pregnancies ending in PTB. Only metabolites with a Mann-Whitney U test *P* lower than the threshold were selected as part of the significant pathways. FDR was estimated by a permutation-based method (permutation N=1000).

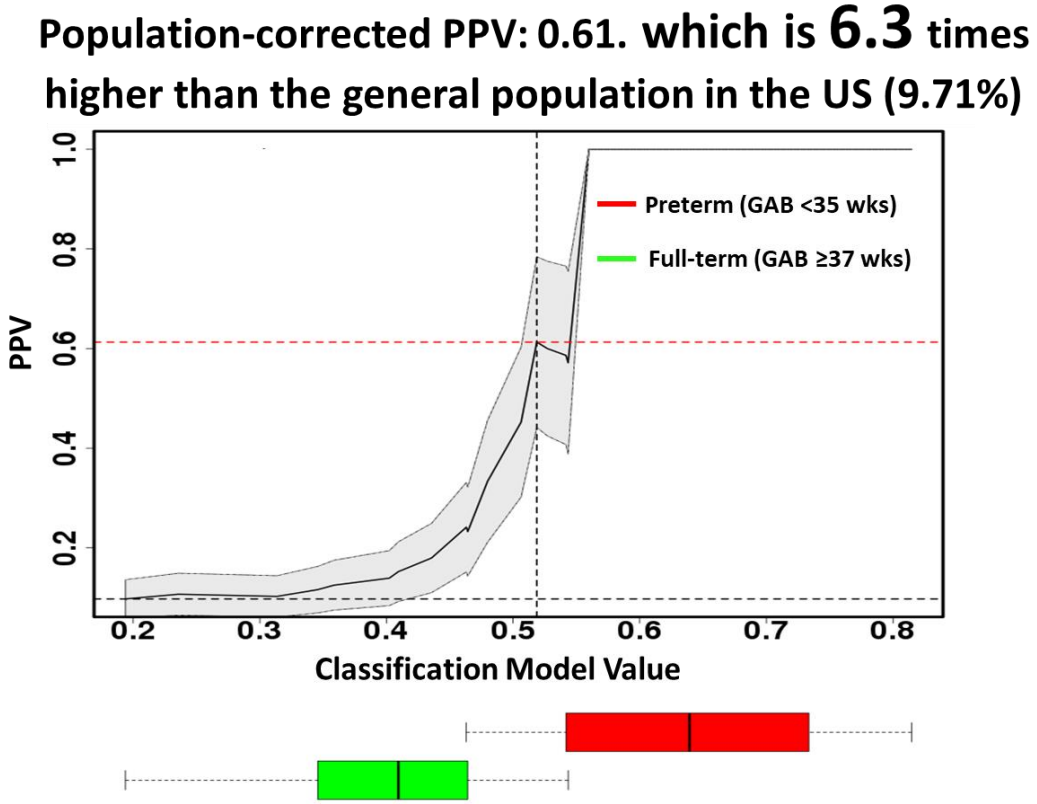



**Fig. A.6**. Stratification of patients by the classification model prediction on the UAB cohort. PPV was corrected by bootstrapping the full-term patients to reach the population PTB prevalence of 9.71% on singleton births. Two horizontal dashed lines represent the population mean of PTB risk that is 9.71% (black) and the PPV (= 0.61; red) at the high-risk cutoff. The grey dashed line indicates the high-risk cutoff value (= 0.52). The grey area represents the 95% confidence interval of the PPV. The box plot at the bottom shows the classification model value distribution stratified by the samples. GAB: gestational age at birth. wks: weeks' GA.

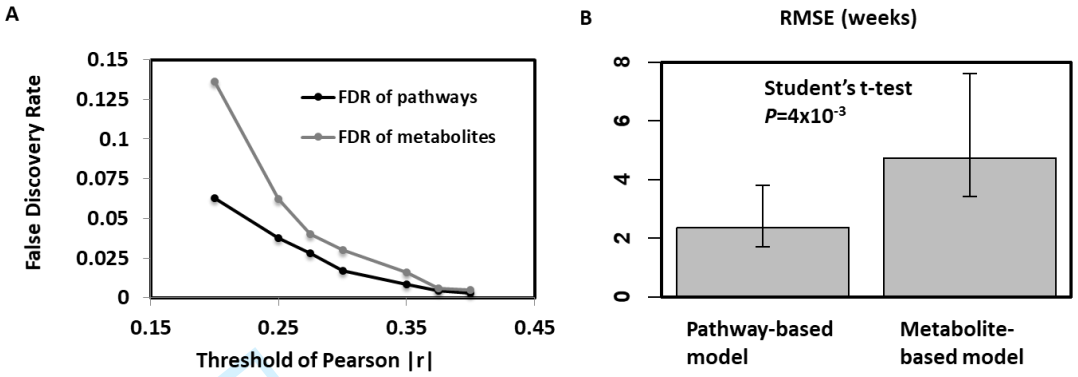



**Fig. A.7.** (A) False discovery rate (FDR) analysis of the metabolites and metabolic pathways significantly associated with the GA in full-term pregnancies. Pearson |r| was calculated as the correlation between metabolite serological abundance and GA. Only the metabolites with a Pearson |r| higher than the threshold (=0.35) would be selected as part of the significant pathways. FDR was estimated by a permutation-based method (permutation N=1000). (B) A comparison of RMSE of the GA estimation model trained by pathways and the model trained by metabolites. All metabolites had a Pearson |r|>0.35. RMSE was measured with the full-term samples of the validation (UAB) cohort.

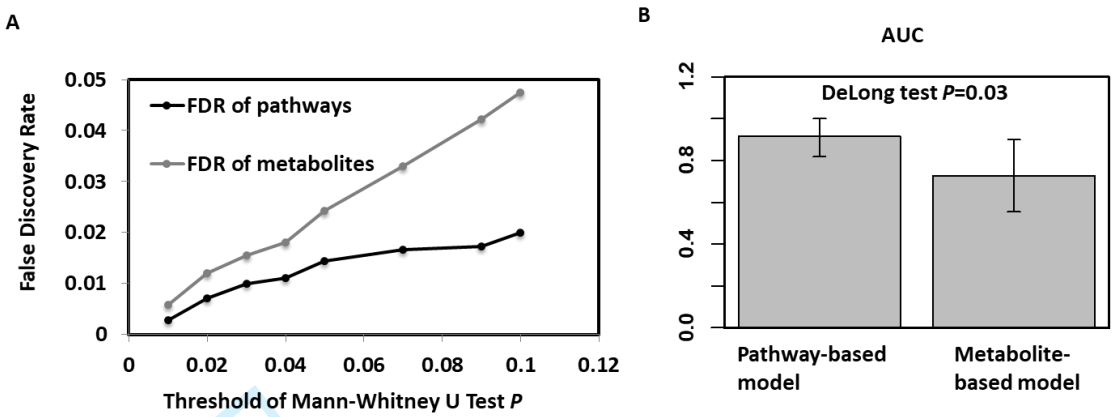



**Fig. A.8.** (A) False discovery rate (FDR) analysis of the metabolites and metabolic pathways significantly associated with the PTB. Mann-Whitney U test *P* measured the difference in metabolite serological abundances between full-term pregnancies and pregnancies ending in PTB. Only the metabolites with a Mann-Whitney U test *P* lower than the threshold (=0.05) would be selected as part of the significant pathways. FDR was estimated by a permutation-based method (permutation N=1000). (B) A comparison of the AUC of the preterm birth classification model utilizing pathways and the model utilizing metabolites. All the metabolites had a Mann-Whitney U test $P < 0.05$. AUC was measured with the samples of the validation (UAB) cohort.

**Table A.1.** Sensitivity and specificity of the XGBoost model with respect to the cutoff point.

| Cutoff | Cohort | Sensitivity | Specificity | Number of preterm samples identified by the model |
|---|---|---|---|---|
| **0.4** | **SU** | 0.94 | 0.78 | 30 |
| | **UAB** | 0.95 | 0.31 | 21 |
| **0.5** | **SU** | 0.88 | 0.94 | 28 |
| | **UAB** | 0.86 | 0.85 | 19 |
| **0.6** | **SU** | 0.81 | 0.98 | 26 |
| | **UAB** | 0.59 | 1 | 13 |
| **0.7** | **SU** | 0.53 | 0.98 | 17 |
| | **UAB** | 0.32 | 1 | 7 |

**Text A.1 Metabolic compound selection, pathway computation, and model development**

*GA estimation*

Metabolites measured by targeted and untargeted MS were aggregated and filtered using Pearson correlation coefficient analyses in relation to GA. The remaining metabolites were mapped to pathways. The value of each pathway was calculated as the weighted sum of the normalized concentrations of metabolites on the pathway divided by the number of metabolites. The weight of each metabolite was the absolute value of the Pearson correlation coefficient in relation to GA. Metabolites having positive or negative coefficients were aggregated separately. That is, a pathway could have two values, one for metabolites positively correlated to GA, and the other for those negatively correlated to GA.

A supervised, cross-validated machine-learning technique XGBoost was developed with the pathway values of samples from full-term patients in the SU cohort. An ensemble of regression trees was generated to give a score estimating the GA. The model was validated on the UAB cohort. For a patient that had multiple samples, an 'integrated' GA estimate was calculated by shifting the GA estimates of every sample to a reference point for obtaining the median. Error distribution of GA estimation based on patients was calculated as the distribution of the differences between the 'integrated' GA estimates and the US measurement.

*PTB prediction*

Samples collected before 35 weeks' GA were selected to build the model to predict PTB. Mann–Whitney U test was used to select the initial candidate metabolites that were then mapped to pathways. The value of each pathway was calculated as the weighted sum of the normalized concentrations of metabolites on the pathway divided by the number of metabolites. The weight of each metabolite was the absolute value of the ratio of median of full-term samples to PTB samples. Like the GA estimation, pathways could have two values that depended on the ratio of median greater or less than 1. An XGBoost model was developed utilizing samples from the SU cohort and validated with the UAB cohort.

**Text A.2 Metabolite model vs. IBP4/SHBG in predicting PTB**

We conducted ELISA tests to evaluate the IBP4/SHBG signature, a predictor that was validated in a prospective study as a predictor of spontaneous PTB. Commercial kits Human IGFBP4 ELISA Kit (Abcam, Burlingame, CA, USA) and Human SHBG Quantikine ELISA Kit (R&D System Inc.) were used. AUC of the predictor was calculated in different GA intervals and with different maternal BMI values, and was compared to the performance of the metabolic model.

With a BMI of >22 and ≤37 kg/m$^2$, the AUC values of the IBP4/SHBG predictor peaked at 15–20 weeks' GA (SU: 0.833; UAB: 1), and dropped rapidly after 20 weeks (Figure A below). The AUC values were lower with extreme BMI (0.7 at BMI ≤20 kg/m$^2$ and 0.63 at BMI >27 kg/m$^2$; see Figure B below). These findings are consistent with the previous validation study. Compared with the IBP4/SHBG predictor, the metabolic model has a more stable AUC performance over the gestation and different BMI values in SU ($P$ = 0.03). In UAB at >18 weeks' GA, the AUC of IBP4/SHBG dropped from 0.6 to 0.3, while the AUC of the metabolic model was above 0.8.

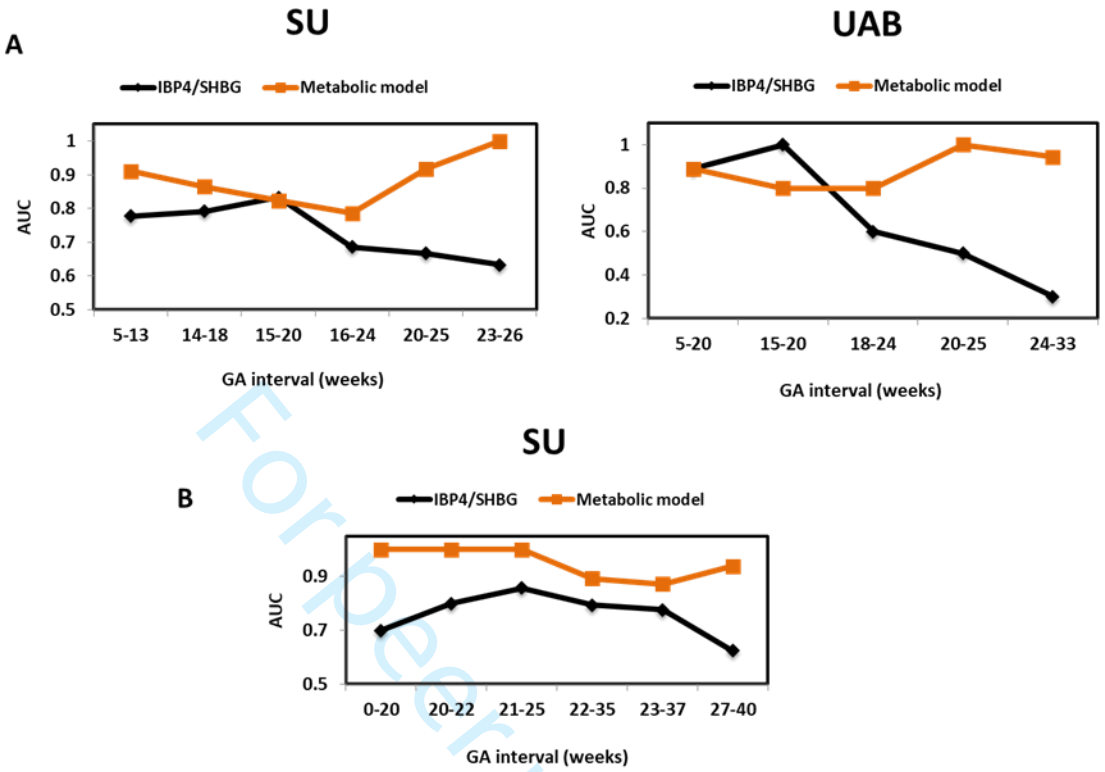



Figure. The performance of the IBP4/SHBG predictor and the metabolic model. The results are stratified by the GA intervals with a BMI at 22–37 kg/m$^2$ (A), and by BMI values with a GA interval of 5–20 weeks (B).

STROBE Statement—checklist of items that should be included in reports of observational studies

| | Item No | Recommendation | Page No |
|---|---|---|---|
| **Title and abstract** | 1 | (*a*) Indicate the study's design with a commonly used term in the title or the abstract | 1 |
| | | (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found | 2-3 |
| **Introduction** | | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported | 5 |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses | 6 |
| **Methods** | | | |
| Study design | 4 | Present key elements of study design early in the paper | 6 |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | 6,7 |
| Participants | 6 | (*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up<br>*Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls<br>*Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants | 6,7 |
| | | (*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed<br>*Case-control study*—For matched studies, give matching criteria and the number of controls per case | |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | 6,7,8 |
| Data sources/ measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | *7* |
| Bias | 9 | Describe any efforts to address potential sources of bias | 8 |
| Study size | 10 | Explain how the study size was arrived at | |
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | 7,8 |
| Statistical methods | 12 | (*a*) Describe all statistical methods, including those used to control for confounding | 8 |
| | | (*b*) Describe any methods used to examine subgroups and interactions | |
| | | (*c*) Explain how missing data were addressed | |
| | | (*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed<br>*Case-control study*—If applicable, explain how matching of cases and controls was addressed<br>*Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy | |
| | | (*e*) Describe any sensitivity analyses | 8 |

Continued on next page

| | | | |
|---|---|---|---|
| **Results** | | | |
| Participants | 13* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed | 9 |
| | | (b) Give reasons for non-participation at each stage | |
| | | (c) Consider use of a flow diagram | |
| Descriptive data | 14* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders | 9 |
| | | (b) Indicate number of participants with missing data for each variable of interest | |
| | | (c) *Cohort study*—Summarise follow-up time (eg, average and total amount) | 9 |
| Outcome data | 15* | *Cohort study*—Report numbers of outcome events or summary measures over time | *10* |
| | | *Case-control study*—Report numbers in each exposure category, or summary measures of exposure | |
| | | *Cross-sectional study*—Report numbers of outcome events or summary measures | |
| Main results | 16 | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included | 11-12 |
| | | (*b*) Report category boundaries when continuous variables were categorized | 11-12 |
| | | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period | |
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses | 12 |
| **Discussion** | | | |
| Key results | 18 | Summarise key results with reference to study objectives | 13 |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias | 16 |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence | 14-15 |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results | 16 |
| **Other information** | | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based | 18 |

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

# A metabolic clock as noninvasive blood tests of preterm birth and for gestational age assessment: a two-center retrospective study in the US



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2020-040647.R1 |
| Article Type: | Original research |
| Date Submitted by the Author: | 19-Oct-2020 |
| Complete List of Authors: | Sylvester, Karl; Stanford University<br>Hao, Shiying; Stanford University<br>You, Jin; Stanford University<br>Zheng, Le; Stanford University<br>Tian, Lu; Stanford University<br>Yao, Xiaoming; West China Hospital of Medicine<br>Mo, Lihong; University of California San Francisco Fresno Center for Medical Education and Research<br>Ladella, Subhashini; University of California San Francisco Fresno Center for Medical Education and Research<br>Wong, Ronald; Stanford University<br>Shaw, Gary M.; Stanford University<br>Stevenson, David; Stanford University<br>Cohen, Harvet; Stanford University<br>Whitin, John; Stanford University<br>McElhinney, Doff ; Stanford University<br>Ling, Xuefeng; Stanford University, |
| Primary Subject Heading: | Obstetrics and gynaecology |
| Secondary Subject Heading: | Paediatrics |
| Keywords: | Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, OBSTETRICS |
| | |

*I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.*

*The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.*

*Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.*

# A metabolic clock as noninvasive blood tests of preterm birth and for gestational age assessment: a two-center retrospective study in the US

Karl G. SYLVESTER, MD[1*†], Shiying HAO, PhD[2,3*], Jin YOU, PhD[1*], Le ZHENG, PhD[2,3], Lu TIAN[4], Xiaoming YAO, MD, PhD[5], Lihong MO, MD[6], Subhashini LADELLA, MD[6], Ronald J. WONG, BA[7], Gary M. SHAW, DrPh[7], David K. STEVENSON, MD[7], Harvey J. COHEN, MD, PhD[7], John C. WHITIN, PhD[7], Doff B. MCELHINNEY, MD[2,3], Xuefeng B. LING, PhD[1,3†]

[1]Department of Surgery, Stanford University School of Medicine, Stanford, CA, USA.
[2]Department of Cardiothoracic Surgery, Stanford University School of Medicine, Stanford, CA, USA.
[3]Clinical and Translational Research Program, Betty Irene Moore Children's Heart Center, Lucile Packard Children's Hospital, Palo Alto, CA, USA.
[4]Department of Health Research and Policy, Stanford University, Stanford, CA, USA.
[5]Translational Medicine Laboratory, West China Hospital, Chengdu, China.
[6]Department of Obstetrics and Gynecology, University of California San Francisco-Fresno, Fresno, CA, USA.
[7]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA.

[†]Correspondence: Xuefeng B. Ling, Department of Surgery, Stanford University, bxling@stanford.edu; Karl G. Sylvester, Department of Surgery, Stanford University, karls@stanford.edu
[*]Co-first authors

**Word count:** 293 for Abstract; 3508 for Main text.

## ABSTRACT

**Objectives** The aim of this study was to develop a single blood test that could determine gestational age and estimate the risk of preterm birth by measuring serum metabolites. We hypothesized that serial metabolic modeling of serum analytes throughout pregnancy could be used to describe fetal gestational age and project preterm birth with a high degree of precision.

**Study design** A retrospective cohort study

**Setting** Two medical centers from US

**Participants** Thirty-six patients (20 full-term, 16 preterm) enrolled at Stanford University were used to develop gestational age and preterm birth risk algorithms, 22 patients (9 full-term, 13 preterm) enrolled at the University of Alabama were used to validate the algorithms.

**Outcome measures** Maternal blood was collected serially throughout pregnancy. Metabolic datasets were generated using mass spectrometry.

**Results** A model to determine gestational age was developed ($R^2 = 0.98$) and validated ($R^2 = 0.81$). 66.7% of the estimates fell within ± 1 week of ultrasound results during model validation. Significant disruptions from full-term pregnancy metabolic patterns were observed in preterm pregnancies ($R^2 = -0.68$). A separate algorithm to predict preterm birth was developed utilizing a set of 10 metabolic pathways that resulted in an area under the curve of 0.96 and 0.92, a sensitivity of 0.88 and 0.86, and a specificity of 0.96 and 0.92 during development and validation testing, respectively.

**Conclusions** In this study metabolic profiling was used to develop and test a model for determining gestational age during full-term pregnancy progression, and to determine risk of preterm birth. With additional patient validation studies, these algorithms may be used to identify at-risk pregnancies prompting alterations in clinical care, and to gain biologic insights into the pathophysiology of preterm birth. Metabolic pathway-based pregnancy modeling is a novel modality for investigation and clinical application development.

**Keywords:** Metabolic, gestational age, preterm birth, pathway

## Strengths and limitations of this study

- The insensitivity of the prediction model to gestational age (GA) window of sample collection increases its flexibility and opportunity for potential clinical use.
- This study is among the first to propose a pathway-based computational methodology to estimate GA and predict preterm birth.
- The overall cohort size is modest, and the distribution of sampling time are different between patients and cohorts.
- It is a retrospective study; a larger prospective cohort study is necessary before applying the estimates and prediction to a broader population for clinical utility.

## INTRODUCTION

Gestational age (GA) dating is a core element of standard prenatal care [1-4]. Prenatal ultrasound (US) is an established modality for estimating GA, monitoring fetal growth, and screening for fetal anomalies [5]. According to the policy statement of the Committee on Obstetric Practice, the American Institute of Ultrasound in Medicine, and the Society for Maternal-Fetal Medicine, a pregnancy is considered optimally dated through a combination of last menstrual period (LMP) and an accurate US obtained prior to 22 0/7 weeks [6]. Accordingly, LMP is dependent on maternal recall and many pregnancies do not present for a first prenatal US evaluation until the second or third trimester. Thus, there is a need for a molecular method that would complement the potential shortcomings of LMP recall and US dating outside the first trimester. Moreover, it is possible that molecular pregnancy dating will provide greater resolution to pregnancy risk then current information based on calendar dating (LMP) and anthropometrics (US). Although experience is accumulating with the use of second and third trimester US for an estimation of risk of preterm birth (PTB) [7-9], to date these measures have not been widely adopted, are subject to user experience and have reported variable performance characteristics. The availability and expertise of US in disadvantaged areas is limited [10]. Therefore, there is a need to develop an alternative measure of fetal progression to estimate GA and pregnancy risk in a variety of settings and especially when US and LMP dates are unavailable or unreliable.

Compared with imaging methodologies, blood-based molecular testing may provide a more reproducible and precise modality in clinical applications for the frequent monitoring of health status and detection of early signs of disease. Genomic, gene

expression, protein, and metabolite profiles measured in human blood have been increasingly utilized for the determination of disease risk and to gain disease specific pathophysiology insight. Attempts at estimating GA using molecular adaptations have included modeling of RNA, protein, or immune cell changes, and most recently metabolites in maternal or newborn blood [11-17]. Similarly, risk prediction of PTB in clinical settings is currently primarily based on maternal history. Biomarkers have been suggested from genetic and proteomic analyses, but less effort has been focused on understanding maternal metabolic signatures of pregnancy [18-24].

In this study, we hypothesized that longitudinal metabolic profiling of pregnancy reflects the temporal progression of fetal development with a high degree of precision. Moreover, we posited that if a normal pregnancy progression profile could be defined in metabolic terms, then aberrations from the normal profile may identify a pregnancy at risk for PTB. Our findings suggest that composite metabolic panel modeling may serve as a reproducible and precision approach to GA dating of pregnancy and prediction of PTB.

## MATERIALS AND METHODS

### Definition

In this study, a full-term pregnancy was defined as a pregnancy ending with a delivery at ≥ 37 weeks. PTB was defined by delivery at < 35 weeks GA in order to make a complete separation from the full-term subjects.

### Study design

The study was conducted in two phases: (1) modeling to devise a metabolite-based estimation of GA during full-term pregnancies; and (2) modeling to devise a metabolic

panel predictive of PTB (Fig. 1). In this study, the 'gold' standard of GA was US measurement based on the crown-rump length at the first trimester [25]. Serum samples were collected in the 1st, 2nd, or 3rd trimester during pregnancy for each individual woman. Each participant had 1 to 4 time-points collected prior to delivery. Samples were provided by Stanford Hospital and Clinics (SU) and the University of Alabama (UAB). Metabolic concentrations in each sample were measured by targeted and untargeted mass spectrometry (MS) analysis. Models that estimated GA or predicted PTB were developed using the SU cohort and validated using the UAB cohort. The study was approved by the Institutional Review Board of both sites (Protocol #21956). All samples were collected after informed consent was obtained. All statistical analyses were done in R software.

**Targeted and global MS analysis**

Samples of full-term and preterm patients as well as quality control (QC) samples were injected into the MS. Targeted MS analysis was done through flow injection methods by using Ultimate 3000 Ultra-High-Performance Liquid Chromatography (UHPLC) system and Quantiva Triple Quadrupole Mass Spectrometer. Global (i.e. untargeted) MS analysis was done by using a Vanquish UHPLC system coupled to a Q Exactive plus mass spectrometer and Q Exactive HF hybrid quadrupole-Orbitrap mass spectrometer.

**Data preprocessing and metabolic identification**

A data pre-processing procedure was conducted to convert the raw data generated by MS analysis into a matrix of relative concentrations of metabolites versus samples [26]. This procedure was done by R package. Metabolic values in each sample were then normalized by the median values measured with QC samples to reduce the batch effects.

Compounds detected by untargeted analyses were matched to metabolites in the Human Metabolome Database by putative identification [27]. Accurate mass was used for the mapping. Metabolites were mapped to pathways using Kyoto Encyclopedia of Genes and Genomes (KEGG) and Human Metabolome Database (HMDB). Only endogenous pathways were considered.

**Metabolic compound selection, pathway computation, and model development**

Metabolites measured by targeted and untargeted MS were aggregated and filtered. The remaining metabolites were mapped to pathways. The value of each pathway was calculated as the weighted sum of the normalized concentrations of metabolites on the pathway divided by the number of metabolites. An XGBoost model was developed with the pathway values of samples from full-term patients to estimate the GA. R-squared ($R^2$; goodness-of-fit of the model), root-mean-square error (RMSE), and error distribution were calculated to evaluate the model performance. A second XGBoost model was developed to predict PTB. To evaluate the model performance, Mann–Whitney U tests were used to compare the distribution of final predictive estimates, i.e., XGBoost model values, on full-term and PTB samples. Additional details of model development were described in Text A.1. ELISA tests were conducted on the SU and UAB cohorts to evaluate the insulin-like growth factor-binding protein 4 (IBP4)/sex hormone-binding globulin (SHBG) signature, a predictor that was validated in a prospective study as a predictor of spontaneous PTB [19]. Serum concentrations were measured using commercial kits Human IGFBP4 ELISA Kit (Abcam, Burlingame, CA, USA) and Human SHBG Quantikine ELISA Kit (R&D System Inc.). Results were compared with our metabolic model.

## Patient and Public Involvement statement

This retrospective research was done without patient involvement. Patients were not invited to comment on the study design and were not consulted to develop patient relevant outcomes or interpret the results. Patients were not invited to contribute to the writing or editing of this document for readability or accuracy.

## RESULTS

### Samples

As shown in Fig. 2, the SU cohort had 20 full-term pregnancies with 57 blood samples (17, 32, and 8 collected in the 1st, 2nd, and 3rd trimesters, respectively) and 16 preterm pregnancies with 32 blood samples (9, 19, and 4 collected in the 1st, 2nd, and 3rd trimesters, respectively). The UAB cohort had 9 full-term pregnancies with 13 blood samples (8 and 5 in the 2nd, and 3rd trimesters, respectively) and 13 preterm pregnancies with 22 blood samples (4 and 18 in the 1st and 2nd trimesters, respectively). In the SU cohort, 2 (12.5%) were extremely preterm (< 28 weeks), and 5 (31.3%) were very preterm (28–31 weeks). In the UAB cohort, 6 (46.2%) were extremely preterm, and 3 (23.1%) were very preterm. Our SU and UAB cohorts were assembled: no complications of pregnancy were included;all deliveries were singleton; and all PTB were spontaneous. Demographics of the two cohorts are shown in Table 1.

**Table 1.** Maternal characteristics in SU and UAB cohorts

| | SU | | | UAB | | | SU vs. UAB |
|---|---|---|---|---|---|---|---|
| **Characteristic** | **Full-term** | **Preterm** | ***P*** | **Full-term (n =** | **Preterm** | ***P*** | ***P*** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | **(n = 20)** | **(n = 16)** | | **9)** | **(n = 13)** | |
| Race, n (%) | | | **<0.001** | | 0.5 | **<0.001** |
| Asian | 0 | 1 (6.3) | | 0 | 0 | |
| White | 20 (100) | 5 (31.3) | | 0 | 2 (15.4) | |
| Black | 0 | 1 (6.3) | | 9 (100) | 10 (76.9) | |
| American Indian | 0 | 2 (12.5) | | 0 | 0 | |
| Pacific Islander | 0 | 1 (6.3) | | 0 | 0 | |
| Other/unknown | 0 | 6 (37.5) | | 0 | 1 (7.7) | |
| Hispanic, n (%) | 0 | 8 (50) | **<0.001** | 0 | 1 (7.7) | 0.9 | 0.1 |
| Maternal Age, year, mean (SD) | 31.9 (4.8) | 29.8 (7.5) | 0.3 | 25.6 (5.0) | 27.5 (4.5) | 0.4 | **0.008** |
| Gestational age at delivery, weeks, median (IQR) | 39.5 (39,41) | 32 (30,33) | **<0.001** | 38 (37,39) | 28 (26,32) | **<0.001** | **0.01** |
| Having previous pregnancy, n (%) | 9 (45) | 6 (37.5) | 0.7 | 9 (100) | 13 (100) | 0.4 | **<0.001** |
| BMI, kg/m$^2$, median (IQR) | 22.3 (20.2,24.7) | 27.6 (23.4,33.9) | **0.003** | 30.4 (22.3,33.1) | 26.5 (22.6,36.5) | 0.8 | 0.06 |
| History of PTB, n (%) | 3 (15) | 8 (50) | **0.03** | 7 (77.8) | 13 (100) | 0.2 | **<0.001** |

## LC-MS/MS metabolomics

The study targeted 315 metabolites by LC-MS/MS, including 13 categories: acyl-carnitine (11, 3.5%), amino acid (9, 2.9%), fatty acid (6, 1.9%), ceramide (12, 3.8%), ceramide 1-phosphate (8, 2.5%), galactosylceramide (5, 1.6%), phosphatidyl acid (15,

4.8%), phosphatidylethanolamine (52, 16.5%), phosphatidylglycerol (5, 1.6%), phosphatidylinositol (11, 3.5%), phophatidylcholine (130, 41.3%), cholesteryl ester (16, 5.1%), and sphingomyelin (35, 11.1%). The study also identified 1627 positively-and 295 negatively-charged compounds through untargeted analyses. Together these formed the initial set of 2237 compounds.

**Feature selection of GA estimation modeling**

Of the 2237 compounds, 118 had an absolute Pearson correlation coefficient of > 0.35 with GA. The cutoff of ± 0.35 was selected based on the false discovery rate (FDR) values of the mapped pathways < 1% (Fig. A.1). The 118 compounds were mapped to 89 pathways, 33 of which were selected by the XGBoost model. The normalized value of each pathway varied over the course of gestation (Fig. A.2). Univariate analysis of the 33 pathways is shown in Fig. A.3, and the top 10 pathways in the model is depicted in Fig. 3. The top 10 pathways included those associated in the metabolisms of: glycerophospholipid, arginine and proline, thiamine, purine, butanoate, galactose, sulfur, phenylalanine, and C5-branched dibasic acid.

**Performance of GA estimation**

The performance of GA estimates on full-term samples was similar in the development phase (SU cohort, $R^2 = 0.98$, RMSE = 1.09) and the validation phase (UAB cohort, $R^2 = 0.81$, RMSE = 2.36) (Fig. 4). In our validation testing, 66.7% of the estimates were within ± 1 week of the US results (Fig. A.4).

Intriguingly, model performance significantly deteriorated when applied to samples from PTB pregnancies ($R^2 = -0.68$ and RMSE = 6.6 in validation; see Fig. 4). It suggested that

the relationships between metabolic parameters and full-term pregnancies were not maintained in PTB pregnancies. Furthermore, such disruptions were notable as early as 10 weeks' GA (Fig. 4) or early to mid-gestation. These findings prompted the development of a metabolic-based model of PTB estimation.

**Performance of PTB prediction**

Samples collected before 35 weeks' GA were used to develop a model that differentiated PTB pregnancies from those full-term. As before, the model was developed with the SU cohort that had 20 full-term (54 samples) and 16 preterm (32 samples) pregnancies, and was validated with the UAB cohort that had 9 full-term (13 samples) and 13 preterm (22 samples) pregnancies. In total, 148 metabolic compounds (with Mann-Whitney U test $P < 0.05$) were mapped to 66 pathways (FDR < 1.5%; see Fig. A.5). Further model development selected 10 pathways as strong predictors covering the metabolisms of glycerophospholipid, sphingolipid, taurine and hypotaurine, arachidonic acid, secondary bile acid biosynthesis, glycerolipid, cysteine and methionine, tryptophan, and arginine and proline (Fig. 5).

The level of prediction accuracy was maintained in the validation cohort ($P = 5\text{x}10^{-5}$, area under the curve [AUC] = 0.92; see Fig. 6). The prevalence-corrected positive predictive values (PPVs) across model values (*i.e.* scores) were plotted based on the PTB prevalence in Alabama in 2018 (12.5%; see Fig. A.6). A threshold value of 0.52 was selected as a high-risk threshold for PTB, which was associated with a PPV of 0.70, a relative risk (RR) of 5.6 compared to the United States population baseline (= 0.70/12.5%), a sensitivity of 0.86 (19 of 22), and a specificity of 0.92 (12 of 13; Fig. 7). The sensitivities and specificities with cutoff values are shown in Table A.1.

In the validation cohort, 12 of 13 full-term samples and 19 of 22 preterm samples were classified correctly. The misclassified full-term sample was from a mother that delivered at 37 weeks' GA. The 19 correctly classified PTB samples were from 13 PTB pregnancies. Of the 13 pregnancies, 9 were identified as high risk at or earlier than 16 weeks' GA. The median gap between the time of identification and the delivery was 11 weeks' GA (IQR: 8, 15.5).

To determine the performance of our metabolic model against existing models, a comparison between the metabolic PTB risk model and the commercially available IBP4/SHBG PTB test was performed and summarized in Text A.2 and Fig. A.7.

**Metabolite-based model and pathway-based model: a comparison**

To determine the effectiveness of model performance based upon robustness of biologic features, we compared model performance using pathway or individual metabolite as selected features in estimating GA and predicting PTB. The performance of the pathway-based models were significantly better than the metabolite-based models, with a lower RMSE (Student's t-test $P = 4x10^{-3}$; Fig. A.8) and a larger AUC (DeLong test $P = 0.03$; Fig. A.9).

## DISCUSSION

### Principal Findings

In this study, we report a panel of metabolic pathways measured in maternal serum that provides an estimation of GA over the course of a full-term pregnancy. A second and distinct set of metabolic pathways was also identified in maternal serum that could distinguish pregnancies ending with PTB (< 35 weeks) from full-term (≥ 37 weeks) with

a high degree of precision. The models were developed and validated using two independent cohorts from two different institutions in order to test the robustness of the biologic features driving the classifications. Intriguingly, PTB pregnancies do not demonstrate the same temporal relationship as term pregnancies upon metabolic modeling across gestation (Fig. 4). Indeed, PTB pregnancies demonstrate a marked departure from the term metabolic profile (Fig. 4) that is not only dramatic ($R^2$= 0.98 train and 0.81 test for term model; compared to $R^2$= 0.50 train and -0.68 test for PTB pregnancy in term model), but is also recognizable as early as 10 weeks' GA as determined by the current standard of US dating. Recognizing the metabolic pathway aberration of PTB pregnancies, a second model was developed using metabolic pathway analyses to quantify the risk of PTB prior to 35 weeks' GA. Once again, metabolic profiling proved to be robust in identifying PTB pregnancies with a high degree of sensitivity (AUC 0.96 training; AUC 0.92 testing) and precision (training PPV 0.93 (0.78-0.99); testing PPV 0.95 (0.75-1). Taken together, this study demonstrated a powerful new, reproducible methodology for monitoring pregnancy progression and identifying abnormal pregnancies.

**Clinical and Research Implications**

The potential clinical utility of developing a test for pregnancy monitoring is appealing. There is a need to develop a more robust method than LMP and an alternative to first trimester US that captures pregnancy progression, a complex relationship of fetal and placental growth, development, and function. To support these processes, there is a need for energy transfer between mother and fetus throughout gestation. We therefore reasoned that metabolic phenotyping would be ideally suited to capture this relationship.

Despite a modest cohort size, the results of metabolic modeling demonstrate a high degree of concordance with clinical standard US dating performed by experts as reflected by 66.7% of model estimates falling within ± 1 week of US results (Fig. A.4). Moreover, unlike the deterioration experienced with US dating of pregnancy, metabolic modeling was shown to achieve near equivalent performance in the 1st, 2nd, and 3rd trimesters, indicating the potential for broad clinical applicability that might achieve independence of reliance on accuracy of LMP or concordance among modality testing. The result of PTB prediction is equally robust demonstrating a high degree of precision. Beyond relying on clinical histories or self-reported symptoms, the model proposed here provides a molecular classification that may be more accurate than current methods and further reflect a comprehensive measure of aberrant pregnancy based on metabolic changes. In practice, clinicians could use the PTB prediction model to differentiate high- from low-risk patients. Low risk patients would then be subject to GA estimation panel testing, all from the same blood draw.

A distinct advantage of the PTB risk prediction developed in this study is that it has a wide window of sampling. Samples were collected broadly before 35 weeks' GA, which is wider than the window of other well-established biomarkers such as fetal fibronectin (between 24 and 34 weeks' GA) [20], IBP4/SHBG (19 to 21 weeks) [19], and inter-alpha-trypsin inhibitor heavy chain 4 protein (24 and 28 weeks) [18]. Relatively stable AUC levels were maintained throughout the diagnostic window (Text A.2). The insensitivity of the prediction model to GA at testing increases its flexibility and opportunity for potential clinical use. An additional advantage of the model herein is the ability for early identification of high-risk women. Although there is no standardized guideline for early-

gestation management of patients at risk of PTB delivery, metabolic modeling for PTB risk may provide a not previously possible opportunity for early gestation risk mitigation. Clinical trials have suggested that hormone treatment and maternal physical activity modifications applied between 16 to 37 weeks' GA reduced the PTB rate of women who were deemed at high risk due to a history of prior PTB delivery [28 29]. In many cases PTB can not be prevented, however any opportunity is deemed highly desirable for even a modest delay (1–2 weeks) in PTB or an enhanced ability to more accurately triage for delivery to centers with the capability to manage profoundly premature neonates [30-32].

This study is among the first to propose a pathway-based computational methodology to estimate GA and predict PTB. Metabolic pathways are linked to chemical functions, and the alteration or disruption of specific functions participate in disease phenotypes, facilitating the use of pathways to function as higher-level biomarkers of diseases [33]. The role of metabolic pathways in disease diagnosis has been explored in several preliminary clinical studies [34 35]. Pathway performance in differentiating patients with disease from healthy controls has been found to be effective compared to using individual metabolites [35]. Similarly, we found the pathway-based models had less variability and higher sensitivity than metabolite-based models that were developed using the same population. One plausible explanation for this observation may be attributed to the calculation of pathway values, which represents the sum of individual metabolites and thus may amplify association to outcome relationships. This hypothesis is supported by the FDR comparison (Fig. A.8 and A.9): pathway-based analysis had lower FDR values than metabolite models. This study adds to the exploration of the feasibility of using pathways for health monitoring and prediction.

In this study glycerophospholipid metabolism was identified as the most significant contributing pathway for both gestational age estimation and preterm birth prediction. Glycerophospholipids consist of fatty acid chains and have been previously cited as strong correlates to birth weight, pregnancy duration, and risk of preterm birth [36]. These same authors also found different polyunsaturated fatty acid components of glycerophospholipid had differential effects on fetal growth. Gao et al has reported a potential association between glycerophospholipid and labor timing in rodent models [37] [38]. The current study extends those prior observations through a quantitative assessment of the relationship between glycerophospholipid metabolism, gestational age and the risk of preterm birth. The leading effect of glycerophospholipid pathway metabolism in the current study was positive in both the assessment of gestational age and risk of preterm birth. These findings add further insight into the role of glycerophospholipid metabolism in human pregnancy. Other contributing pathways for preterm birth prediction such as sphingolipid metabolism, arachidonic acid metabolism, and arginine and proline metabolism were also found associated to preterm. Alterations in plasma sphingolipids were found in women who had spontaneous PTB [39]. Increase of arachidonic acid metabolism might correlate to bacteria activities that led to preterm labor [40]. Plasma level of arginine and citrulline was significantly lowered in preterm babies [41].

Taken together, the analysis of the leading pathways found to significantly contribute to the metabolic pregnancy modeling herein provide ample insights to deepen our understanding of pregnancy progression and may facilitate the identification and interpretation of potential therapeutic targets. Further, we speculate that the platform and approaches outlined herein may be extended to the interrogation of additional conditions

of pregnancy including abnormalities of placentation, gestational diabetes and fetal growth disturbances among others.

**Limitations**

This study has several limitations. First, the overall cohort size was modest, and pregnancies with delivery at 35 or 36 weeks were not included in the study. Second, blood samples were collected in a non-uniform manner with respect to GA timing and time of day. The time between two adjacent samples corresponding to the same patient varied. Third, the distribution of samples throughout pregnancy were different between patients and cohorts. In the SU cohort, none of the full-term patients had samples collected between 30 and 37 weeks. In the UAB cohort, none of the full-term patients had sampling in the 1st trimester, and none of the PTB patients had sampling in the 3rd trimester. Fourth, for methodologic reasons, not all serum analytes could be identified and mapped to known metabolites. Fifth, baseline characteristics of patients were not included in the analysis. Sixth, the study was retrospective, and the participants were solely from California and Alabama. A larger prospective cohort study with a reasonable ratio of full-term to preterm is necessary before applying the estimates and prediction to a broader population for clinical utility.

## CONCLUSION

The present study demonstrates that maternal serum based metabolic profiling is a highly sensitive and accurate method for determining GA and prediction of PTB. The pathway-based analysis supports the hypothesis of the orderly metabolic progression of pregnancy that can be reproducibly captured using metabolic profiling. The robustness of the modeling reinforces the potential appeal for further clinical development and as a

platform to investigate the pathophysiology associated with aberrant fetal development and pregnancy progression. This study is the first to report a single blood test for metabolic pathway-based determination of GA dating, and early detection of PTB risk.

**Acknowledgments:** The authors thank colleagues at the Stanford University Pediatric Proteomics group and the March of Dimes Prematurity Research Center at Stanford University for critical discussions.

**Funding:** This work was supported in part by the March of the Dimes Prematurity Research Center at Stanford University, and Stanford Child Health Research Institute. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Grant numbers are N/A.

**Conflict of Interest:** The authors report no conflict of interest.

**Author contributions:** XBL, KGS, and HJC contributed to concept development and design.

JY, RJW, and DKS contributed to the acquisition of data.

KGS, SH, LZ, XY, LT, LM, SL, RJW, GMS, DKS, JCW and DBM contributed to the analysis and interpretation of data.

KGS and SH drafted the manuscript.

JY, LZ, LT, XY, LM, SL, RJW, GMS, DKS, HJC, JCW, DBM, and XBL critically revised the manuscript.

All the authors gave final approval of the version to be submitted and agreed to be accountable for all aspects of the work.

**Data and materials availability:** The datasets used and/or analyzed in this study are available upon request to the corresponding author.

## References

1. Brownfoot FC, Gagliardi DI, Bain E, et al. Different corticosteroids and regimens for accelerating fetal lung maturation for women at risk of preterm birth. *Cochrane Database Syst Rev* 2013(8):CD006764. doi: 10.1002/14651858.CD006764.pub3
2. Raju TN, Mercer BM, Burchfield DJ, et al. Periviable birth: executive summary of a Joint Workshop by the Eunice Kennedy Shriver National Institute of Child Health and Human Development, Society for Maternal-Fetal Medicine, American Academy of Pediatrics, and American College of Obstetricians and Gynecologists. *Journal of perinatology : official journal of the California Perinatal Association* 2014;34(5):333-42. doi: 10.1038/jp.2014.70
3. Vohr B. Long-term outcomes of moderately preterm, late preterm, and early term infants. *Clin Perinatol* 2013;40(4):739-51. doi: 10.1016/j.clp.2013.07.006
4. Pereira AP, Dias MA, Bastos MH, et al. Determining gestational age for public health care users in Brazil: comparison of methods and algorithm creation. *BMC Res Notes* 2013;6:60. doi: 10.1186/1756-0500-6-60
5. Peek MJ, Devonald KJ, Beilby R, et al. The value of routine early pregnancy ultrasound in the antenatal booking clinic. *Aust N Z J Obstet Gynaecol* 1994;34(2):140-3.
6. Committee Opinion No 700: Methods for Estimating the Due Date. *Obstetrics and gynecology* 2017;129(5):e150-e54. doi: 10.1097/aog.0000000000002046 [published Online First: 2017/04/21]
7. Roman A, Saccone G, Dude CM, et al. Midtrimester transvaginal ultrasound cervical length screening for spontaneous preterm birth in diamniotic twin pregnancies according to chorionicity. *European journal of obstetrics, gynecology, and reproductive biology* 2018;229:57-63. doi: 10.1016/j.ejogrb.2018.08.006 [published Online First: 2018/08/15]
8. Erkamp JS, Voerman E, Steegers EAP, et al. Second and third trimester fetal ultrasound population screening for risks of preterm birth and small-size and large-size for gestational age at birth: a population-based prospective cohort study. *BMC Med* 2020;18(1):63. doi: 10.1186/s12916-020-01540-x [published Online First: 2020/04/08]
9. Dziadosz M, Bennett TA, Dolin C, et al. Uterocervical angle: a novel ultrasound screening tool to predict spontaneous preterm birth. *American journal of obstetrics and gynecology* 2016;215(3):376.e1-7. doi: 10.1016/j.ajog.2016.03.033 [published Online First: 2016/03/29]
10. Jehan I, Zaidi S, Rizvi S, et al. Dating gestational age by last menstrual period, symphysis-fundal height, and ultrasound in urban Pakistan. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics* 2010;110(3):231-4. doi: 10.1016/j.ijgo.2010.03.030
11. Wilson K, Hawken S, Potter BK, et al. Accurate prediction of gestational age using newborn screening analyte data. *American journal of obstetrics and gynecology* 2016;214(4):513 e1-13 e9. doi: 10.1016/j.ajog.2015.10.017

12. Knight AK, Craig JM, Theda C, et al. An epigenetic clock for gestational age at birth based on blood methylation data. *Genome biology* 2016;17(1):206. doi: 10.1186/s13059-016-1068-z
13. Jelliffe-Pawlowski LL, Norton ME, Baer RJ, et al. Gestational dating by metabolic profile at birth: a California cohort study. *American journal of obstetrics and gynecology* 2016;214(4):511 e1-11 e13. doi: 10.1016/j.ajog.2015.11.029
14. Aghaeepour N, Ganio EA, McIlwain D, et al. An immune clock of human pregnancy. *Sci Immunol* 2017;2(15) doi: 10.1126/sciimmunol.aan2946
15. Liang L, Rasmussen M-LH, Piening B, et al. Metabolic Dynamics and Prediction of Gestational Age and Time to Delivery in Pregnant Women. *Cell* 2020;181(7):1680-92.e15. doi: https://doi.org/10.1016/j.cell.2020.05.002
16. Ngo TTM, Moufarrej MN, Rasmussen MH, et al. Noninvasive blood tests for fetal development predict gestational age and preterm delivery. *Science* 2018;360(6393):1133-36. doi: 10.1126/science.aar3819
17. Aghaeepour N, Lehallier B, Baca Q, et al. A proteomic clock of human pregnancy. *American journal of obstetrics and gynecology* 2018;218(3):347 e1-47 e14. doi: 10.1016/j.ajog.2017.12.208
18. Esplin MS, Merrell K, Goldenberg R, et al. Proteomic identification of serum peptides predicting subsequent spontaneous preterm birth. *American journal of obstetrics and gynecology* 2011;204(5):391 e1-8. doi: 10.1016/j.ajog.2010.09.021
19. Saade GR, Boggess KA, Sullivan SA, et al. Development and validation of a spontaneous preterm delivery predictor in asymptomatic women. *American journal of obstetrics and gynecology* 2016;214(5):633 e1-33 e24. doi: 10.1016/j.ajog.2016.02.001
20. Peaceman AM, Andrews WW, Thorp JM, et al. Fetal fibronectin as a predictor of preterm birth in patients with symptoms: a multicenter trial. *American journal of obstetrics and gynecology* 1997;177(1):13-8.
21. Strauss JF, 3rd, Romero R, Gomez-Lopez N, et al. Spontaneous preterm birth: advances toward the discovery of genetic predisposition. *American journal of obstetrics and gynecology* 2018;218(3):294-314 e2. doi: 10.1016/j.ajog.2017.12.009
22. Virgiliou C, Gika HG, Witting M, et al. Amniotic Fluid and Maternal Serum Metabolic Signatures in the Second Trimester Associated with Preterm Delivery. *Journal of proteome research* 2017;16(2):898-910. doi: 10.1021/acs.jproteome.6b00845
23. Li J, Lu YP, Reichetzeder C, et al. Maternal PCaaC38:6 is Associated With Preterm Birth - a Risk Factor for Early and Late Adverse Outcome of the Offspring. *Kidney Blood Press Res* 2016;41(3):250-7. doi: 10.1159/000443428
24. Hawdon JM, Ward Platt Mp Fau - Aynsley-Green A, Aynsley-Green A. Patterns of metabolic adaptation for preterm and term infants in the first neonatal week. (1468-2044 (Electronic))
25. Robinson HP. Sonar measurement of fetal crown-rump length as means of assessing maturity in first trimester of pregnancy. (0007-1447 (Print))
26. Dunn WB, Broadhurst D, Begley P, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid

chromatography coupled to mass spectrometry. *Nat Protoc* 2011;6(7):1060-83. doi: 10.1038/nprot.2011.335
27. Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 2007;3(3):211-21. doi: 10.1007/s11306-007-0082-2
28. Meis PJ, Klebanoff M, Thom E, et al. Prevention of recurrent preterm delivery by 17 alpha-hydroxyprogesterone caproate. *The New England journal of medicine* 2003;348(24):2379-85. doi: 10.1056/NEJMoa035140
29. Evenson KR, Siega-Riz AM, Savitz DA, et al. Vigorous leisure activity and pregnancy outcome. *Epidemiology* 2002;13(6):653-9. doi: 10.1097/01.EDE.0000021463.45041.95
30. McIntire DD, Leveno KJ. Neonatal mortality and morbidity rates in late preterm births compared with births at term. *Obstetrics and gynecology* 2008;111(1):35-41. doi: 10.1097/01.AOG.0000297311.33046.73
31. Henderson-Smart DJ. The effect of gestational age on the incidence and duration of recurrent apnoea in newborn babies. *Aust Paediatr J* 1981;17(4):273-6.
32. Khashu M, Narayanan M, Bhargava S, et al. Perinatal outcomes associated with preterm birth at 33 to 36 weeks' gestation: a population-based cohort study. *Pediatrics* 2009;123(1):109-13. doi: 10.1542/peds.2007-3743
33. Lee DS, Park J, Kay KA, et al. The implications of human metabolic network topology for disease comorbidity. *Proceedings of the National Academy of Sciences of the United States of America* 2008;105(29):9880-5. doi: 10.1073/pnas.0802208105
34. Baumgartner C, Bohm C, Baumgartner D, et al. Supervised machine learning techniques for the classification of metabolic disorders in newborns. *Bioinformatics* 2004;20(17):2985-96. doi: 10.1093/bioinformatics/bth343
35. Huang S, Chong N, Lewis NE, et al. Novel personalized pathway-based metabolomics models reveal key metabolic pathways for breast cancer diagnosis. *Genome medicine* 2016;8(1):34. doi: 10.1186/s13073-016-0289-9
36. Grootendorst-van Mil NH, Tiemeier H, Steenweg-de Graaff J, et al. Maternal plasma n-3 and n-6 polyunsaturated fatty acids during pregnancy and features of fetal health: Fetal growth velocity, birth weight and duration of pregnancy. *Clin Nutr* 2018;37(4):1367-74. doi: 10.1016/j.clnu.2017.06.010
37. Menon R, Bonney EA, Condon J, et al. Novel concepts on pregnancy clocks and alarms: redundancy and synergy in human parturition. *Human reproduction update* 2016;22(5):535-60. doi: 10.1093/humupd/dmw022
38. Gao L, Rabbitt EH, Condon JC, et al. Steroid receptor coactivators 1 and 2 mediate fetal-to-maternal signaling that initiates parturition. *The Journal of clinical investigation* 2015;125(7):2808-24. doi: 10.1172/JCI78544
39. Morillon A-C, Yakkundi S, Thomas G, et al. Association between phospholipid metabolism in plasma and spontaneous preterm birth: a discovery lipidomic analysis in the cork pregnancy cohort. *Metabolomics* 2020;16(2):19. doi: 10.1007/s11306-020-1639-6
40. Bennett PR, Rose MP, Myatt L, et al. Preterm labor: Stimulation of arachidonic acid metabolism in human amnion cells by bacterial products. *American*

*journal of obstetrics and gynecology* 1987;156(3):649-55. doi: https://doi.org/10.1016/0002-9378(87)90070-6

41. Contreras MT, Gallardo MJ, Betancourt LR, et al. Correlation between plasma levels of arginine and citrulline in preterm and full-term neonates: Therapeutical implications. *Journal of Clinical Laboratory Analysis* 2017;31(6):e22134. doi: 10.1002/jcla.22134

**Figure Legends**

**Fig. 1.** Study design. Models were developed separately to estimate gestational age during full-term pregnancy, and to predict the risk of preterm birth. Both models were developed with the SU cohort and validated with the UAB cohort.

**Fig. 2.** Cohort construction. Each line represents an individual patient. Diamond and triangle markers indicate sample collection dates and delivery dates, respectively. The red dashed line represents 37 weeks' gestational age.

**Fig. 3.** The importance of the top 10 metabolic pathways in the gestational age estimation model. Pathways either positively or negatively correlated gestational age.

**Fig. 4.** Gestational age estimates of the gestational age model with the SU ($R^2$=0.98, RMSE=1.09 weeks) and UAB cohorts ($R^2$ = 0.81, RMSE = 2.36 weeks).

**Fig. 5.** (A) Univariate analysis of the 10 metabolic pathways in the preterm birth prediction model. Odds ratio of each pathway was calculated. **P*<0.05, ***P*<0.01, ****P*<0.005. (B) The importance of the metabolic pathways in the preterm birth prediction model. Pathways were either up- or down-regulated in relation to preterm birth.

**Fig. 6.** (A) Prediction of preterm birth risk grouped by full-term and preterm birth patients (top) and over the course of gestation (bottom). (B) AUC performance of the

prediction in SU and UAB cohorts. *P* was calculated using Mann–Whitney U test. wks: weeks' gestational age.

**Fig. 7.** Performance of the preterm birth prediction model. (A) A contingency table showing the number of samples in each category. (B) Sensitivity, specificity, PPV, and NPV together with the 95% confidence intervals.

## Appendix Captions

**Fig. A.1** False discovery rate (FDR) analysis of the metabolic pathways significantly associated with the GA in full-term pregnancies. Pearson |r| was calculated as the correlation between metabolite serological abundance and GA. Only the metabolites with a Pearson |r| higher than the threshold would be selected as part of the significant pathways. FDR was estimated by a permutation-based method (permutation N=1000).

**Fig. A.2** Profile of the metabolic pathways in the GA estimation model over the course of gestation on SU cohort. All pathways are (A) positively or (B) negatively correlated to the GA (FDR<1%). Profile of each pathway was calculated as the weighted sum of the z-score normalized metabolite serological abundances divided by the number of metabolites. Means ± standard errors at each time point were plotted.

**Fig. A.3** Univariate analysis of the 33 metabolic pathways in the GA estimation model. Pearson correlation coefficient of each pathway to GA was calculated. **P*<0.05, ***P*<0.01, ****P*<0.005.

**Fig. A.4** Comparison of GA estimates using the model and US measurements. (A) Distributions of differences between GA measured by US and GA estimated by the model, in T2 (weeks 14–27), T3 (weeks 28–40), and T2+T3. n represents the number of full-term patients included. (B) Error distribution of GA estimation on a combination of SU and UAB cohorts in T2, T3, and T2+T3.

**Fig. A.5** False discovery rate (FDR) analysis of the metabolic pathways significantly associated with PTB. Mann-Whitney U test *P* measured the difference in metabolite serological abundances between full-term pregnancies and pregnancies ending in PTB.

Only metabolites with a Mann-Whitney U test *P* lower than the threshold were selected as part of the significant pathways. FDR was estimated by a permutation-based method (permutation N=1000).

**Fig. A.6** Stratification of patients by the classification model prediction on the UAB cohort. PPV was corrected by bootstrapping the full-term patients to reach the population PTB prevalence of 12.5% on singleton births. Two horizontal dashed lines represent the population mean of PTB risk that is 12.5% (black) and the PPV (= 0.70; red) at the high-risk cutoff. The grey dashed line indicates the high-risk cutoff value (= 0.52). The grey area represents the 95% confidence interval of the PPV. The box plot at the bottom shows the classification model value distribution stratified by the samples. GAB: GA at birth. wks: weeks of gestation.

**Fig. A.7** The performance of the IBP4/SHBG predictor and the metabolic model. The results are stratified by the GA intervals with a BMI at 22–37 kg/m2 (A), and by BMI values with a GA interval of 5–20 weeks (B).

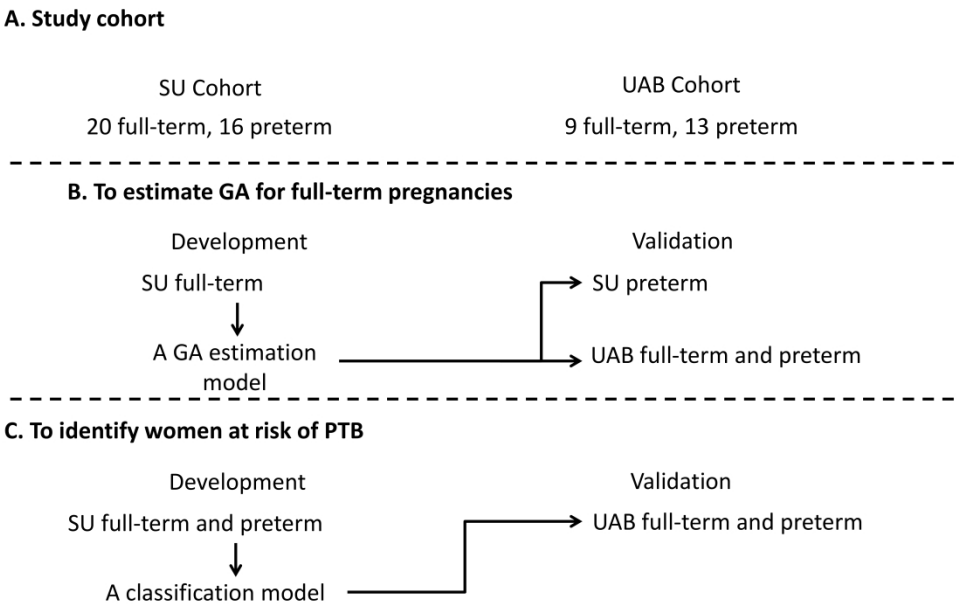
**Fig. A.8** (A) False discovery rate (FDR) analysis of the metabolites and metabolic pathways significantly associated with GA in full-term pregnancies. Pearson |r| was calculated as the correlation between metabolite serological abundance and GA. Only the metabolites with a Pearson |r| higher than the threshold (=0.35) would be selected as part of the significant pathways. FDR was estimated by a permutation-based method (permutation N=1000). (B) A comparison of RMSE of the GA estimation model trained by pathways and the model trained by metabolites. All metabolites had a Pearson |r|>0.35. RMSE was measured with the full-term samples of the validation (UAB) cohort.

**Fig. A.9** (A) False discovery rate (FDR) analysis of the metabolites and metabolic pathways significantly associated with the PTB. Mann-Whitney U test *P* measured the difference in metabolite serological abundances between full-term pregnancies and pregnancies ending in PTB. Only the metabolites with a Mann-Whitney U test *P* lower than the threshold (=0.05) would be selected as part of the significant pathways. FDR was estimated by a permutation-based method (permutation N=1000). (B) A comparison of the AUC of the PTB classification model utilizing pathways and the model utilizing metabolites. All the metabolites had a Mann-Whitney U test $P < 0.05$. AUC was measured with the samples of the validation (UAB) cohort.

**Table A.1** Sensitivity and specificity of the XGBoost model with respect to the cutoff point.

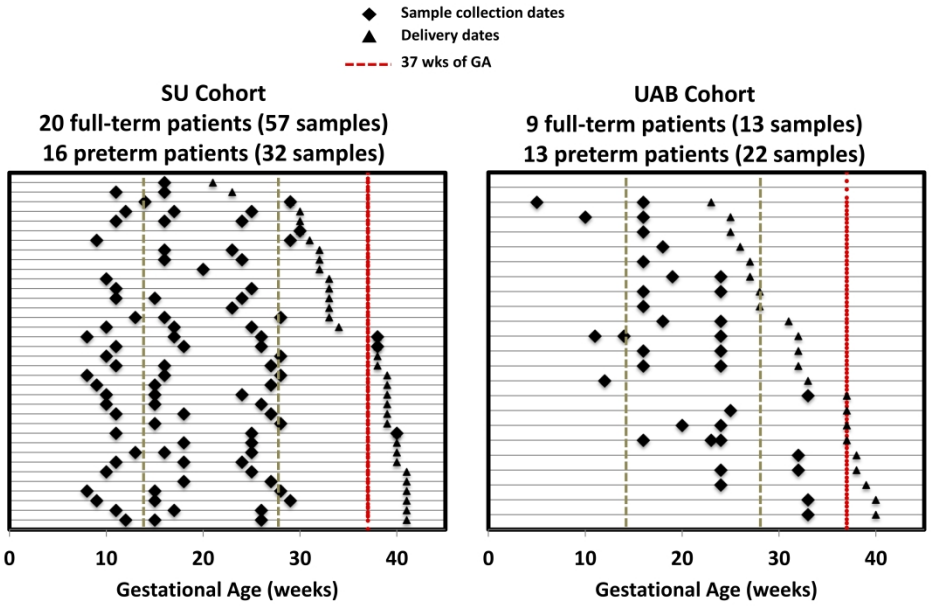
**Text A.1** Metabolic compound selection, pathway computation, and model development

**Text A.2** Metabolite model vs. IBP4/SHBG in predicting PTB



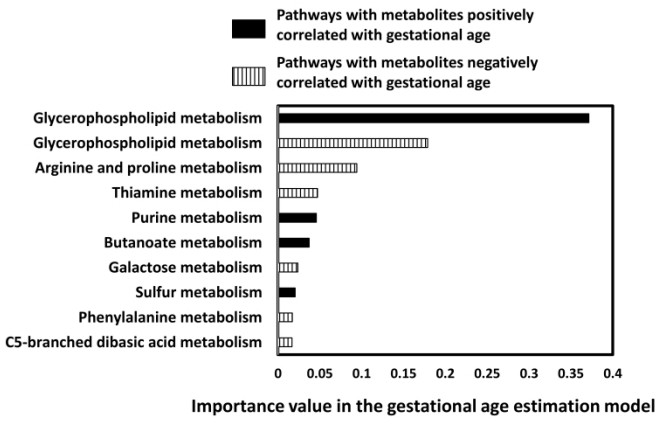

Study design. Models were developed separately to estimate gestational age during full-term pregnancy, and to predict the risk of preterm birth. Both models were developed with the SU cohort and validated with the UAB cohort.

254x190mm (600 x 600 DPI)



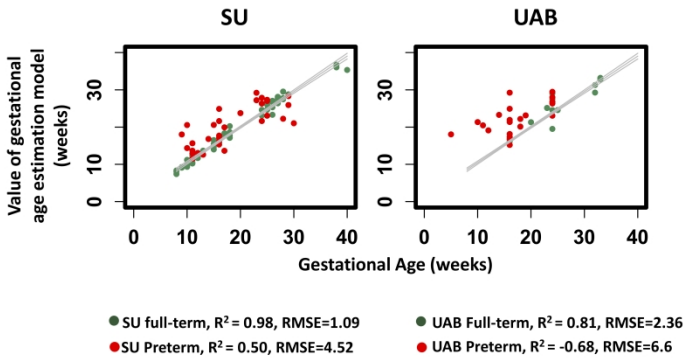

Cohort construction. Each line represents an individual patient. Diamond and triangle markers indicate sample collection dates and delivery dates, respectively. The red dashed line represents 37 weeks' gestational age.

254x190mm (600 x 600 DPI)



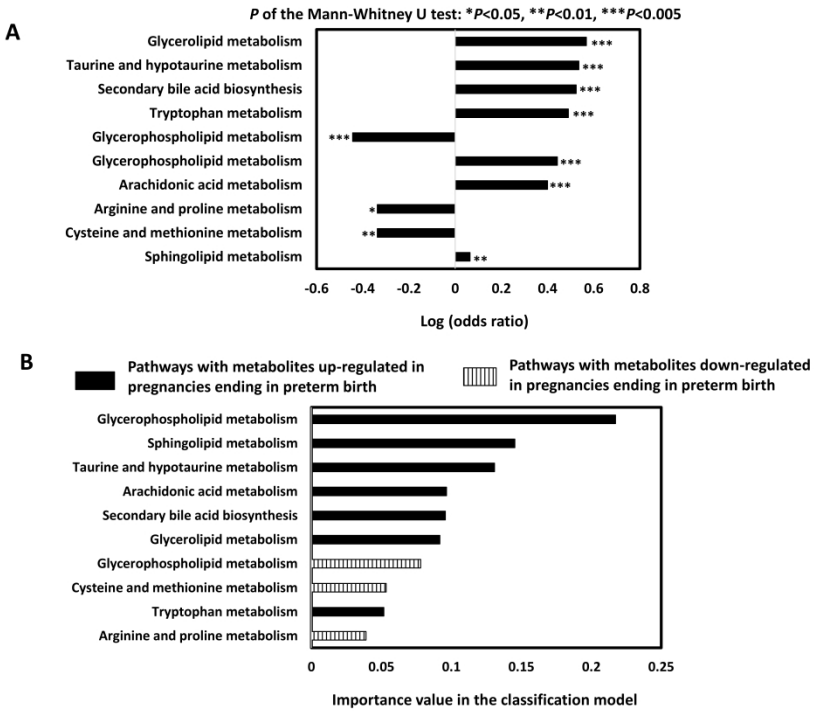

The importance of the top 10 metabolic pathways in the gestational age estimation model. Pathways either positively or negatively correlated gestational age.

254x190mm (600 x 600 DPI)



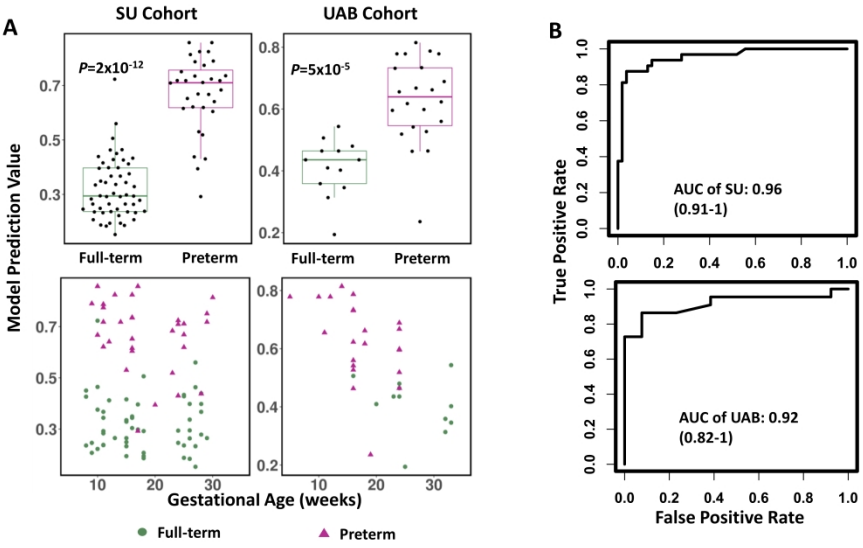

Gestational age estimates of the gestational age model with the SU (R2=0.98, RMSE=1.09 weeks) and UAB cohorts (R2 = 0.81, RMSE = 2.36 weeks).

254x190mm (600 x 600 DPI)




(A) Univariate analysis of the 10 metabolic pathways in the preterm birth prediction model. Odds ratio of each pathway was calculated. *P<0.05, **P<0.01, ***P<0.005. (B) The importance of the metabolic pathways in the preterm birth prediction model. Pathways were either up- or down-regulated in relation to preterm birth.

254x190mm (600 x 600 DPI)




(A) Prediction of preterm birth risk grouped by full-term and preterm birth patients (top) and over the course of gestation (bottom). (B) AUC performance of the prediction in SU and UAB cohorts. P was calculated using Mann–Whitney U test. wks: weeks' gestational age.

254x190mm (600 x 600 DPI)

A

**SU**

| | Preterm | Full-term |
|---|---|---|
| Classified as Preterm | 28 | 2 |
| Classified as Full-term | 4 | 52 |

**UAB**

| | Preterm | Full-term |
|---|---|---|
| Classified as Preterm | 19 | 1 |
| Classified as Full-term | 3 | 12 |

B

| Cohort | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) |
|---|---|---|---|---|
| **SU** | 0.88 (0.71-0.97) | 0.96 (0.87-1) | 0.93 (0.78-0.99) | 0.93 (0.83-0.98) |
| **UAB** | 0.86 (0.65-0.97) | 0.92 (0.64-1) | 0.95 (0.75-1) | 0.80 (0.52-0.96) |

Performance of the preterm birth prediction model. (A) A contingency table showing the number of samples in each category. (B) Sensitivity, specificity, PPV, and NPV together with the 95% confidence intervals.

254x190mm (600 x 600 DPI)

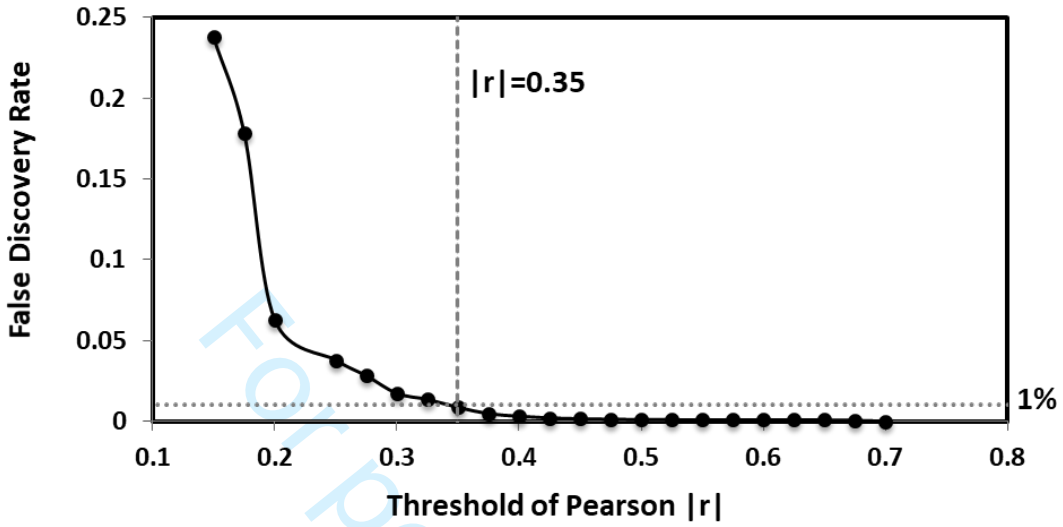



**Fig. A.1.** False discovery rate (FDR) analysis of the metabolic pathways significantly associated with the GA in full-term pregnancies. Pearson |r| was calculated as the correlation between metabolite serological abundance and GA. Only the metabolites with a Pearson |r| higher than the threshold would be selected as part of the significant pathways. FDR was estimated by a permutation-based method (permutation N=1000).

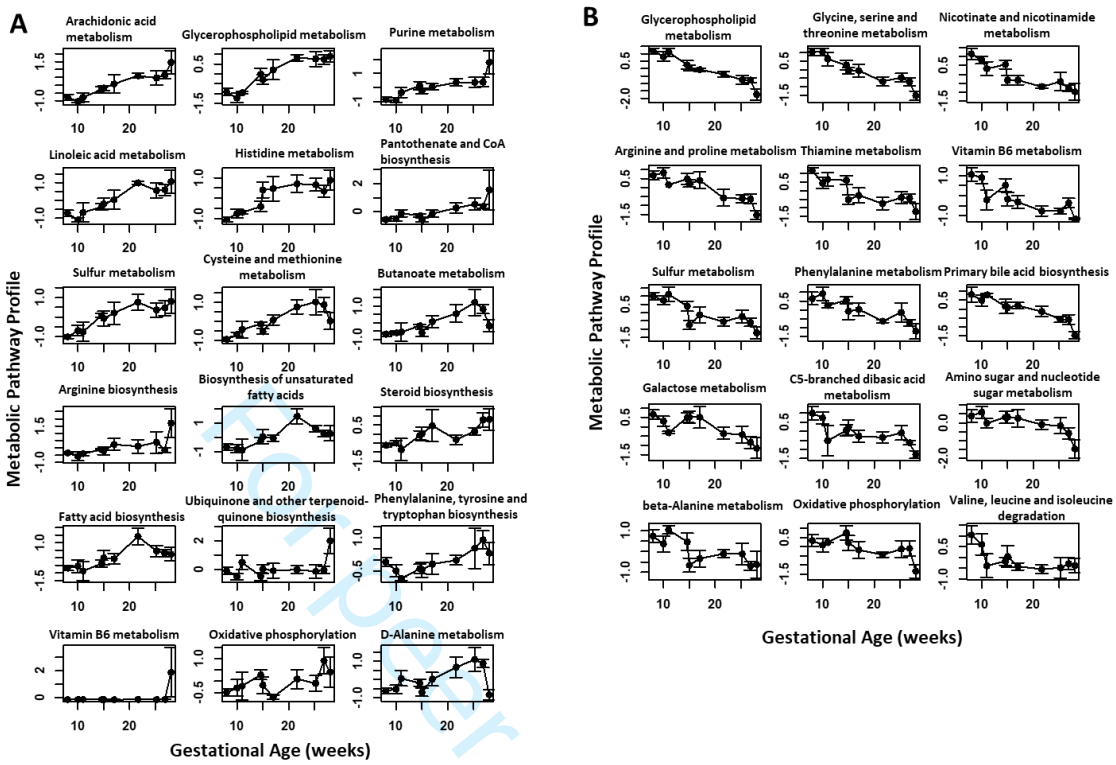



**Fig. A.2.** Profile of the metabolic pathways in the GA estimation model over the course of gestation on SU cohort. All pathways are (A) positively or (B) negatively correlated to the GA (FDR<1%). Profile of each pathway was calculated as the weighted sum of the z-score normalized metabolite serological abundances divided by the number of metabolites. Mean ± standard error of the mean at each time point was plotted.

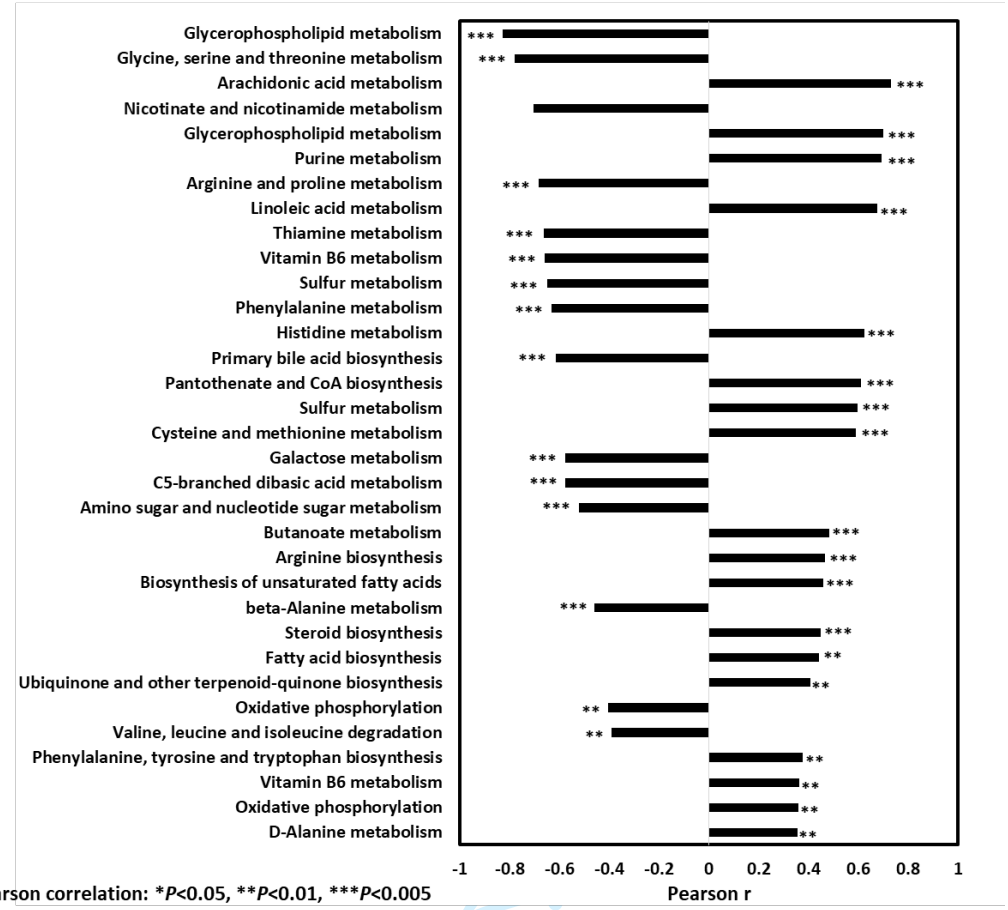



**Fig. A.3.** Univariate analysis of the 33 metabolic pathways in the GA estimation model. Pearson correlation coefficient r of each pathway to GA was calculated. *$P$<0.05, **$P$<0.01, ***$P$<0.005.

A

| Trimester and subject number | Δ [model estimation – ultrasound measurements (weeks)] (%) | | | | |
|---|---|---|---|---|---|
| | < -2 | -1 to -2 | ±1 | +1 to +2 | > +2 |
| SU (T2, n = 19) | 0 | 0 | 84.2 | 15.8 | 0 |
| SU (T3, n = 8) | 12.5 | 25 | 50 | 12.5 | 0 |
| SU (All, n = 20) | 0 | 5 | 85 | 10 | 0 |
| UAB (T2, n = 5) | 0 | 0 | 60 | 0 | 40 |
| UAB (T3, n = 5) | 20 | 0 | 80 | 0 | 0 |
| UAB (All, n = 9) | 11.1 | 0 | 66.7 | 11.1 | 11.1 |
| SU and UAB (T2, n = 24) | 0 | 0 | 79.2 | 12.5 | 8.3 |
| SU and UAB (T3, n = 13) | 15.4 | 15.4 | 61.5 | 7.7 | 0 |
| SU and UAB (All, n = 29) | 3.4 | 3.4 | 79.3 | 10.3 | 3.4 |

B

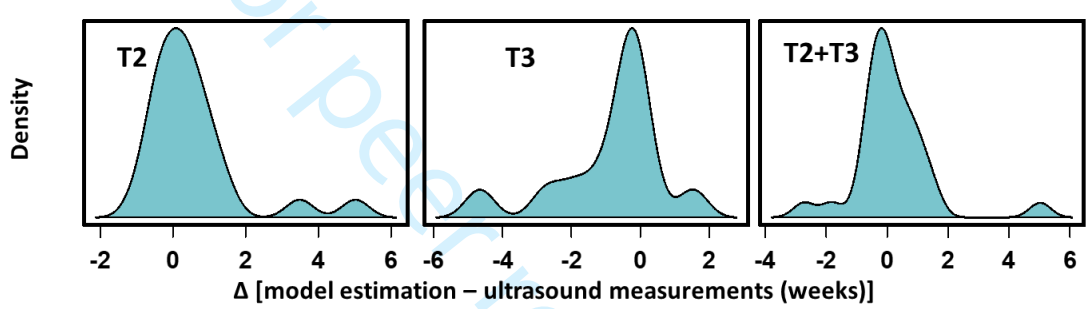



**Fig. A.4.** Comparison of GA estimates using the model and US measurements. (A) Distributions of differences between GA measured by US and GA estimated by the model, in T2 (weeks 14–27), T3 (weeks 28–40), and T2+T3. n represents the number of full-term patients included. (B) Error distribution of GA estimation on a combination of SU and UAB cohorts in T2, T3, and T2+T3.

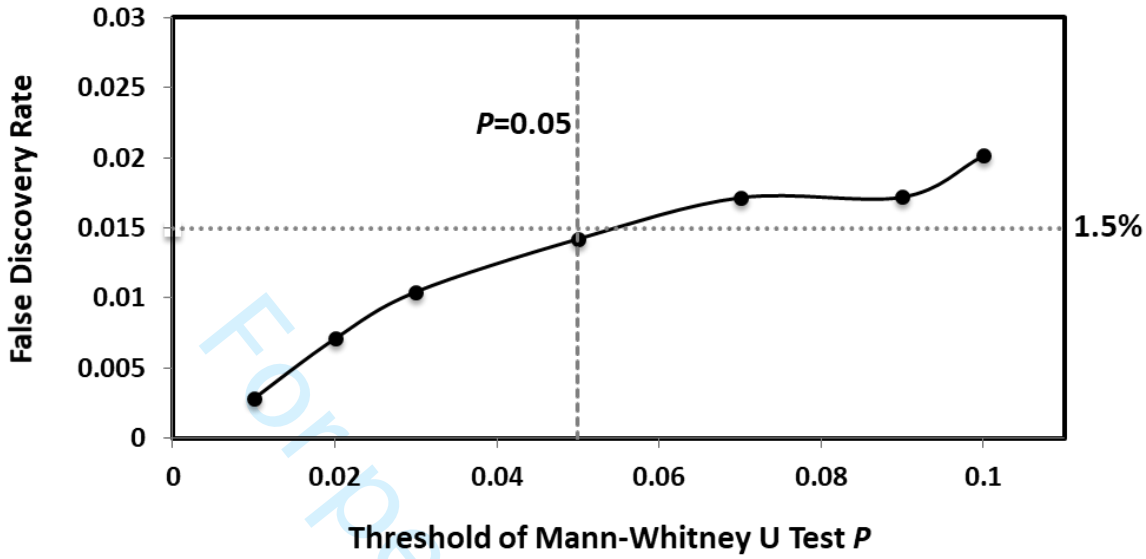



**Fig. A.5.** False discovery rate (FDR) analysis of the metabolic pathways significantly associated with PTB. Mann-Whitney U test *P* measured the difference in metabolite serological abundances between full-term pregnancies and pregnancies ending in PTB. Only metabolites with a Mann-Whitney U test *P* lower than the threshold were selected as part of the significant pathways. FDR was estimated by a permutation-based method (permutation N=1000).

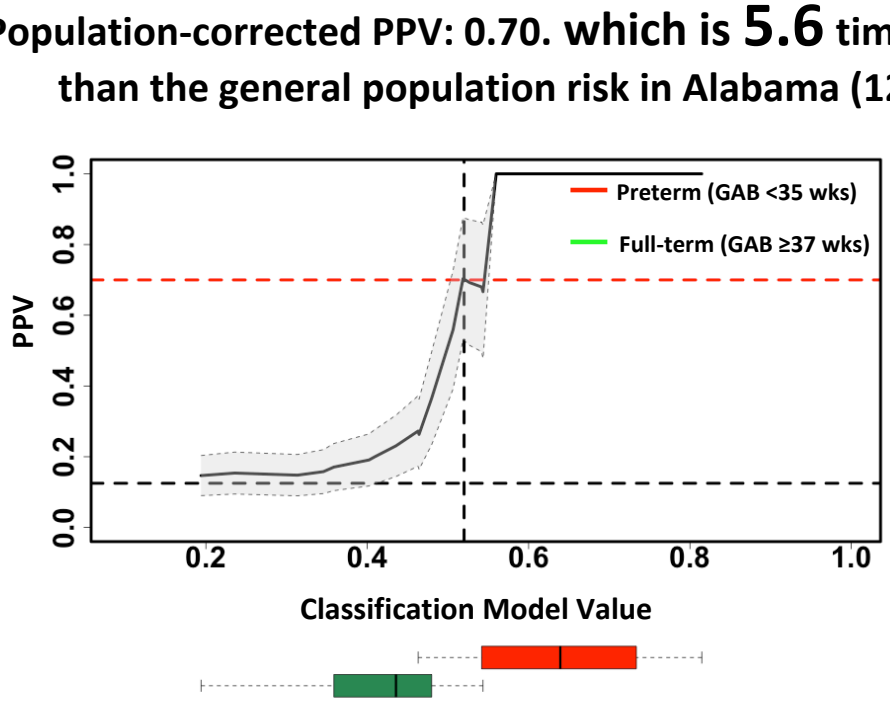



**Fig. A.6**. Stratification of patients by the classification model prediction on the UAB cohort. PPV was corrected by bootstrapping the full-term patients to reach the population PTB prevalence of 12.5% on singleton births in Alabama. Two horizontal dashed lines represent the population mean of PTB risk that is 12.5% (black) and the PPV (= 0.70; red) at the high-risk cutoff. The grey dashed line indicates the high-risk cutoff value (= 0.52). The grey area represents the 95% confidence interval of the PPV. The box plot at the bottom shows the classification model value distribution stratified by the samples. GAB: gestational age at birth. wks: weeks' GA.

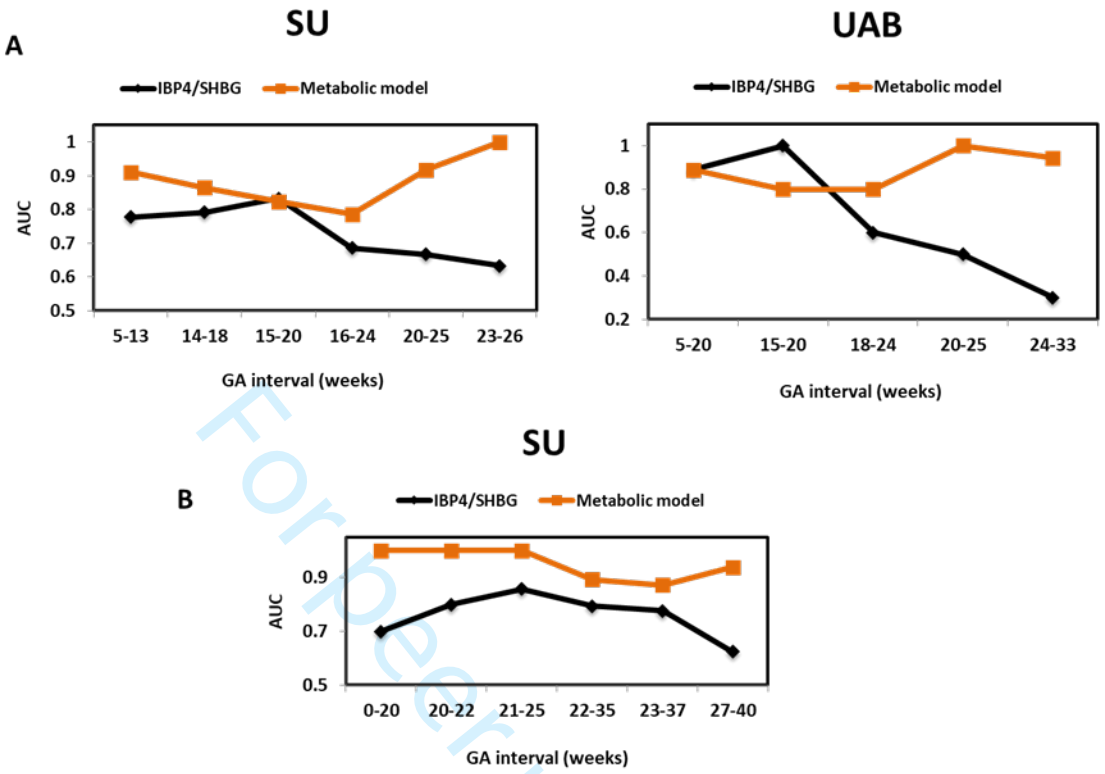



**Fig. A.7.** The performance of the IBP4/SHBG predictor and the metabolic model. The results are stratified by the GA intervals with a BMI at 22–37 kg/m$^2$ (A), and by BMI values with a GA interval of 5–20 weeks (B).

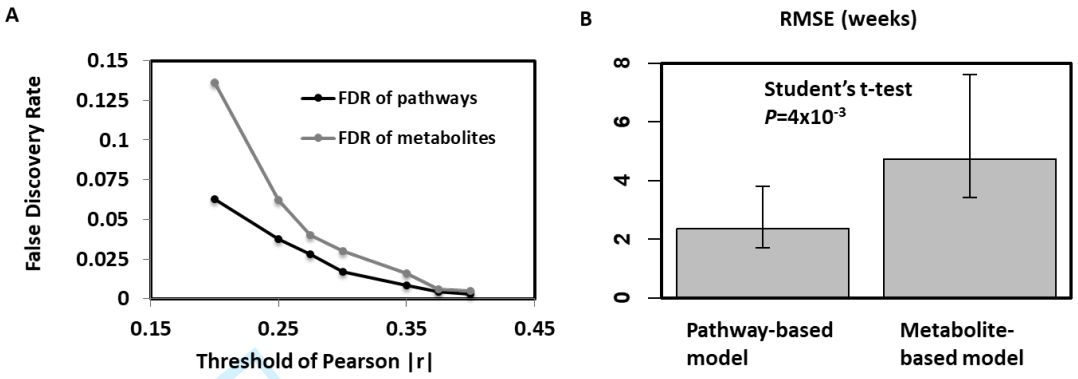



**Fig. A.8.** (A) False discovery rate (FDR) analysis of the metabolites and metabolic pathways significantly associated with the GA in full-term pregnancies. Pearson |r| was calculated as the correlation between metabolite serological abundance and GA. Only the metabolites with a Pearson |r| higher than the threshold (=0.35) would be selected as part of the significant pathways. FDR was estimated by a permutation-based method (permutation N=1000). (B) A comparison of RMSE of the GA estimation model trained by pathways and the model trained by metabolites. All metabolites had a Pearson |r|>0.35. RMSE was measured with the full-term samples of the validation (UAB) cohort.

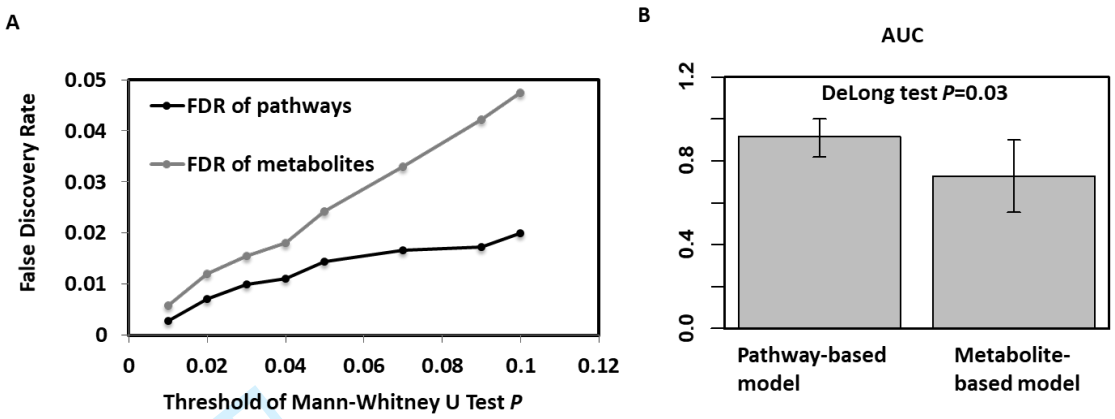



**Fig. A.9.** (A) False discovery rate (FDR) analysis of the metabolites and metabolic pathways significantly associated with the PTB. Mann-Whitney U test *P* measured the difference in metabolite serological abundances between full-term pregnancies and pregnancies ending in PTB. Only the metabolites with a Mann-Whitney U test *P* lower than the threshold (=0.05) would be selected as part of the significant pathways. FDR was estimated by a permutation-based method (permutation N=1000). (B) A comparison of the AUC of the preterm birth classification model utilizing pathways and the model utilizing metabolites. All the metabolites had a Mann-Whitney U test $P < 0.05$. AUC was measured with the samples of the validation (UAB) cohort.

**Table A.1.** Sensitivity and specificity of the XGBoost model with respect to the cutoff point.

| Cutoff | Cohort | Sensitivity | Specificity | Number of preterm samples identified by the model |
|---|---|---|---|---|
| **0.4** | **SU** | 0.94 | 0.78 | 30 |
| | **UAB** | 0.95 | 0.31 | 21 |
| **0.5** | **SU** | 0.88 | 0.94 | 28 |
| | **UAB** | 0.86 | 0.85 | 19 |
| **0.6** | **SU** | 0.81 | 0.98 | 26 |
| | **UAB** | 0.59 | 1 | 13 |
| **0.7** | **SU** | 0.53 | 0.98 | 17 |
| | **UAB** | 0.32 | 1 | 7 |

**Text A.1 Metabolic compound selection, pathway computation, and model development**

*GA estimation*

Metabolites measured by targeted and untargeted MS were aggregated and filtered using Pearson correlation coefficient analyses in relation to GA. The remaining metabolites were mapped to pathways. The value of each pathway was calculated as the weighted sum of the normalized concentrations of metabolites on the pathway divided by the number of metabolites. The weight of each metabolite was the absolute value of the Pearson correlation coefficient in relation to GA. Metabolites having positive or negative coefficients were aggregated separately. That is, a pathway could have two values, one for metabolites positively correlated to GA, and the other for those negatively correlated to GA.

A supervised, cross-validated machine-learning technique XGBoost was developed with the pathway values of samples from full-term patients in the SU cohort. An ensemble of regression trees was generated to give a score estimating the GA. The model was validated on the UAB cohort. For a patient that had multiple samples, an 'integrated' GA estimate was calculated by shifting the GA estimates of every sample to a reference point for obtaining the median. Error distribution of GA estimation based on patients was calculated as the distribution of the differences between the 'integrated' GA estimates and the US measurement.

*PTB prediction*

Samples collected before 35 weeks' GA were selected to build the model to predict PTB. Mann–Whitney U test was used to select the initial candidate metabolites that were then mapped to pathways. The value of each pathway was calculated as the weighted sum of the normalized concentrations of metabolites on the pathway divided by the number of metabolites. The weight of each metabolite was the absolute value of the ratio of median of full-term samples to PTB samples. Like the GA estimation, pathways could have two values that depended on the ratio of median greater or less than 1. An XGBoost model was developed utilizing samples from the SU cohort and validated with the UAB cohort.

**Text A.2 Metabolite model vs. IBP4/SHBG in predicting PTB**

We conducted ELISA tests on the SU and UAB cohorts to evaluate the IBP4/SHBG signature, a predictor that was validated in a prospective study as a predictor of spontaneous PTB. Commercial kits Human IGFBP4 ELISA Kit (Abcam, Burlingame, CA, USA) and Human SHBG Quantikine ELISA Kit (R&D System Inc.) were used. AUC of the predictor was calculated in different GA intervals and with different maternal BMI values, and was compared to the performance of the metabolic model.

With a BMI of >22 and ≤37 kg/m$^2$, the AUC values of the IBP4/SHBG predictor peaked at 15–20 weeks' GA (SU: 0.833; UAB: 1), and dropped rapidly after 20 weeks (Figure A below). The AUC values were lower with extreme BMI (0.7 at BMI ≤20 kg/m$^2$ and 0.63 at BMI >27 kg/m$^2$; see Figure B below). These findings are consistent with the previous validation study. Compared with the IBP4/SHBG predictor, the metabolic model has a more stable AUC performance over the gestation and different BMI values in SU ($P$ = 0.03). In UAB at >18 weeks' GA, the AUC of IBP4/SHBG dropped from 0.6 to 0.3, while the AUC of the metabolic model was above 0.8.

STROBE Statement—checklist of items that should be included in reports of observational studies

| | Item No | Recommendation | Page No |
|---|---|---|---|
| **Title and abstract** | 1 | (*a*) Indicate the study's design with a commonly used term in the title or the abstract | 1 |
| | | (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found | 2-3 |
| **Introduction** | | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported | 5 |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses | 6 |
| **Methods** | | | |
| Study design | 4 | Present key elements of study design early in the paper | 6 |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | 6,7 |
| Participants | 6 | (*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up<br>*Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls<br>*Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants | 6,7 |
| | | (*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed<br>*Case-control study*—For matched studies, give matching criteria and the number of controls per case | |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | 6,7,8 |
| Data sources/ measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | *7* |
| Bias | 9 | Describe any efforts to address potential sources of bias | 8 |
| Study size | 10 | Explain how the study size was arrived at | |
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | 7,8 |
| Statistical methods | 12 | (*a*) Describe all statistical methods, including those used to control for confounding | 8 |
| | | (*b*) Describe any methods used to examine subgroups and interactions | |
| | | (*c*) Explain how missing data were addressed | |
| | | (*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed<br>*Case-control study*—If applicable, explain how matching of cases and controls was addressed<br>*Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy | |
| | | (*e*) Describe any sensitivity analyses | 8 |

Continued on next page

| | | | |
|---|---|---|---|
| **Results** | | | |
| Participants | 13* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed | 9 |
| | | (b) Give reasons for non-participation at each stage | |
| | | (c) Consider use of a flow diagram | |
| Descriptive data | 14* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders | 9 |
| | | (b) Indicate number of participants with missing data for each variable of interest | |
| | | (c) *Cohort study*—Summarise follow-up time (eg, average and total amount) | 9 |
| Outcome data | 15* | *Cohort study*—Report numbers of outcome events or summary measures over time | *10* |
| | | *Case-control study*—Report numbers in each exposure category, or summary measures of exposure | |
| | | *Cross-sectional study*—Report numbers of outcome events or summary measures | |
| Main results | 16 | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included | 11-12 |
| | | (*b*) Report category boundaries when continuous variables were categorized | 11-12 |
| | | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period | |
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses | 12 |
| **Discussion** | | | |
| Key results | 18 | Summarise key results with reference to study objectives | 13 |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias | 16 |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence | 14-15 |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results | 16 |
| **Other information** | | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based | 18 |

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.