European Journal of Medical Research

RESEARCH

Open Access

Metabolic profiling and early prediction models for gestational diabetes mellitus in PCOS and non-PCOS pregnant women

Jin Wang[1,2†], Can Cui[1,2†], Fei Hou[1,2], Zhiyan Wu[3], Yingying Peng[1,2] and Hua Jin[1,2*]

Abstract

Background Gestational diabetes mellitus (GDM) is the most common pregnancy complication, significantly affecting maternal and neonatal health. Polycystic ovary syndrome (PCOS) is a common endocrine disorder characterized by metabolic abnormalities, which notably elevates the risk of developing GDM during pregnancy.

Methods In this study, we utilized ultra-high-performance liquid chromatography for untargeted metabolomics analysis of serum samples from 137 pregnant women in the early-to-mid-pregnancy. The cohort consisted of 137 participants, including 70 in the PCOS group (36 who developed GDM in mid-to-late pregnancy and 34 who did not) and 67 in the non-PCOS group (37 who developed GDM and 30 who remained GDM-free). The aim was to investigate metabolic profile differences between PCOS and non-PCOS patients and to construct early GDM prediction models separately for the PCOS and non-PCOS groups.

Results Our findings revealed significant differences in the metabolic profiles of PCOS patients, which may help elucidate the higher risk of GDM in the PCOS population. Moreover, tailored early GDM prediction models for the PCOS group demonstrated high predictive performance, providing strong support for early diagnosis and intervention in clinical practice.

Conclusions Untargeted metabolomics analysis revealed distinct metabolic patterns between PCOS patients and non-PCOS patients, particularly in pathways related to GDM. Based on these findings, we successfully constructed GDM prediction models for both PCOS and non-PCOS groups, offering a promising tool for clinical management and early intervention in high-risk populations.

Keywords Gestational diabetes mellitus, Polycystic ovary syndrome, Untargeted metabolomics analysis, Prediction models

Background

Gestational diabetes mellitus (GDM), one of the most common complications of pregnancy, is an abnormality in glucose tolerance of varying degrees that occurs for the first time during pregnancy [1]. GDM is rising globally, year by year, ranging from 5.8% to 14.0% [2–5]. Studies have shown that, as the most common pregnancy complication in the second and third trimesters, it is associated with a higher risk of adverse perinatal outcomes, including preeclampsia, infection, premature birth, increased cesarean section rates, and premature rupture

†Jin Wang and Can Cui have contributed equally to this work.

*Correspondence:
Hua Jin
prenatal123@163.com
[1] Prenatal Diagnosis Center, Jinan Maternal and Child Health Care Hospital, No. 2, Rd. Jianguo Xiaojing, Jinan 250002, Shandong Province, People's Republic of China
[2] Shandong First Medical University, Jinan, Shandong Province, People's Republic of China
[3] Department of Gynecology, Qingzhou People's Hospital, Weifang, Shandong Province, People's Republic of China

of membranes [6, 7]. Furthermore, women with GDM and their children are more likely to develop type 2 diabetes in their later life [8]. Pregnant women with GDM have a 52.2% risk of developing type 2 diabetes within the first decade after giving birth [9]. Recent studies showed that the overall relative risk of type 2 diabetes in women with DGM is increased by 7–10 times compared to women without GDM [10–12].

Polycystic ovary syndrome (PCOS) is a common endocrine disorder characterized by insulin resistance, hyperandrogenism, and metabolic abnormalities, affecting ~ 5–15% of women of reproductive age [13, 14]. Patients with PCOS have a significantly higher risk of developing GDM during pregnancy compared to the general pregnant population [15]. Some studies have shown that PCOS increases the risk of developing gestational diabetes by two to three times [16]. Previous studies have shown that the combination of PCOS and insulin resistance may accelerate the progression of glucose metabolism disorders [17]. Thus, the mechanism of GDM in patients with PCOS may differ from that in the general pregnant population.

As two of the most common metabolic diseases affecting women of reproductive age, both are risk factors for future metabolic and cardiovascular abnormalities, especially type 2 diabetes. The prevalence of GDM and PCOS, as well as diabetes itself, is increasing worldwide, posing a significant public health challenge. Together, GDM and PCOS, which act and affect up to 20% of generally young women, are two critical risk factors for the future development of type 2 diabetes [18, 19].

The current gold standard for diagnosing GDM is the oral glucose tolerance test (OGTT) [20]. Women at high risk for GDM undergo an OGTT at the earliest opportunity, whereas women who do not meet the high-risk criteria are usually only examined after 24–28 weeks of gestation [21, 22]. This means that the opportunity for early detection and intervention for GDM is delayed. In addition, the OGTT has a specific false-positive and false-negative rate, which may lead to inaccurate predictions, especially if individuals have different levels of insulin resistance. Existing diagnostic tools are mainly based on routine glucose monitoring but do not allow an in-depth exploration of the whole metabolic pictures of the pregnant women. Machine learning (ML) models are increasingly used to identify risk factors and early predict GDM [23].

Metabolomics can provide the instantaneous state of cellular metabolic activity and reflect the combined effects of environmental, genetic, and physiological changes on the organism. Compared with genomics and proteomics, metabolites are downstream products closer to disease manifestations and are, therefore, more suitable for predicting metabolic-related diseases. Non-targeted metabolomics analysis can comprehensively cover the metabolite spectrum, thereby revealing more potential disease biomarkers and facilitating early detection of disease risks [24]. Metabolites are recognized as potential biomarkers of disease, and the ability to reveal metabolic abnormalities before the onset of disease symptoms can help to identify at-risk populations in early pregnancy and provide a basis for early intervention. For example, cardiometabolic diseases are associated with specific lipid metabolites [25], and branched-chain amino acids and their metabolites can predict the incidence of type 2 diabetes [26]. In the current study, although patients with PCOS are known to be at higher risk of GDM, fewer studies have been conducted on metabolite differences between patients with PCOS and ordinary pregnant women, especially in the early stages of pregnancy. The prediction models that have been developed are usually based on the whole group of pregnant women without making a clear distinction between PCOS and ordinary pregnant women. Such models may ignore the unique metabolic profile of patients with PCOS, thus limiting their applicability to high-risk populations.

In this study, we analyzed plasma samples from 137 pregnant women in early-to-mid-gestation, divided into two groups based on PCOS status: the PCOS group ($n = 70$) and the non-PCOS group ($n = 67$). Using ultra-high-performance liquid chromatography (UHPLC) for untargeted metabolomics, we constructed early GDM prediction models for PCOS and non-PCOS patients and compared the metabolite patterns between the two groups. We hypothesize that the metabolic characteristics of PCOS patients differ significantly from those of non-PCOS patients, particularly in pathways related to GDM. This study aims to provide insights into the unique metabolic profiles of PCOS patients, establish specific GDM prediction models for this population, and promote the application of metabolomics in the early diagnosis of GDM.

Methods

This study is a prospective study based on the birth cohort in Jinan, aimed at identifying biomarkers of the risk for GDM in the mid-to-late pregnancy among women with PCOS and non-PCOS individuals, using plasma metabolomics during early pregnancy (Clinical Trial no. ChiCTR2400085868). Women with singleton pregnancies at 11^{+0}–13^{+6} weeks of gestation were re-recruited at Jinan Maternity and Child Care Hospital. The study was approved by the Ethics Review Committee of Jinan Maternity and Child Care Hospital and conducted in accordance with the Declaration of Helsinki

(No. 2024 -1- 011). All participants signed a written informed consent.

Study population and design

The inclusion and exclusion criteria for participants are as follows. Inclusion criteria: (1) gestational age $11–13^{+6}$ weeks at inclusion; (2) no history of diabetes mellitus before pregnancy; and (3) age 18–45 years. Exclusion criteria: (1) baseline fasting blood glucose ≥ 6.1 mmol/L and (2) patients with severe acute or chronic diseases, such as severe liver and kidney dysfunction, cardiovascular and cerebrovascular diseases, autoimmune diseases, blood system diseases, etc.

Diagnosis of GDM and PCOS

Hundred thirty seven pregnant women were recruited and provided clinical and laboratory data at $11–13^{+6}$ weeks of gestation. They underwent a 75 g OGTT at 24–28 weeks of gestation. The diagnosis of GDM was made according to the International Association for Diabetes and Pregnancy Study Group (IADPSG) criteria [27]: fasting glucose ≥ 5.11 mmol/L, 1-h glucose ≥ 10.00 mmol/L, or 2-h glucose ≥ 8.50 mmol/L. Using the modified Rotterdam criteria [28], PCOS can be diagnosed if any two of the following are met with the exclusion of other relevant disorders: (1) clinical or biochemical hyperandrogenism; (2) evidence of sporadic anovulation; and (3) ultrasonographic findings of a polycystic ovarian pattern.

Sample collection

5 mL peripheral vein ethylene diamine tetra-acetic acid (EDTA) blood samples were collected and centrifuged at 1600×*g* for 10 min at 4 °C. The upper plasmas were aspirated into eppendorf (EP) tubes, followed by centrifugation at 16,000×*g* for 10 min at 4 °C and stored at − 80 °C.

Instrument analysis and quality control

One hundred μL of plasma was placed in an EP tube, and added 400 μL of 80% methanol in water for each sample. The mixture was vortexed and placed in an ice bath for 5 min, then centrifuged at 15,000×*g*, 4 ℃ for 20 min. The supernatants were diluted with mass spectrometry grade water to a methanol content of 53%, then centrifuged at 15,000×*g* and 4 °C for 20 min, and the supernatants were collected and injected into liquid chromatography–mass spectrometer (LC–MS) for analysis. The samples were separated using a Vanquish UHPLC system (Thermo Scientific, Bremen, Germany) using a Hypesil Gold column (100 × 2.1 mm, 1.9 μm). Each sample was detected by electrospray ionization (ESI) in both positive and negative ion modes. The samples were separated by UHPLC and analyzed by mass spectrometry using Q Exactive™ HF/Q Exactive™ HF-X spectrometer (Thermo Scientific, Bremen, Germany). All samples were injected in a random order.

Principal component analysis (PCA) was performed to determine the separation between control (CON) and GDM groups and the separation between PCOS and PCOS–GDM groups. Differential metabolites were screened based on the VIP value of orthogonal partial least squares–discriminant analysis (OPLS–DA) and the p value of the T test. $p < 0.05$, VIP > 1, and FC > 1.2 are up-regulated, and $p < 0.05$, VIP value > 1, and FC < 1/1.2 were down-regulated. Enrichment analysis of differential metabolites was performed based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to obtain significantly enriched pathways ($p < 0.05$).

The quality control (QC) samples were prepared by mixing equal volumes of the experimental samples. The first three QCs before injection were used to monitor the instrument status and balance the chromatography–mass spectrometry system, and the next three QCs were scanned in segments. QCs inserted in the middle of the sample test were used to evaluate the system stability during the entire experiment and perform data quality control analysis.

Identification of differential metabolites

The offline data file was imported into CD 3.3 library search software for processing and simple screening of parameters such as retention time and mass-to-charge ratio was performed for each metabolite. The molecular formula was then predicted based on the molecular ion peaks and fragment ions and compared with the mzCloud, mzVault, and Masslist databases. The original quantitative results were standardized according to the formula to obtain the relative peak area, and finally, the identification and relative quantitative results of the metabolites were determined.

Model building strategy

The Boruta algorithm was used to select metabolites that were essential in classifying the two comparison groups. The top 3 metabolites were selected according to their importance in drawing receiver operating characteristic curve (ROC curve), and classification models were constructed based on the importance of the top 3 metabolites using a multivariate logistic regression algorithm. The performance of the models was evaluated by the ROC curve and the area under the curve (AUC value).

Statistical analysis

SPSS 27.0 statistical analysis software was used for clinical data analysis. The normality of continuous variables was evaluated using the Shapiro–Wilk test. Student's

t test or chi-square test and logistic regression analysis were used to evaluate the differences between GDM cases and controls in terms of continuous and categorical variables. Continuous variables with approximately normal distribution were expressed as mean ± SD. When the data did not conform to normal distribution, the median and interquartile range were used to describe the central tendency and dispersion of the data. Independent samples *t* tests were used for variables that conformed to a normal distribution and met the Chi-square assumption. When the data did not conform to normal distribution or did not meet the assumption of homogeneity of variance, the Mann–Whitney *U* test was used for inter-group comparison.

Data availability

Metabolomics data were deposited at (https://ngdc.cncb.ac.cn/omix: accession no.OMIX008162).

Results

Clinical characteristics of pregnant women in each group

The clinical characteristics of pregnant women in the normal PCOS and non-PCOS groups (participants without PCOS) are shown in Table 1. A total of 137 participants were enrolled in this prospective study. In the PCOS group, 36 participants were diagnosed with GDM in the second and third trimesters, while 34 did not develop GDM. In the non-PCOS group, 37 participants were diagnosed with GDM in the second and third trimesters, and 30 served as controls without GDM. There were no statistically significant difference in the gestational week at the time of blood sampling and fasting insulin between pregnant women without PCOS and those with PCOS. However, pregnant women with PCOS had a significantly higher body mass index (BMI) compared to the non-PCOS group. In addition, fasting blood glucose and insulin resistance indicators (IRI) were also higher in the PCOS group than in the non-PCOS group, with a statistically significant difference ($p < 0.05$).

Overall experimental process

Figure 1 shows the flow chart of the experiment. A total of 137 plasma samples were collected from pregnant women with and without PCOS, of which the PCOS population included those with PCOS alone and those with PCOS combined with GDM, and the non-PCOS group consisted of ordinary pregnant women and those with GDM. Pregnant women's plasmas were collected at $11–13^{+6}$ weeks, centrifuged, and stored in a − 80 °C refrigerator. A 75 g OGTT was performed at 24–28 weeks. Subsequently, metabolomics analysis was conducted, and differences in metabolic patterns between groups were analyzed, as well as the construction of disease prediction models.

Analysis of metabolite pattern differences between PCOS and non-PCOS groups

Retention time–accurate mass pairs were extracted from each sample profile in positive and negative ion modes. The quality control samples showed a strong correlation with $R^2 > 0.9$ (Figure S1), indicating good stability of the metabolomics detection process and high data quality. We performed unsupervised PCA analysis using the metabolic profiles of all samples. The metabolic characteristics of the PCOS were significantly different from the non-PCOS group (Fig. 2A). The OPLS–DA plot also

Table 1 Clinical characteristics in the PCOS and non-PCOS groups

Maternal characteristic	PCOS		Non-PCOS		*p* value
	PCOS–GDM	PCOS	GDM	CON	
	n = 36	*n* = 34	*n* = 37	*n* = 30	
Age (years)	30.86 ± 4.58	31.26 ± 5.09	34.0 (31.5–37.0)	35.0 (31.0–36.0)	< 0.001
BMI	26.25 ± 4.45	25.45 ± 4.43	25.42 ± 3.26	22.23 ± 3.03	0.007
Parity					
0	23 (63.9%)	23 (67.6%)	11 (29.7%)	14 (46.7%)	< 0.001
1	13 (36.1%)	9 (26.5%)	25 (67.6%)	11 (36.7%)	
≥ 2	0 (0.0%)	2 (5.9%)	1 (2.7%)	5 (16.6%)	
Gestational week at time of blood sampling	12.5 (12.10–13.20)	12.3 (11.50–12.50)	12.4 (12.10–13.15)	12.3 (11.90–13.30)	0.432
Fasting blood glucose	4.60 (4.50–4.80)	4.65 (4.58–4.80)	4.60 (4.40–4.80)	4.40 (4.10–4.60)	< 0.001
IRI	2.24 ± 0.52	1.87 ± 0.58	1.93 ± 0.68	1.72 ± 0.39	0.02
Fasting insulin	10.71 ± 2.52	9.00 ± 0.58	8.70 (7.26–12.30)	8.91 (7.46–10.30)	0.156

Data are presented as mean ± SD or median (interquartile range) or number (%)

IRI = fasting insulin × fasting glucose/22.5, *PCOS* polycystic ovary syndrome, *GDM* gestational diabetes mellitus, *CON* control, *BMI* body mass index

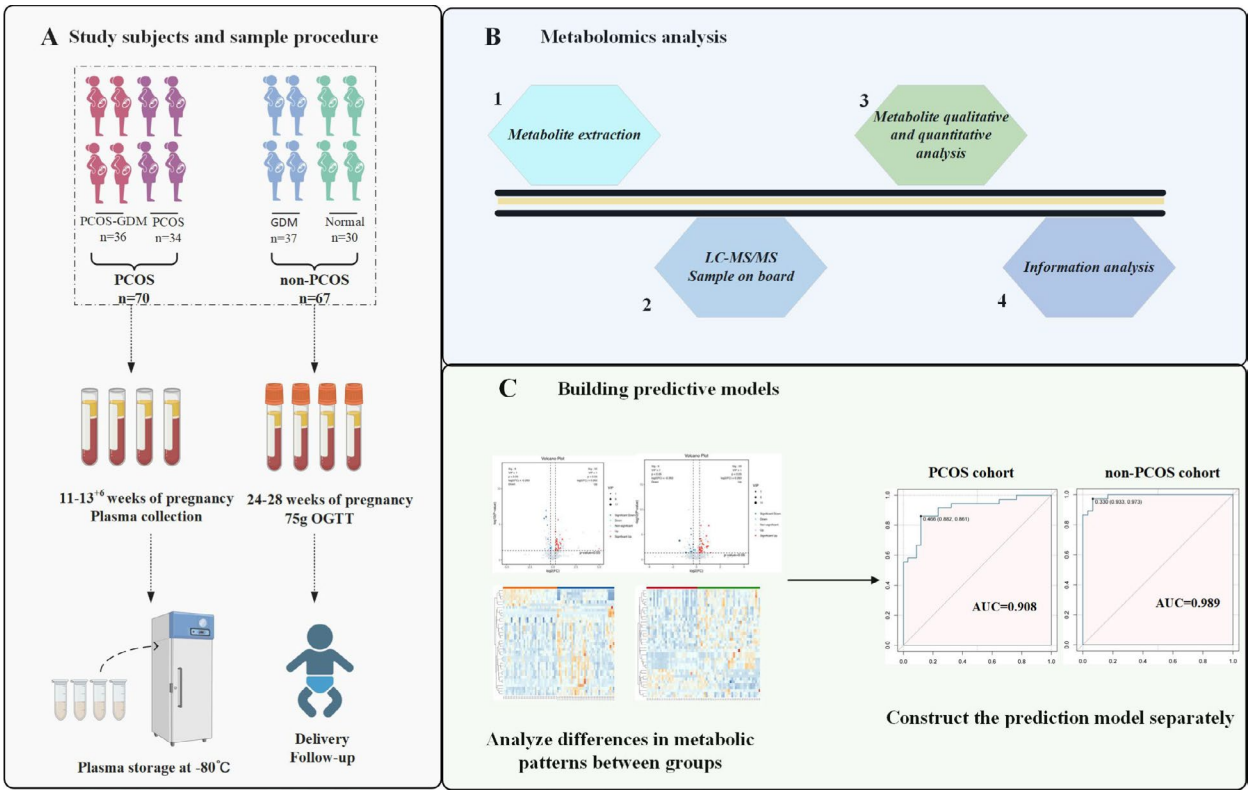

Fig. 1 Overall experimental process. **A** Study subjects and sample procedure; **B** metabolotics analysis; **C** building predictive models

showed differences in metabolic profiles between the two groups (Fig. 2B). A total of 1114 metabolites were annotated in the PCOS vs. non-PCOS groups. We set the significance threshold to screen as differential metabolites as VIP > 1.0, FC > 1.2, or FC < 0.833, and p value < 0.05. Twenty-nine metabolites were upregulated, and three were downregulated (Fig. 2C, Table S1). In addition, in the heatmaps, we observed differences in metabolite profiles between PCOS patients and non-PCOS patients (Fig. 2D). We found that the metabolic profile of PCOS patients was significantly different from that of pregnant women without PCOS.

Differences in metabolic patterns between PCOS patients and control with GDM

We further analyzed the differences that between metabolites in normal subjects and PCOS patients after developing GDM. In the PCOS–GDM vs. PCOS groups, 1114 metabolites were annotated. We set the significance threshold as VIP > 1.0, FC > 1.2, or FC < 0.833, and p value < 0.05; 32 metabolites were upregulated, and 6 were downregulated. 856 and 258 metabolites were annotated in the GDM vs. CON groups, respectively. With a significance threshold, 29 metabolites were upregulated, and 9 were downregulated (Fig. 3A, B, Tables S2, 3). Furthermore, in the heatmaps, we observed differences in metabolite profiles between PCOS patients with GDM and GDM patients (Fig. 3C, D). All differential metabolites in different comparison groups were matched to the KEGG database to obtain information on the pathways in which the metabolites were involved. Enrichment analysis was performed on the annotated results to obtain the pathways with higher enrichment of differential metabolites. The differential metabolites between the GDM group and the CON group were mainly annotated and enriched in glutathione metabolism, glyoxylate, and dicarboxylate metabolism, proximal tubule bicarbonate reclamation, citrate cycle (TCA cycle), taste transduction, etc. (Fig. 3E). Differential metabolites in the PCOS–GDM and PCOS groups were mainly annotated and enriched in the glyoxylate and dicarboxylate metabolism, citrate cycle (TCA cycle), taste transduction, central carbon metabolism in cancer, long-term depression, etc. (Fig. 3F). There were overlaps in some metabolic pathways among these comparison groups, e.g., glyoxylate and dicarboxylate metabolism, citrate cycle (TCA cycle), taste transduction. These metabolic pathways were closely related to the research objectives. We observed a clear difference in the differential metabolites between GDM and normal controls patients and between PCOS

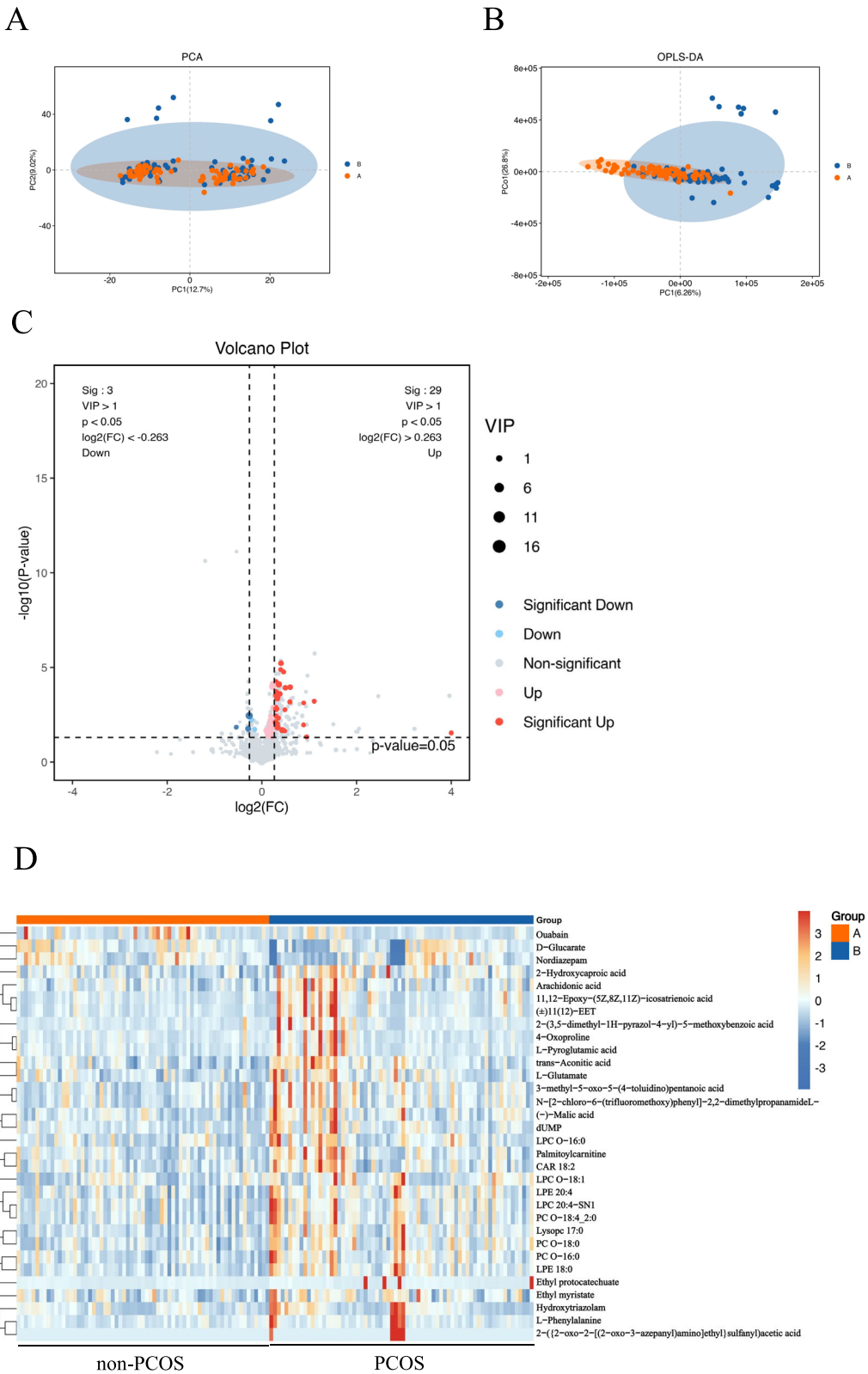

Fig. 2 Metabolite pattern differences between PCOS and non-PCOS groups. **A** PCA score plots of PCOS (blue dots) vs. non-PCOS groups (orange dots); **B** PCOS vs. non-PCOS, cumulative $R^2X = 0.453$ and $R^2Y = 0.369$; **C** volcano plots of different metabolites in PCOS vs. non-PCOS groups; **D** heatmap: hierarchical clustering analysis was performed on differential metabolites between PCOS and non-PCOS.

patients with GDM patients and PCOS. Thus, unlike the GDM prediction model for normal subjects, we needed to construct a specific GDM prediction model for PCOS patients.

Identification of biomarkers using machine learning method

Based on the Boruta algorithm, 17 variables were screened and played an essential role in separating of GDM vs. CON groups (Table S4). The ROC curves for the top five metabolites, ranked by importance, were drawn individually. The AUC values of these five metabolites were 0.901, 0.881, 0.813, 0.819, and 0.814 (Figure S2). When the top three metabolites were used to predict the separation of the two groups using logistic regression, an AUC value of 0.989 (AUC > 0.9) indicated good predictive ability (Fig. 4). Based on the boruta algorithm, a total of 12 metabolites that played an essential role in the separation of PCOS–GDM vs. PCOS groups were identified (Table S5). The ROC curves of the top five critical metabolites were drawn separately. Their AUC values were 0.815, 0.836, 0.826, 0.826, and 0.721 (Figure S3). Using the top three metabolites, ranked by importance, to predict the separation of the two groups with the logistic regression method, an AUC value of 0.908 (AUC > 0.9) demonstrated good predictive ability.

Discussion

GDM is the most common complication of pregnancy and usually manifests in the middle and late stages of pregnancy. Clinical risk factors for GDM include weight gain, increasing age, cardiovascular disease, past medical history, and PCOS [29]. GDM has been shown to have a significant impact on maternal and neonatal outcomes, including timing of delivery, birth weight, and neonatal health status. Women with PCOS are considered a high-risk group and have a significantly increased risk of developing gestational diabetes if the condition persists during pregnancy [30]. The current prevalence of GDM in pregnant women with PCOS is ~ 26%, which is significantly higher than that observed in healthy controls [31, 32]. Previous studies have demonstrated that newborns born to women with PCOS combined with GDM are more likely to experience fetal growth restriction [33, 34], amniotic fluid, premature rupture of membranes, and moderate to severe ovarian hyperstimulation syndrome [35, 36]. Predicting and preventing GDM requires identifying its associated risk factors and understanding the mechanistic links between PCOS and GDM, beyond recognizing the already established elevated risks.

Several GDM prediction models have been developed by integrating patients'general characteristics, clinical examination results, genetic information, and other relevant data. Sweeting et al. constructed a multivariate prediction model combining clinical risk factors with novel biomarkers, such as PAPP-A, triglycerides, and lipocalin- 2, achieving high predictive accuracy for early GDM with an AUC of 0.93 [37]. In addition, a meta-analysis revealed that the use of metformin can reduce the incidence of GDM, highlighting its efficacy in PCOS patients [12]. Understanding the risk factors for GDM in PCOS patients not only enables early prevention but also provides clinicians with critical tools for timely interventions.

This study identified significant metabolic differences between PCOS patients and normal pregnant women with GDM. PCA revealed distinct metabolic profiles in both the PCOS–GDM vs. PCOS and GDM vs. CON groups, which were further corroborated by supervised OPLS–DA. Among the significantly different metabolites screened, 17 were significantly altered in the GDM vs. CON group, while 13 were significantly altered in the PCOS–GDM vs. PCOS group. Interestingly, cis-aconitic acid was the only metabolite shared between the two groups, and it was upregulated in both.

In the GDM vs. CON group, downregulated metabolites included THJ2201 *N*-pentanoic acid metabolite, DL-stachydrine, spectinomycin, and orotic acid. Upregulated metabolites included compounds, such as 2-(3-fluoroanilino)-5,6-dihydro- 1,3-thiazepin-7(4*H*)-one, *N*-methyldioctylamine, PE 18:0_22:6, (2*R*)-2-[(2*R*,5*S*)-5-[(2*S*)-2-hydroxybutyl]oxolan-2-yl] propanoic acid, L-(-)-Malic acid, LPE 22:6, Tert-ButylN-[1-(aminoca- 3-methylbutyl]carbamate, 2,4-dihydroxy-heptadec- 16-en-1-yl acetate, PE 18:0_20:4, L-Glutathione oxidized, 3-Indoleacrylic acid, PC 16:0_18:0. In contrast, the PCOS-GDM vs. PCOS group demonstrated downregulation of metabolites, such as D-Glucarate,

(See figure on next page.)

Fig. 3 Metabolite pattern differences between PCOS and control patients with GDM. **A** Volcano plots of different metabolites in GDM vs. CON groups; **B** volcano plots of different metabolites in PCOS–GDM vs. PCOS groups; **C** heatmap: hierarchical clustering analysis was performed on differential metabolites between GDM and CON; **D** heatmap: hierarchical clustering analysis was performed on differential metabolites between PCOS–GDM and PCOS; **E** significantly enriched pathways in GDM vs. CON groups; **F** significantly enriched pathways in PCOS–GDM vs. PCOS groups. For bubble plots, the color of the dots reflects the enrichment *p* value, and the size reflects the count of the enriched metabolite. The impact value shows the weight of the metabolite on the pathway

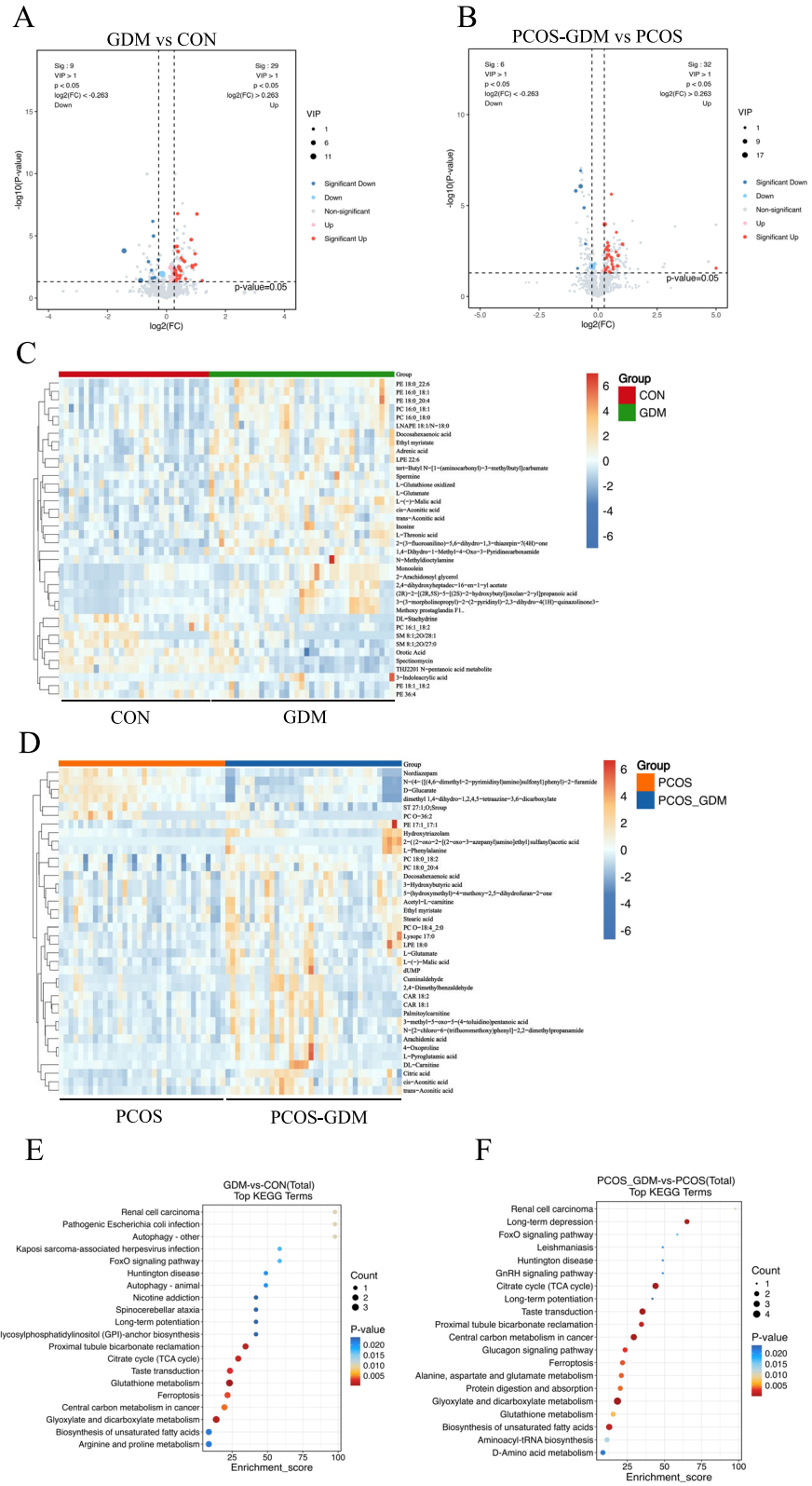

Fig. 3 (See legend on previous page.)

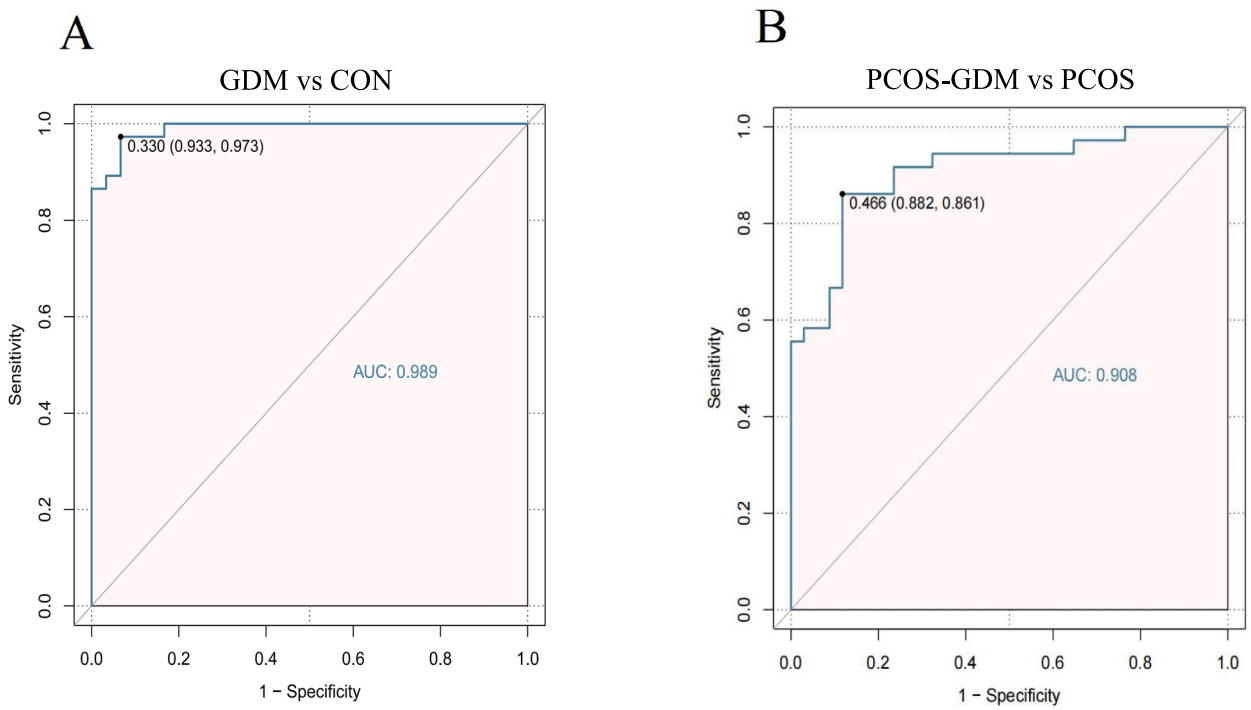

Fig. 4 Identification of biomarkers using machine learning method. **A** ROC curve of the top three metabolites in terms of importance in GDM vs. CON groups: AUC value = 0.989; **B** ROC curve of the top three metabolites in terms of importance in PCOS–GDM vs. PCOS groups: AUC value = 0.908

Nordiazepam, Dimethyl1,4-dihydro-1,2,4,5-tetraazine-3,6-dicarboxylate, *N*-(4-{[(4,6-dimethyl-2-pyrimidinyl) amino]sulfonyl}phenyl)-2-furamide, ST 27:1;O;S. Upregulated metabolites included citric acid, hydroxytriazolam, PC 18:0_20:4, trans-aconitic acid, dUMP, L-glutamate. Most of these differential metabolites were classified as lipids, organic acids, or amino acids, suggesting their involvement in pathways related to energy metabolism and glucose regulation.

Our findings align with those of previous studies. Pinto et al. reported significant changes in plasma metabolites, such as betaine, alanine, methanol, and proline in GDM patients, suggesting disruptions glycolysis, tricarboxylic acid cycle (TCA), amino acids metabolism, urea cycle, and lipid homeostasis [38]. The TCA cycle, which was significantly altered in this study, also played a critical role in the findings of O'Neill et al., who demonstrated pronounced amino acid metabolite dysregulation in the amniotic fluid of GDM patients during the second trimester, with significant changes observed in glucose, amino acids, glutathione, and fatty acids. Similarly, Lu et al. identified 13 lipid metabolites, 10 of which were strongly associated with impaired glucose tolerance [39]. Anderson et al. found that phosphatidylcholines (PC) and lysophosphatidylcholines (LPC) were positively associated with the risk of developing GDM [40]. These findings, combined with ours, suggest that PC, as a key lipid, may contribute to disruptions in blood glucose regulation.

The disadvantage of single variable prediction of GDM is that the prediction accuracy is not higher enough. Therefore, combining multiple variables, i.e., a machine learning-based approach, achieves high accuracy for GDM prediction. Research has shown that machine learning-based models can improve clinical diagnosis [41]. Studies have demonstrated that the model composed of α-hydroxybutyric acid, β-hydroxybutyric acid, and myristic acid had a strong ability to diagnose GDM in mid-pregnancy, with an AUC of 0.828 [42]. Similarly, Liu et al. developed a model with l-phenylalanyl-l-proline, hydroxylauroylcarnitine, and levoglucosan, which achieved an AUC of 0.89 in early pregnancy [43]. McMichael et al. employed targeted plasma metabolomics to predict GDM in overweight and obese pregnant women, achieving an internal validation AUC of 0.833 with α-hydroxybutyrate, sphingo-myelin 14:0, xanthine, and hypoxanthine combined [44]. In this study, we constructed a GDM prediction model for normal pregnant women using (2R)− 2-[(2R,5S)− 5-[(2S)− 2-hydroxybutyl]oxolan- 2-yl]propanoic acid,

2-(3-fluoroanilino)− 5,6-dihydro- 1,3-thiazepin- 7(4H)-one), and spectinomycin, achieving an AUC value of 0.989, indicative of high predictive performance. In addition, a specific prediction model for PCOS patients was constructed using D-glucarate, *N*-(4-{[(4,6-dimethyl-2-pyrimidinyl)amino]sulfonyl}phenyl)− 2-furamide, and trans-aconitic acid, achieving an AUC of 0.908. These results highlight the value of tailored predictive models for different subpopulations.

Despite these findings, this study has several limitations. First, the sample size was relatively small. Second, factors, such as diet and exercise, which could influence metabolic differences, were not accounted for. Finally, this study relied on untargeted metabolomics, and validation with larger; targeted data sets are warranted to confirm these findings.

Conclusions

This study identified significant differences in metabolic patterns between PCOS patients and non-PCOS patients through non-targeted metabolomics analysis, with particular emphasis on pathways related to GDM. Notably, the metabolic profiles differed substantially between PCOS and normal pregnant women, highlighting the need for tailored prediction approaches. We successfully constructed GDM prediction models for both PCOS patients and normal pregnant women, each utilizing the top three metabolic markers. The models demonstrated high predictive power, with strong potential for clinical application in early identification and management of GDM. Future studies with larger cohorts and targeted validation are warranted to refine these models and further elucidate the metabolic mechanisms underlying GDM.

Abbreviations

GDM	Gestational diabetes mellitus
PCOS	Polycystic ovary syndrome
OGTT	Oral glucose tolerance test
ML	Machine learning
IADPSG	International Association for Diabetes and Pregnancy Study Group
UHPLC	Ultra-high-performance liquid chromatography system
PCA	Principal component analysis
OPLS–DA	Orthogonal Partial Least Squares–Discriminant Analysis
KEGG	Kyoto Encyclopedia of Genes and Genomes
QC	Quality control
ROC curve	Operating characteristic curve
AUC value	The area under the curve
IRI	Insulin resistance indicators
BMI	Body mass index

Acknowledgements

Not applicable.

Author contributions

HJ, JW and CC conceived and designed the study. HJ, FH and ZW contributed to analysis and interpretation of data. FH, ZW and YP contributed to technical support. JW and CC wrote the paper. FH and ZW provided experimental techniques.

Funding

This research was funded by the Science and Technology Development Project of Jinan (No. 202225064), and the Shandong Province Medical and Health Science and Technology Project (No.202305020433).

Availability of data and materials

Metabolomics data were deposited at (https://ngdc.cncb.ac.cn/omix: accession no.OMIX008162).

Declarations

Ethics approval and consent to participate

Ethics committee approval for this study was obtained from the ethics committee of Jinan Maternity and Child Care Hospital (study no: 2024 - 1- 011).

Consent for publication

All authors gave consent for the publication.

Competing interest

The authors declare no competing interests.

Received: 9 January 2025 Accepted: 27 March 2025
Published online: 05 April 2025

References

1. Sweeting A, Hannah W, Backman H, Catalano P, Feghali M, Herman WH, et al. Epidemiology and management of gestational diabetes. The Lancet. 2024;404(10448):175–92.
2. Zhu Y, Zhang C. Prevalence of gestational diabetes and risk of progression to type 2 diabetes: a global perspective. Curr Diab Rep. 2016;16(1):7.
3. Wang H, Li N, Chivese T, Werfalli M, Sun H, Yuen L, et al. Idf diabetes atlas: estimation of global and regional gestational diabetes mellitus prevalence for 2021 by international association of diabetes in pregnancy study group's criteria. Diabetes Res Clin Pract. 2022;183: 109050.
4. Saeedi M, Cao Y, Fadl H, Gustafson H, Simmons D. Increasing prevalence of gestational diabetes mellitus when implementing the iadpsg criteria: a systematic review and meta-analysis. Diabetes Res Clin Pract. 2021;172: 108642.
5. Scheuer CM, Jensen DM, McIntyre HD, Ringholm L, Mathiesen ER, Nielsen C, et al. Applying who2013 diagnostic criteria for gestational diabetes mellitus reveals currently untreated women at increased risk. Acta Diabetol. 2023;60(12):1663–73.
6. Catalano PM, McIntyre HD, Cruickshank JK, McCance DR, Dyer AR, Metzger BE, et al. The hyperglycemia and adverse pregnancy outcome study: associations of gdm and obesity with pregnancy outcomes. Diabetes Care. 2012;35(4):780–6.
7. Simmons D, Immanuel J, Hague WM, Teede H, Nolan CJ, Peek MJ, et al. Treatment of gestational diabetes mellitus diagnosed early in pregnancy. N Engl J Med. 2023;388(23):2132–44.
8. Hannah W, Bhavadharini B, Beks H, Deepa M, Anjana RM, Uma R, et al. Global burden of early pregnancy gestational diabetes mellitus (egdm): a systematic review. Acta Diabetol. 2022;59(3):403–27.
9. Lowe WJ, Scholtens DM, Lowe LP, Kuang A, Nodzenski M, Talbot O, et al. Association of gestational diabetes with maternal disorders of glucose metabolism and childhood adiposity. JAMA. 2018;320(10):1005–16.
10. Benhalima K, Lens K, Bosteels J, Chantal M. The risk for glucose intolerance after gestational diabetes mellitus since the introduction of the iadpsg criteria: a systematic review and meta-analysis. J Clin Med. 2019;8(9):1431.
11. Bellamy L, Casas JP, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. Lancet. 2009;373(9677):1773–9.

12. Vounzoulaki E, Khunti K, Abner SC, Tan BK, Davies MJ, Gillies CL. Progression to type 2 diabetes in women with a known history of gestational diabetes: systematic review and meta-analysis. BMJ. 2020;369: m1361.
13. Deswal R, Narwal V, Dang A, Pundir CS. The prevalence of polycystic ovary syndrome: a brief systematic review. J Hum Reprod Sci. 2020;13(4):261–71.
14. Atiomo W, Rizwan M, Bajwa MH, Furniturewala HJ, Hazari KS, Harab D, et al. Prevalence and diagnosis of pcos using electronic health records: a scoping review and a database analysis. Int J Environ Res Public Health. 2024;21(3):354.
15. Foroozanfard F, Asemi Z, Bazarganipour F, Taghavi SA, Allan H, Aramesh S. Comparing pregnancy, childbirth, and neonatal outcomes in women with different phenotypes of polycystic ovary syndrome and healthy women: a prospective cohort study. Gynecol Endocrinol. 2020;36(1):61–5.
16. Kouhkan A, Khamseh ME, Moini A, Pirjani R, Valojerdi AE, Arabipoor A, et al. Predictive factors of gestational diabetes in pregnancies following assisted reproductive technology: a nested case-control study. Arch Gynecol Obstet. 2018;298(1):199–206.
17. Hao M, Yuan F, Jin C, Zhou Z, Cao Q, Xu L, et al. Overexpression of lnk in the ovaries is involved in insulin resistance in women with polycystic ovary syndrome. Endocrinology. 2016;157(10):3709–18.
18. Behboudi-Gandevani S, Amiri M, Bidhendi YR, Ramezani TF. The impact of diagnostic criteria for gestational diabetes on its prevalence: a systematic review and meta-analysis. Diabetol Metab Syndr. 2019;11:11.
19. Skiba MA, Islam RM, Bell RJ, Davis SR. Understanding variation in prevalence estimates of polycystic ovary syndrome: a systematic review and meta-analysis. Hum Reprod Update. 2018;24(6):694–709.
20. Simmons D, Sweeting A. Defining gestational diabetes: not just about cutoffs. Lancet Diabetes Endocrinol. 2023;11(5):303–4.
21. Spaight C, Gross J, Horsch A, Puder JJ. Gestational diabetes mellitus. Endocr Dev. 2016;31:163–78.
22. Zhao H, Li H, Chung A, Xiang L, Li X, Zheng Y, et al. Large-scale longitudinal metabolomics study reveals different trimester-specific alterations of metabolites in relation to gestational diabetes mellitus. J Proteome Res. 2019;18(1):292–300.
23. Zhang Z, Yang L, Han W, Wu Y, Zhang L, Gao C, et al. Machine learning prediction models for gestational diabetes mellitus: meta-analysis. J Med Internet Res. 2022;24(3): e26634.
24. Newgard CB. Metabolomics and metabolic diseases: where do we stand? Cell Metab. 2017;25(1):43–56.
25. Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. Cell Metab. 2009;9(4):311–26.
26. Chen T, Ni Y, Ma X, Bao Y, Liu J, Huang F, et al. Branched-chain and aromatic amino acid profiles and diabetes risk in Chinese populations. Sci Rep. 2016;6:20594.
27. Metzger BE, Gabbe SG, Persson B, Buchanan TA, Catalano PA, Damm P, et al. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care. 2010;33(3):676–82.
28. Dumesic DA, Oberfield SE, Stener-Victorin E, Marshall JC, Laven JS, Legro RS. Scientific statement on the diagnostic criteria, epidemiology, pathophysiology, and molecular genetics of polycystic ovary syndrome. Endocr Rev. 2015;36(5):487–525.
29. Quintanilla RB, Vadakekut ES, Mahdy H. Gestational diabetes. 2024.
30. Alfadhli EM. Gestational diabetes mellitus. Saudi Med J. 2015;36(4):399–406.
31. Li G, Huang W, Zhang L, Tian Z, Zheng W, Wang T, et al. A prospective cohort study of early-pregnancy risk factors for gestational diabetes in polycystic ovarian syndrome. Diabetes Metab Res Rev. 2018;34(5): e3003.
32. Ouyang P, Duan S, You Y, Jia X, Yang L. Risk prediction of gestational diabetes mellitus in women with polycystic ovary syndrome based on a nomogram model. Bmc Pregnancy Childbirth. 2023;23(1):408.
33. Yu T, Wu D, Cao Y, Zhai J. Association between menstrual patterns and adverse pregnancy outcomes in patients with polycystic ovary syndrome. Front Endocrinol (Lausanne). 2021;12: 740377.
34. Fougner SL, Vanky E, Lovvik TS, Carlsen SM. No impact of gestational diabetes mellitus on pregnancy complications in women with pcos, regardless of gdm criteria used. PLoS ONE. 2021;16(7): e254895.
35. Li X, Liu X, Zuo Y, Gao J, Liu Y, Zheng W. The risk factors of gestational diabetes mellitus in patients with polycystic ovary syndrome: what should we care. Medicine (Baltimore). 2021;100(31): e26521.
36. Qiu M, Qu J, Tian Y, Wang Y. The influence of polycystic ovarian syndrome on obstetric and neonatal outcomes after frozen-thawed embryo transfer. Reprod Biomed Online. 2022;45(4):745–53.
37. Sweeting AN, Wong J, Appelblom H, Ross GP, Kouru H, Williams PF, et al. A novel early pregnancy risk prediction model for gestational diabetes mellitus. Fetal Diagn Ther. 2019;45(2):76–84.
38. Pinto J, Almeida LM, Martins AS, Duarte D, Barros AS, Galhano E, et al. Prediction of gestational diabetes through nmr metabolomics of maternal blood. J Proteome Res. 2015;14(6):2696–706.
39. Lu L, Koulman A, Petry CJ, Jenkins B, Matthews L, Hughes IA, et al. An unbiased lipidomics approach identifies early second trimester lipids predictive of maternal glycemic traits and gestational diabetes mellitus. Diabetes Care. 2016;39(12):2232–9.
40. Anderson SG, Dunn WB, Banerjee M, Brown M, Broadhurst DI, Goodacre R, et al. Evidence that multiple defects in lipid regulation occur before hyperglycemia during the prodrome of type-2 diabetes. PLoS ONE. 2014;9(9): e103217.
41. Goecks J, Jalili V, Heiser LM, Gray JW. How machine learning will transform biomedicine. Cell. 2020;181(1):92–101.
42. Raczkowska BA, Mojsak P, Rojo D, Telejko B, Paczkowska-Abdulsalam M, Hryniewicka J, et al. Gas chromatography-mass spectroscopy-based metabolomics analysis reveals potential biochemical markers for diagnosis of gestational diabetes mellitus. Front Pharmacol. 2021;12: 770240.
43. Liu X, Wang X, Sun H, Guo Z, Liu X, Yuan T, et al. Urinary metabolic variation analysis during pregnancy and application in gestational diabetes mellitus and spontaneous abortion biomarker discovery. Sci Rep. 2019;9(1):2605.
44. McMichael LE, Heath H, Johnson CM, Fanter R, Alarcon N, Quintana-Diaz A, et al. Metabolites involved in purine degradation, insulin resistance, and fatty acid oxidation are associated with prediction of gestational diabetes in plasma. Metabolomics. 2021;17(12):105.

Publisher's Note