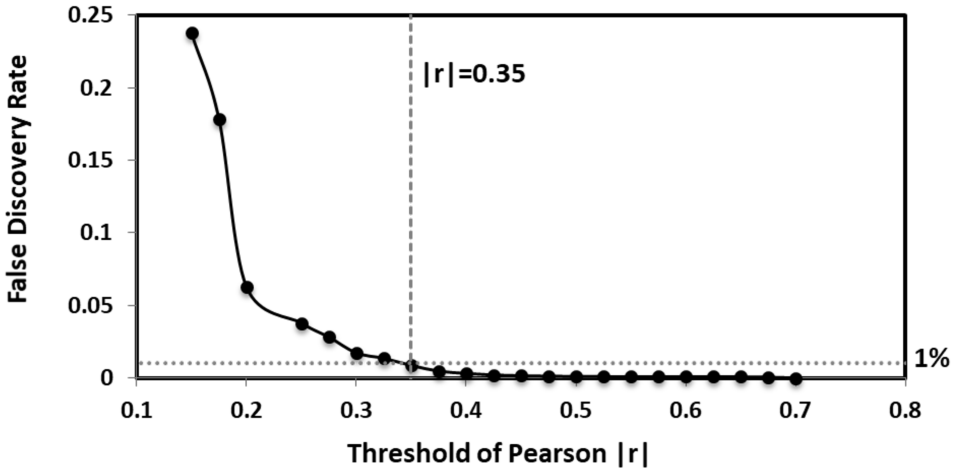

Fig. A.1. False discovery rate (FDR) analysis of the metabolic pathways significantly associated with the GA in full-term pregnancies. Pearson |r| was calculated as the correlation between metabolite serological abundance and GA. Only the metabolites with a Pearson |r| higher than the threshold would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

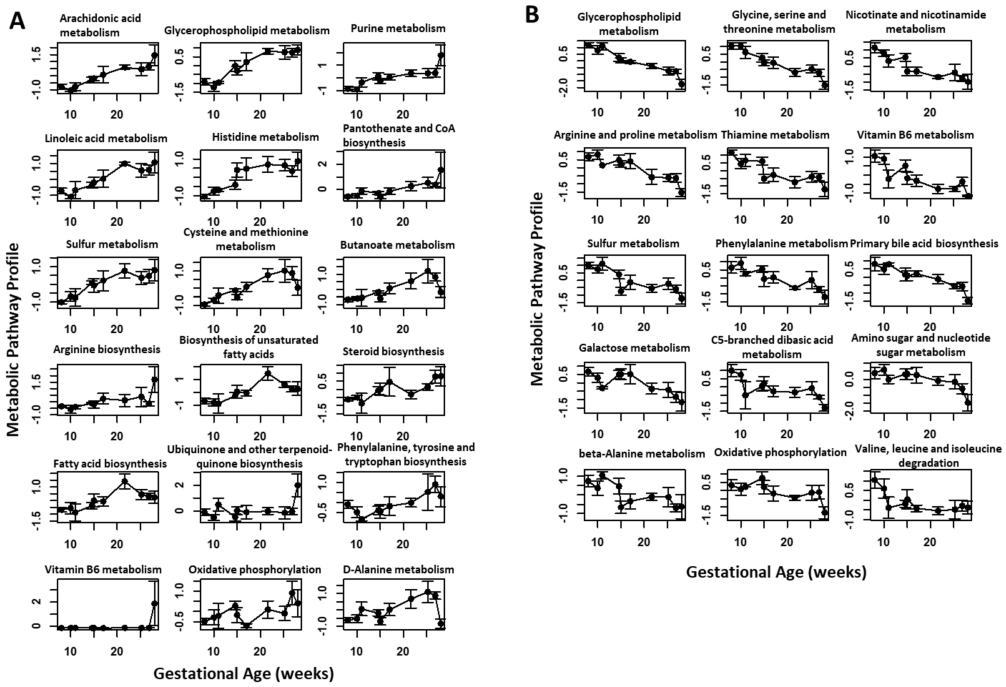

Fig. A.2. Profile of the metabolic pathways in the GA estimation model over the course of gestation on SU cohort. All pathways are (A) positively or (B) negatively correlated to the GA (FDR<1%). Profile of each pathway was calculated as the weighted sum of the z-score normalized metabolite serological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

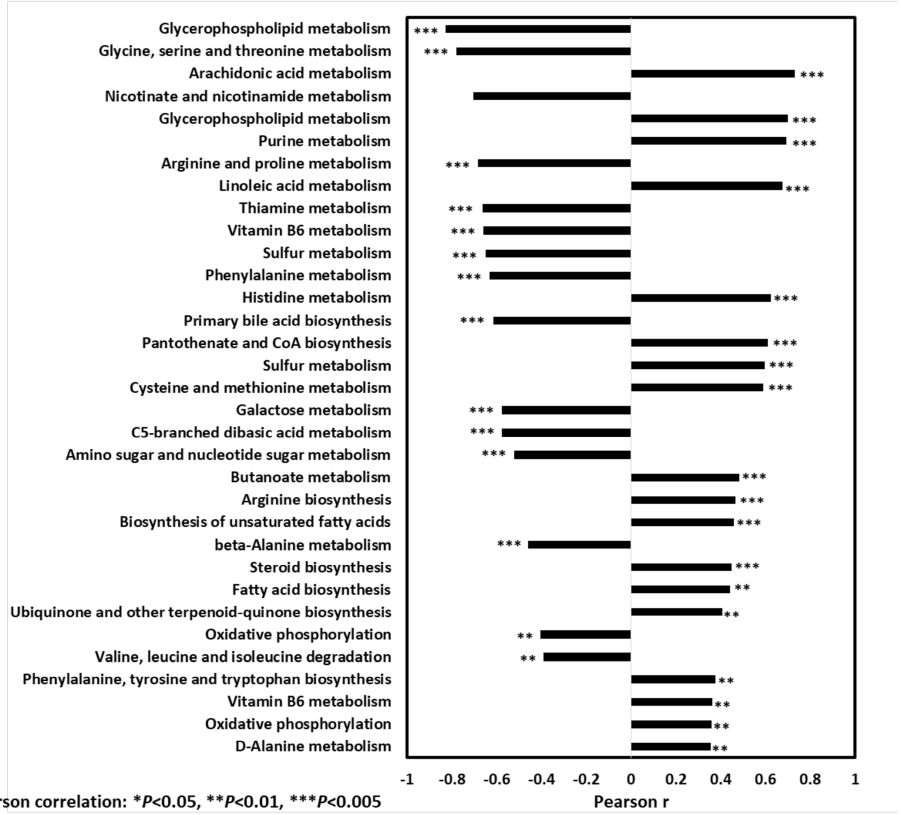

Fig. A.3. Univariate analysis of the 33 metabolic pathways in the GA estimation model. Pearson correlation coefficient r of each pathway to GA was calculated. **P*<0.05, ***P*<0.01, ****P*<0.005.

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

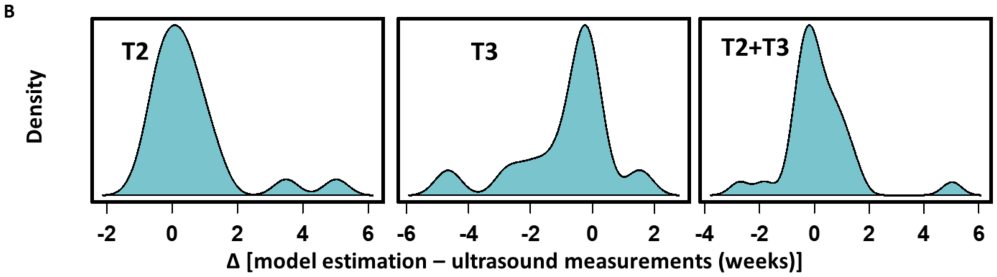

Fig. A.4. Comparison of GA estimates using the model and US measurements. (A) Distributions of differences between GA measured by US and GA estimated by the model, in T2 (weeks 14–27), T3 (weeks 28–40), and T2+T3. n represents the number of full-term patients included. (B) Error distribution of GA estimation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

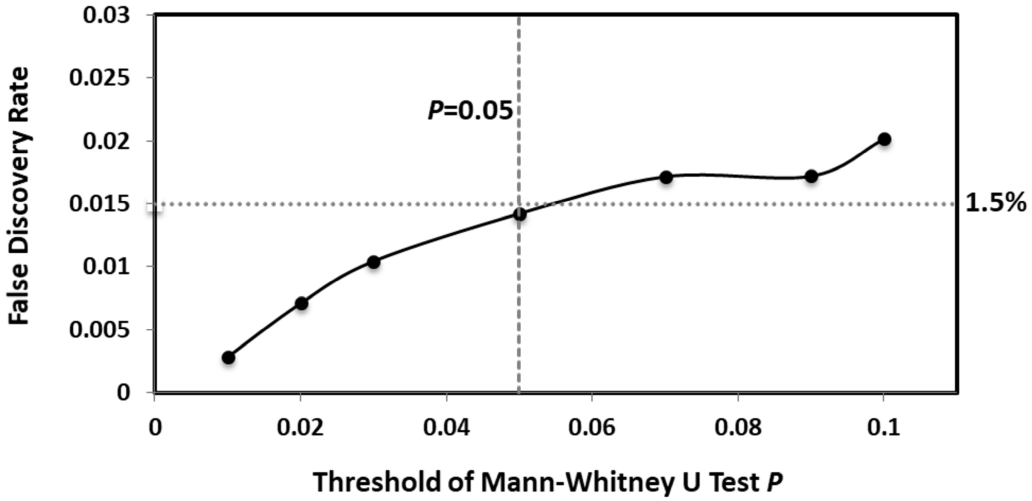

Fig. A.5. False discovery rate (FDR) analysis of the metabolic pathways significantly associated with PTB. Mann-Whitney U test *P* measured the difference in metabolite serological abundances between full-term pregnancies and pregnancies ending in PTB. Only metabolites with a Mann-Whitney U test *P*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

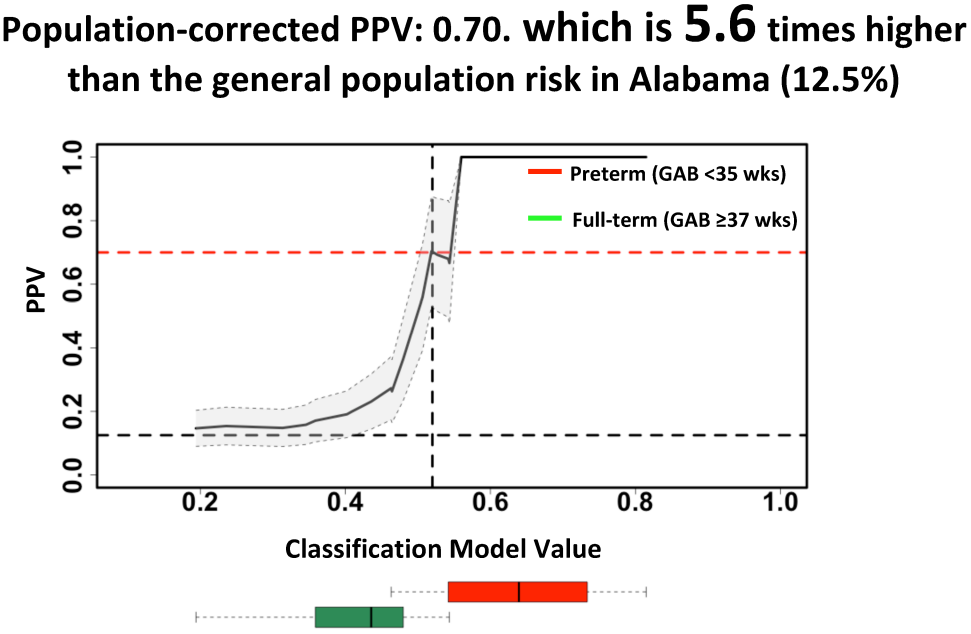

Fig. A.6. Stratification of patients by the classification model prediction on the UAB cohort. PPV was corrected by bootstrapping the full-term patients to reach the population PTB prevalence of 12.5% on singleton births in Alabama. Two horizontal dashed lines represent the population mean of PTB risk 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.

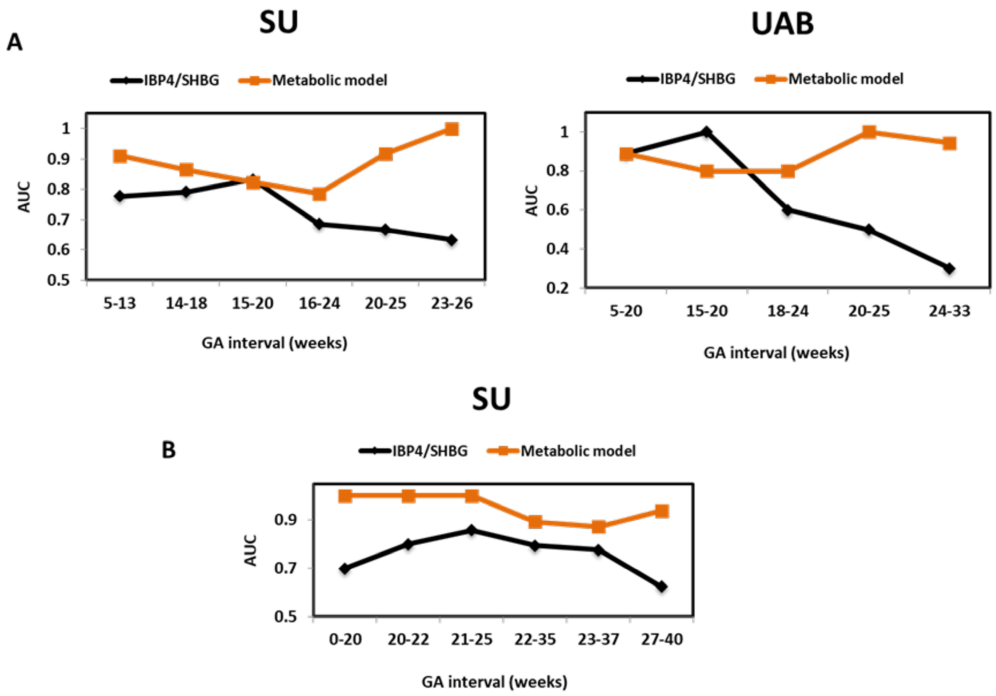

Fig. A.7. The performance of the IBP4/SHBG predictor and the metabolic model. The results are stratified by the GA intervals with a BMI at 22–37 kg/m^2 (A), and by BMI values with a GA interval of 5–20 weeks (B).

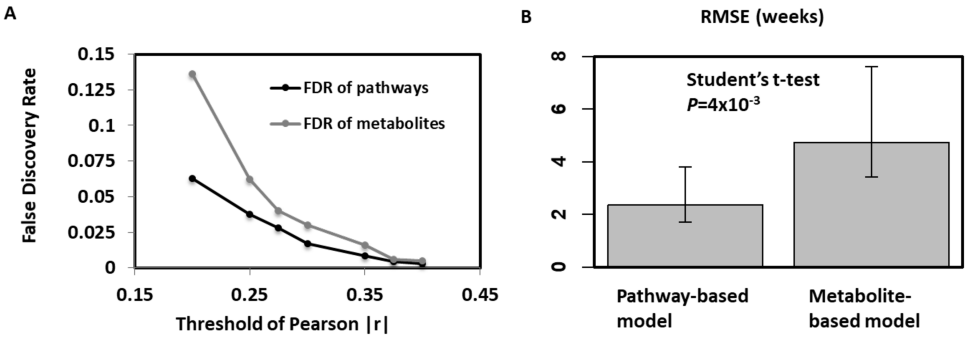

Fig. A.8. (A) False discovery rate (FDR) analysis of the metabolites and metabolic pathways significantly associated with the GA in full-term pregnancies. Pearson |r| was calculated as the correlation between metabolite serological abundance and GA. Only the metabolites with a Pearson |r| higher than 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.

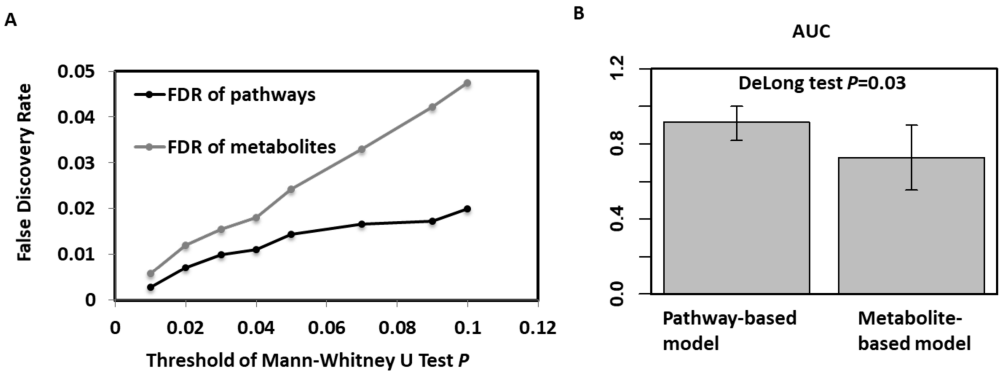

Fig. A.9. (A) False discovery rate (FDR) analysis of the metabolites and metabolic pathways significantly associated with the PTB. Mann-Whitney U test *P* measured the difference in metabolite serological abundances between full-term pregnancies and pregnancies ending in PTB. Only the metabolites with 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.