BMC Pregnancy and Childbirth

RESEARCH

Open Access

Biomarkers for congenital ventricular outflow tract malformations based on maternal serum lipid metabolomics analysis

Xuelian Yuan[1,2†], Hong Kang[1,2†], Yuqin Qin[3], Haibo Li[4], Lu Li[2], Yuting Li[2], Meixian Wang[1,2], Nana Li[1,2], Ying Deng[1,2], Xiaohong Li[1,2], Ping Yu[1,2], Yanping Wang[1,2*] and Zhen Liu[1,2*]

Abstract

Background The congenital ventricular outflow tract malformations (CVOTMs) is a major congenital heart diseases (CHDs) subtype, and its pathogenesis is complex and unclear. Lipid metabolic plays a crucial role in embryonic cardiovascular development. However, due to the limited types of detectable metabolites in previous studies, findings on lipid metabolic and CHDs are still inconsistent, and the possible mechanism of CHDs remains unclear.

Methods The nest case-control study obtained subjects from the multicenter China Teratology Birth Cohort (CTBC), and maternal serum from the pregnant women enrolled during the first trimester was utilized. The subjects were divided into a discovery set and a validation set. The metabolomics of CVOTMs and normal fetuses were analyzed by targeted lipid metabolomics. Differential comparison, random forest and lasso regression were used to screen metabolic biomarkers.

Results The lipid metabolites were distributed differentially between the cases and controls. Setting the selection criteria of *P* value < 0.05, and fold change (FC) > 1.2 or < 0.833, we screened 70 differential metabolites. Within the prediction model by random forest and lasso regression, DG (14:0_18:0), DG (20:0_18:0), Cer (d18:2/20:0), Cer (d18:1/20:0) and LPC (0:0/18:1) showed good prediction effects in discovery and validation sets. Differential metabolites were mainly concentrated in glycerolipid and glycerophospholipids metabolism, insulin resistance and lipid & atherosclerosis pathways, which may be related to the occurrence and development of CVOTMs.

Conclusion Findings in this study provide a new metabolite data source for the research on CHDs. The differential metabolites and involved metabolic pathways may suggest new ideas for further mechanistic exploration of CHDs, and the selected biomarkers may provide some new clues for detection of COVTMs.

Keywords Congenital heart defects, Lipid metabolomics, Biomarker, Machine learning

†Xuelian Yuan and Hong Kang contributed equally to this work.

*Correspondence:
Yanping Wang
wyxyanping@163.com
Zhen Liu
Jenny_liu@scu.edu.cn
[1]National Center for Birth Defects Monitoring of China, West China Second University Hospital, Sichuan University, Chengdu, Sichuan, China
[2]Key Laboratory of Birth Defects and Related Diseases of Women and Children (Sichuan University), Ministry of Education, Chengdu, Sichuan, China
[3]Department of Obstetrics and Gynecology, Maternal and Child Healthcare Hospital of Guangxi Zhuang Autonomous Region, Nanning, Guangxi, China
[4]Department of Obstetrics and Gynecology, Fujian provincial Maternal and Child Healthcare Hospital, Fuzhou, Fujian, China

Background

Congenital heart diseases (CHDs) with an estimated prevalence of 9 cases per 1000 births [1], was the leading cause of newborn death [2], and have complex phenotypes [3]. The congenital ventricular outflow tract malformations (CVOTMs) involving intrapericardial arterial trunks, the arterial roots, and the subvalvar ventricular outflow tracts, is a major CHDs subtype and account for at least one third of abnormal cardiac development [4]. Some severe CVOTMs types, such as hypoplastic left heart, Ebstein anomaly, pulmonary atresia, can lead to fetal death or neonatal cyanosis [5]. It is reported that less than 30% of CHDs are caused by clear risks, and the unknown causes are considered as the result of interactions between environmental and genetic factors [6, 7], which needs to be explored from a new perspective [8]. Biomarkers for prenatal diagnosis of CHDs have not been identified, and ultrasound echocardiography is the major diagnosis method [9], the accuracy of which highly depends on operator experience, equipment quality, and some uncontrollable conditions [10–12]. Therefore, whether a relatively objective inspection method could be developed to improve the detection of CVOTMs and the prediction effect in later progress needs to be considered.

Serving as the main energy source for cardiac physiological activities and an important source for the maturation and differentiation of cardiomyocytes [13], Fatty acids play a crucial role in embryonic development, such as cardiogenesis. As cardiomyocytes grow, mitochondria's oxidative capacity rises while glycolysis diminishes, leading the heart to increasingly rely on fatty acid β-oxidation for energy [14]. Simultaneously, lipids crucial for cell membrane synthesis and as signaling molecules rise notably during heart maturation [15]. Hence, there may be a potential link between lipid metabolism disorders and CHDs development.

Metabolomics quantitatively detects changes in all metabolites post external stimuli or genetic modification [16], and thereby discovers disease-related small molecule compounds, providing a theoretical basis for exploring the pathogenesis of maternal-fetal diseases and searching for disease-related biomarkers [17, 18]. Metabolomics has been used in etiological studies CHDs [19, 20]. Previous studies have used biological samples such as maternal serum [21], urine [22, 23], and amniotic fluid (AF) [24, 25] to detect metabolites for exploring biomarkers of CHDs. However, because of the limited types of detectable metabolites in these studies, findings are still inconsistent, and the possible mechanism CHDs remains unclear. Besides, untargeted metabolomics used in previous studies, does not focus on the quantitative detection of certain metabolites, and its detection accuracy may be insufficient.

In the current study, based on maternal serum during the first trimester, we use a new metabolomic research method involving quantitatively targeted lipid metabolomic assays to analyze the specific metabolic markers. Furthermore, we conduct machine learning algorithm to select metabolic markers that most relevant to the occurrence of CVOTMs, which aimed to provide some evidence for the etiological study of CVOTMs.

Materials and methods

Study population and sampling

Subjects in the nested case-control study were obtained from the China Teratology Birth Cohort (CTBC), established between August 2018 and December 2022 [26]. Pregnant women were recruited to participate in the cohort in the first trimester and were followed up until at least 42 days after birth. A structured questionnaire-based interview was conducted to collect the subjects' information, then fasting venous blood was collected at the beginning of enrollment after informed consent. The venous blood was centrifuged, and the supernatant and hemocyte was stored respectively in aliquots at -80 °C until analysis [26]. This study was approved by the Medical Ethics Committee of Sichuan University (K2017045) and West China Second University Hospital (2022(113)).

In this study, mothers whose fetuses were prenatally diagnosed with CVOTMs and without any other anomalies were initially chosen as the cases. The phenotype was diagnosed by sonographers, pathologists, and pediatricians through systematic ultrasound, autopsy, or postnatal follow-up. The CHDs was coded using ICD-10 coding rules. Cases with Q22, Q23, and Q25 in the codes were included in the CVOTMs cases. The subtypes of CVOTMs cases are listed in Table S1. Pregnant women who were matched for the next enrolled normal fetus and delivered at term served as controls. Both cases and controls were singletons without family history of CHDs. All information and serum samples were obtained from the project biobank. The subjects from Guangxi, Fujian, and Sichuan alliance hospitals were used as the discovery set, and the subjects from the West China Second University Hospital were used as the validation set. The grouping and analysis process are shown in Fig. 1.

Sample pretreatment

The serum samples were removed from the refrigerator at -80℃ and thawed in the ice box. After vortex mixing and centrifuged, 50μL supernatant was added into the numbered centrifuge tube, and 1mL lipid extraction solution (including the internal standard mixture) was added. After vortex for 15 min, 200 μL water was added to the mixture, which was then centrifuged at 12,000 rpm and 4℃ for 10 min. Extract 500μL of the supernatant and concentrate it. Dissolve the powder in 200 μL of the

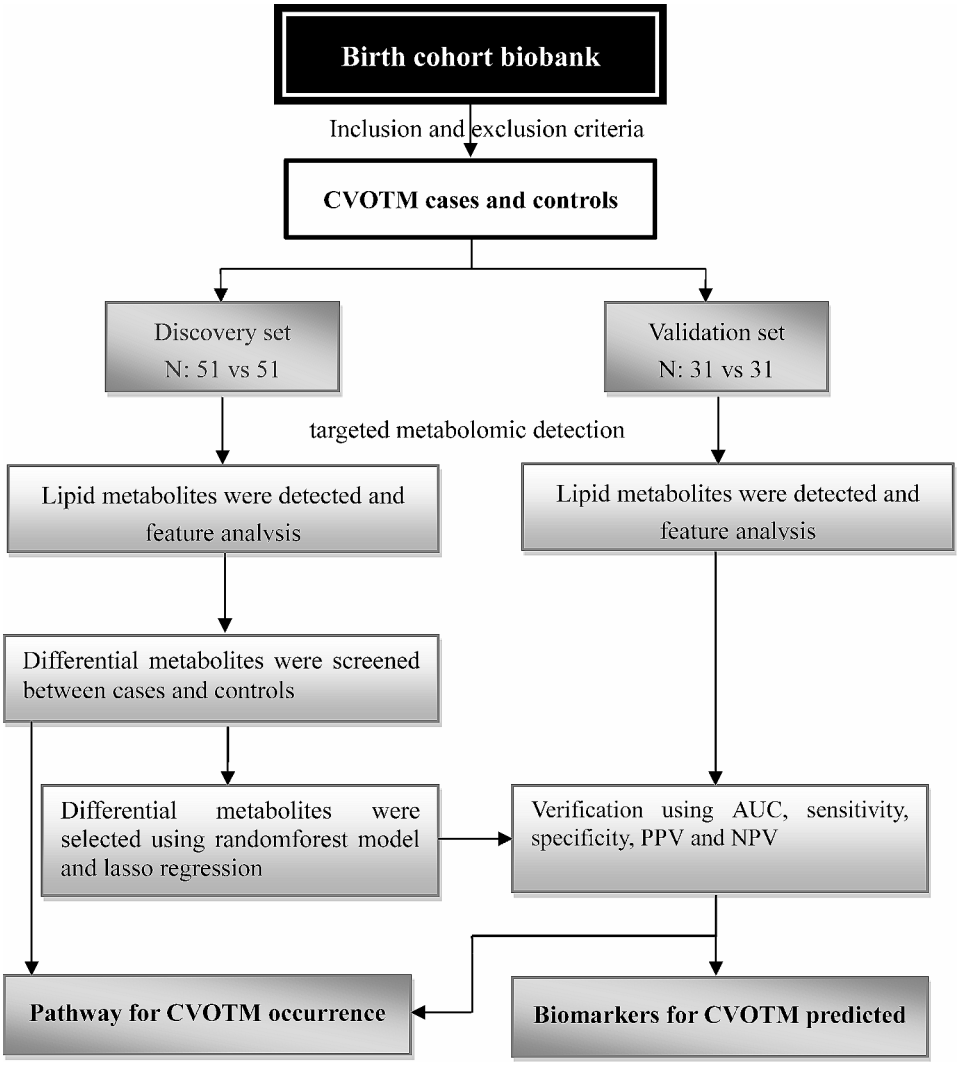

Fig. 1 The flow chart of the design and analysis of this study

reconstituted solution, then store the mixture at -80℃. Redissolved with 200 μL lipid resolution for LC-MS/MS analysis.

Metabolomic measurements

The sample extracts were analyzed using an LC-ESI-MS/MS system, which consisted of a UPLC (ExionLC AD, https://sciex.com.cn/) and a QTRAP® 6500+System MS/MS (https://sciex.com/). The liquid chromatographic column was UPLC column, Thermo Accucore™C30 (2.6 μm, 2.1 mm×100 mm i.d.). Solvent system, A: acetonitrile/water (60/40, v/v, 0.1% formic acid, 10 mM ammonium formate), B: acetonitrile/isopropanol (10/90, v/v, 0.1% formic acid, 10 mM ammonium formate); gradient program, A/B (80:20, v/v) at 0 min, 70:30 v/v at 2.0 min, 40:60 v/v at 4 min, 15:85 v/v at 9 min, 10:90 v/v at 14 min, 5:95 v/v at 15.5 min, 5:95 v/v at 17.3 min, 80:20 v/v at 17.3 min, 80:20 v/v at 20 min; flow rate, 0.35 mL/min; temperature, 45 °C; injection volume: 2 μL; the effluent was alternatively connected to an Electrospray Ionization (ESI)-triple quadrupole-linear ion trap (QTRAP) MS. LIT and triple quadrupole (QQQ) scans were obtained using a QTRAP® 6500+LC-MS/MS System. The detection rang of the instrument is 50-1200Da. The mass spectrum conditions mainly include: ESI source temperature 500 ℃, mass ionization voltage 5500 V in positive ion mode, -4500 V in negative ion mode, Ion source gas 1 (GS1) 45 psi, gas 2 (GS2) 55 psi, Curtain Gas (CUR) 35 psi. Each ion pair in the triple quadrupole is scanned and detected in accordance with their optimized Declustering Potential (DP) and Collision Energy (CE).

The detection of metabolomics was conducted at Metware Biotechnology Inc (https://www.metware.cn/). The qualitative analysis was based on the Metware database (MWDB), and was conducted according to Retention time (RT) and parent-parent ion pairs of detected

substances. Lipid quantitative analysis was performed using Multiple Reaction Monitoring (MRM) by triple quadrupole mass spectrometry.

Statistical analysis

Descriptive statistics

The personalized features included maternal age, gestational age (day), maternal prepregnancy body mass index (ppBMI), fetal gender (male/female), maternal education level (middle school or lower/high school/college or above), maternal ethnicity (Han/minority), parity (primipara/multipara), conception mode (spontaneous/assisted reproductive technology), medications taken (yes/no) and folate supplement (yes/no) during early pregnancy. Frequency (n, %) was used to describe qualitative data, and mean and standard deviation (SD) was used to describe quantitative data. Differences in the distributions of these factors between cases and controls were assessed using the chi-square test or Student's t-test.

Metabolomics data processing

Mass spectrometry (MS) raw data (.wiff) files were converted to the mzXML format by ProteoWizard. The process including peak deconvolution, alignment and integration, was processed by Analyst 1.6.3. In-house MS2 database was applied for metabolite identification.

Metabolomics data was unit variance scaled, and then analyzed using the differentially grouped principal component analysis (PCA) and unsupervised clustering analysis. The data was log transform (log_2) and mean centering before orthogonal partial least squares discriminant analysis (OPLS-DA). Significantly regulated metabolites between groups were determined by *P* value ($P<0.05$), fold change (FC, FC > 1.2 or FC < 0.83).

Metabolic biomarker screening

The dataset was composed of discovery set and validation set. Based on the significantly regulated metabolites, we conducted random forest (RF) and lasso regression in the discovery set, respectively, to select the most influential biomarkers. RF was implemented using the 'randomForest' package in R [27]. In order to obtain the optimal parameter combination of ntree and mtry, we calculated all the out-of-bag (OOB) estimate of error rate of RF models with various of combinations of mtry and ntree. The mtry and ntree parameters of the RF model were set to 62 and 23 respectively, according to the minimum OOB estimate of error rate (0.245). Markers were ranked by mean-decrease gini, and the number of important markers was determined based on the minimum error calculated by the 10-fold cross-validation method. Lasso regression was implemented using the 'glmnet' package in R. Based on the minimum mean-squared error (MSE) (0.174) calculated by the 10-fold cross-validation method (figure S5), the λ was set to 0.027, and markers with coefficients > 0 would be selected as the important markers. The intersection of markers identified by RF and lasso regression would be taken as the selected markers.

We performed logistic regression models based on demographic characteristics, selected markers, and the combination of the both, respectively. To evaluate the predictive performance of these selected markers, sensitivity, specificity, and positive predictive value (PPV), negative predictive value (NPV), and area under the curve (AUC) score were used to evaluate the performance of these models for classifying the subjects. The true positive rate (TPR) against the false positive rate (FPR) at varying classification thresholds was also presented by a ROC curve.

Metabolic pathway analysis

The classification information of differential metabolites was annotated by using the HMDB (Human Metabolome Database, https://hmdb.ca/) and KEGG ((Kyoto Encyclopedia of Genes and Genomes, https://www.kegg.jp/) databases. Enrichment analysis and statistical drawing of the annotated differential metabolites were performed.

All analyses were carried out using R version 4.2.2 (http://www.r-project.org). Two-tailed values of $P<0.05$ were considered significant.

Results

Major characteristics of the participants

After applying the inclusion and exclusion criteria to subjects with qualified biological samples, 164 samples (102 in discovery set and 62 in validation set) were ultimately recruited in the present study (Fig. 1). The demographic characteristics are presented in Table 1 and the individual information of the 164 subjects are listed in appendix table S1. We can see that maternal age, gestational days, ppBMI, ethnicity, and pregnancy status were similar between the case and control groups except for different education levels (Table 1).

Metabolomics characteristic

In the discovery set, 927 lipid metabolites were detected. The results analyzed by PCA of the total sample showed no significant difference, but the total content of lipid molecules was slightly higher in case than in the control group. This result had similar findings in the validation set. The results are shown in Supplementary figure S1.

The lipid composition was divided into 37 subtypes, and the content and chain length of DG-O, DG and MG in case group were generally higher than those in control group (Fig. 2A). For the analysis of unsaturation, the overall unsaturation of the cases were slightly higher than that of the controls, especially for LPC, SPH, Cer, Hexcer

Table 1 Comparison of demographic characteristics between two groups

Variables	Discovery set			Validation set		
	Cases	Controls	*P*value	Cases	Controls	*P*value
Total number	51	51		31	31	
Maternal age, y [a]						
Mean (SD)	30.20 (4.59)	29.70 (4.11)	0.57	31.59 (3.54)	32.15 (4.09)	0.57
Gestational age, day [a]						
Mean (SD)	79.69 (12.42)	79.29 (12.20)	0.87	91.43 (9.39)	90.69 (5.54)	0.71
Prepregnancy BMI, kg/m^2 [a]						
Mean (SD)	22.32 (3.44)	21.33 (3.22)	0.14	21.16 (2.39)	21.52 (1.85)	0.51
Fetal sex, n(%) [b]			0.12			0.79
Male	31 (64.58)	24 (47.06)		13 (41.94)	11 (35.48)	
Female	17 (35.42)	27 (52.94)		18 (58.06)	20 (64.52)	
Education, n(%)			0.03			0.34
Middle school or lower	12 (23.53)	3 (5.88)		0 (0)	1 (3.22)	
High school	10 (19.61)	17 (33.33)		1 (3.23)	3 (9.68)	
College or above	29 (56.86)	31 (60.78)		30 (96.77)	27 (87.10)	
Ethnicity, n(%)			0.27			1.00
Han	49 (96.08)	45 (88.24)		29 (93.55)	28 (90.30)	
Other	2 (3.92)	6 (11.76)		2 (6.45)	3 (9.70)	
Parity, n(%)			1.00			1.00
Primipara	31 (60.78)	30 (58.82)		22 (70.97)	21 (67.74)	
Multipara	20 (39.22)	21 (41.18)		9 (29.03)	10 (32.26)	
Conception mode, n(%)			1.00			1.00
Spontaneous	48 (94.12)	49 (96.08)		26 (83.87)	25(80.65)	
Assisted reproductive technology	3 (5.88)	2 (3.92)		5 (16.13)	6(19.35)	
Medications taken, *n* (%)			0.84			0.80
Yes	20 (39.22)	22 (43.1)		17 (54.84)	19 (61.29)	
No	31 (60.78)	29 (56.9)		14 (45.16)	12 (38.71)	
Folate supplement, n(%)			1.00			1.00
Yes	48 (94.12)	47 (92.2)		31 (100)	31 (100)	
No	3 (5.88)	4 (7.8)		0	0	

a. one case in validation set has no data on maternal age, gestational age and pre pregnancy BMI

b. three cases in discovery set have no data on fetal sex

(Fig. 2B). Similar findings were found in the validation set (figure S2).

Differential metabolite screening

Since there was no difference in results of PCA between the case and control groups, Orthogonal projections to latent structures-discriminant analysis (OPLS-DA) could not distinguish the differences between the two groups (Fig. S3). Therefore, the Variable Importance in Projection (VIP) value was not considered when analyzing the differences between the two groups. According to the screening criteria (FC>1.2, and *P*value<0.05), a total of 70 differential metabolites were screened in discovery set, most of which were upregulated (table S2). The results are shown with the volcano map in Fig. 3A.

Metabolic pathway analysis

Differential metabolites in discovery set were annotated by the KEGG (Kyoto Encyclopedia of Genes and Genomes) database and HMDB (Human Metabolome Database). The hypergeometric test was used in ClusterProfiler to perform enrichment analysis on the annotation results of the KEGG differential metabolites [28]. The results showed that the differential metabolites were mainly concentrated in glycerolipid and glycerophospholipids metabolism, insulin resistance and lipid & atherosclerosis (Fig. 3B). These metabolites were enriched ininositol phosphate metabolism, phosphatidylinositol signaling system and glycerolipid metabolism etc. pathways (Fig. 3C).

Biomarker screening and validation

Results of RF and LASSO

Figure 4A shows the mean decrease gini of the top 15 markers calculated by RF, and based on the minimum error (0.305) calculated by the 10-fold cross-validation method (figure S4), the top 9 markers were selected as the important markers, which were:

LIPID.P.0170, P.0183, P.0389, N.0234, P.0137, P.0113, N.0014, P.0357, and P.0370. Figure 4B shows the 12

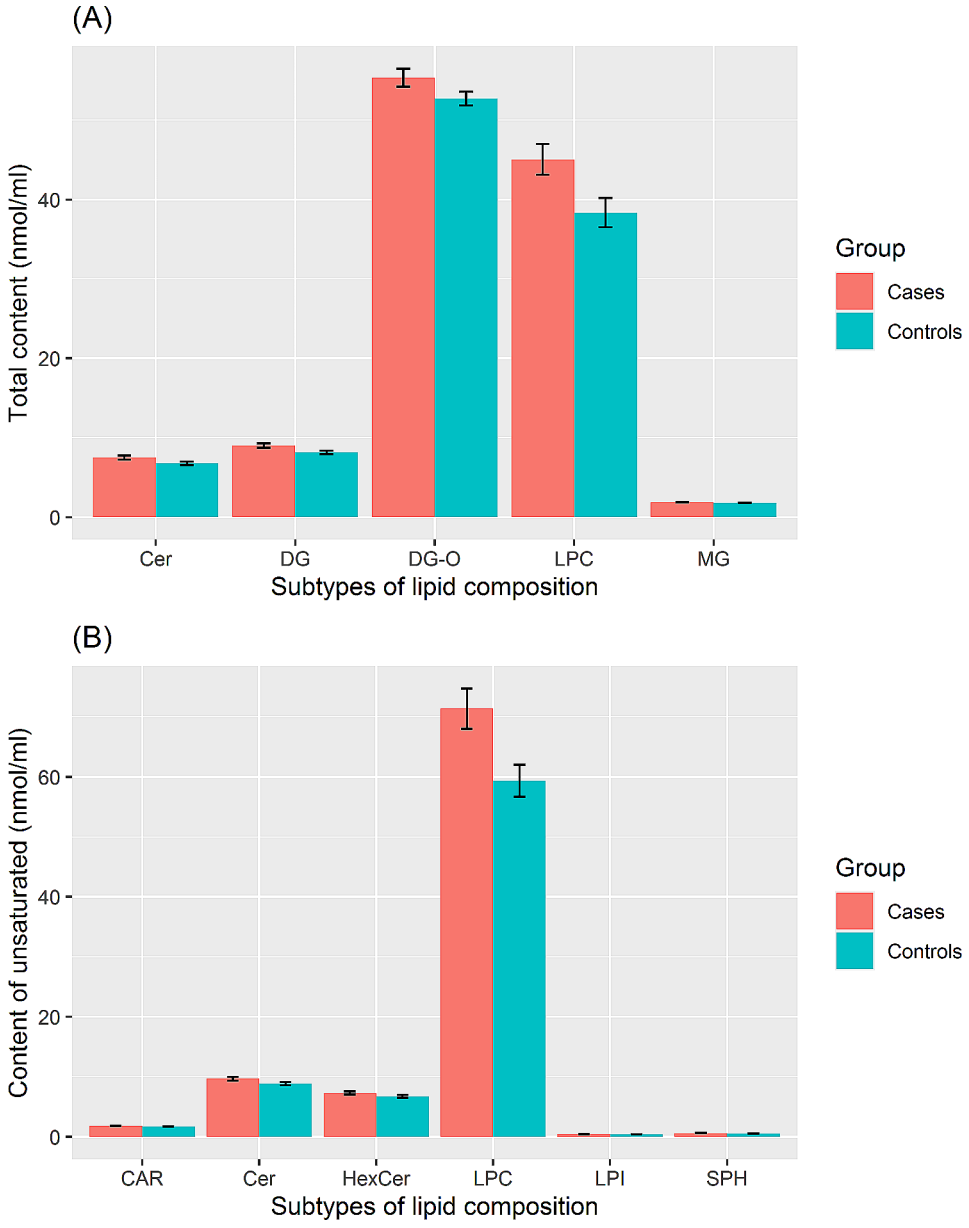

Fig. 2 Metabolomics subclass analysis between cases and controls. (**A**) the content in subclass in discovery set. (**B**) the unsaturated in subclass in discovery set between cases and control groups

markers with coefficients > 0 in lasso regression, and they were LIPID.P.0170, P.0183, N.0539, P.0370, P.0137, P.0303, N.0014, N.0836, N.0751, P.0113, P.0390 and N.0234.

The intersection of 9 markers selected by RF and 12 markers selected by lasso regression was taken, resulting in a total of 7 markers. Pathway analysis of the 7 selected biomarkers shows that they were mainly concentrated in glycerophospholipid metabolism pathways (figure S6). Table 2 shows the detailed information of each selected markers, such as compounds, class I, class II, FC, *P* value and VIP value. Considering that marker selection should consider not only the ability to distinguish cases and controls, but also the operability of the screened markers used in practical work, we further selected these markers with VIP value ≥ 1.0 as the final selected markers for prediction. Finally, 5 biomarkers were selected and they were DG (14:0_18:0), DG (20:0_18:0), Cer (d18:2/20:0), Cer (d18:1/20:0) and LPC (0:0/18:1).

Evaluation of the selected markers

Table 3 shows the performance of the logistic regression models based on demographic characteristics, selected markers, and the combination of the both, respectively. The AUC score of logistic regression models based on selected markers was 0.84 discovery set, while the AUC score of logistic regression models on demographic

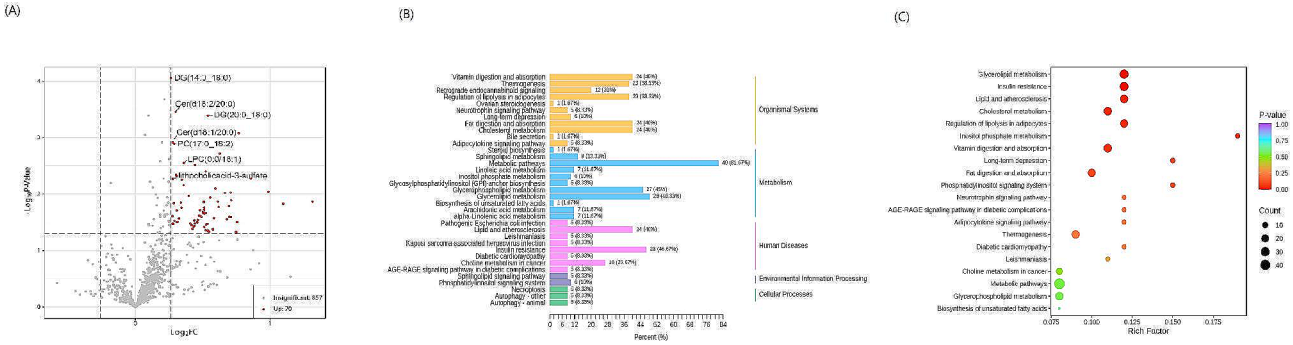

Fig. 3 Differential metabolite enrichment pathways. (**A**) the volcano plot of differential metabolite screening. (**B**) HMDB classification map of the differential metabolites in each group. (**C**) KEGG enrichment map of the differential metabolites

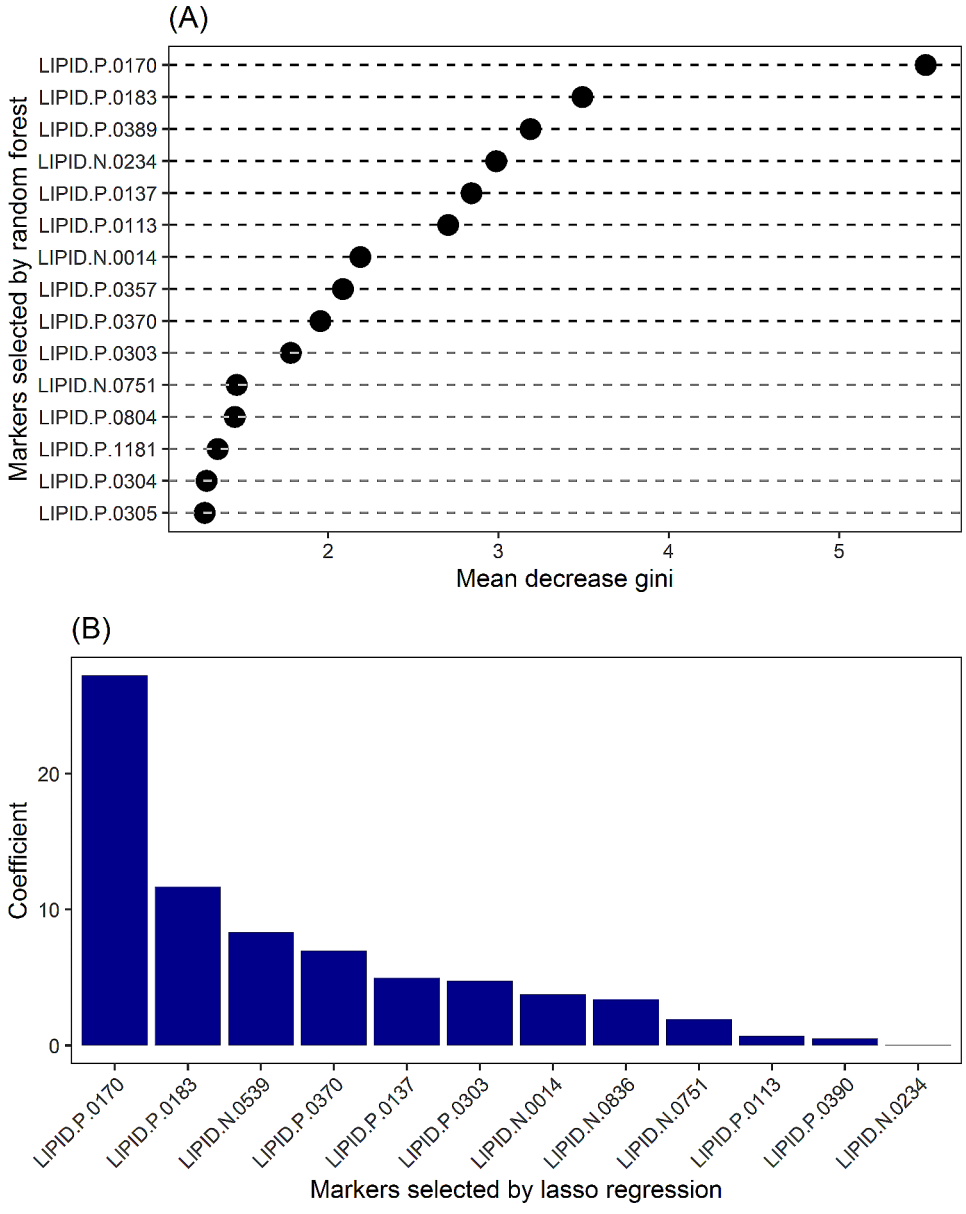

Fig. 4 The top biomarkers ranked by random forest and lasso regression. (**A**) The top 15 markers ranked by mean decrease Gini estimated by the random forest model. (**B**) The top 12 markers ranked by coefficients estimated by the lasso regression

Table 2 Common metabolic markers selected by the two methods

Index	Compounds	Class I	Class II	FC	*P*-value	VIP
LIPID.P.0170	DG (14:0_18:0)	GL	DG	1.202	< 0.0001	1.527
LIPID.P.0183	DG (20:0_18:0)	GL	DG	1.454	0.0004	1.660
LIPID.P.0137	Cer (d18:2/20:0)	SP	Cer	1.231	0.0004	1.632
LIPID.P.0113	Cer (d18:1/20:0)	SP	Cer	1.211	0.0012	1.515
LIPID.P.0370	LPC (0:0/18:1)	GP	LPC	1.282	0.0028	1.860
LIPID.N.0234	PC (17:0_18:2)	GP	PC	1.221	0.0013	0.988
LIPID.N.0014	lithocholicacid-3-sulfate	ST	BA	1.212	0.0054	0.864

FC: fold change; VIP: Variable Importance in Projection; DG: Diacylglycerol; GL: Glycerolipid; Cer: Ceramide; SP: Sphingolipid; LPC: Lysophosphatidylcholine; GP: Glycerophospholipids; PC: phosphatidylcholine; SP: Sphingolipid; ST: Sterols

Table 3 Performance of the model in discovery set and validation set

Assessment indicator	Model in discovery set			Model in validation set		
	Demographic characteristics	Markers	Markers and demographic characteristics	Demographic characteristics	Markers	Markers and demographic characteristics
AUC	0.70	0.84	0.89	0.65	0.79	0.82
Sensitivity (%)	68.63	74.51	74.51	77.42	64.52	67.74
Specificity (%)	62.75	78.43	80.39	58.06	70.97	77.42
PPV (%)	64.81	77.55	79.17	64.86	68.97	75.00
NPV (%)	66.67	75.47	75.93	72.00	66.67	70.59

AUC: area under the curve; PPV: positive predictive value; NPV: negative predictive value

characteristics was 0.70, and the AUC score increased to 0.89 when the model included the combination of demographic characteristics and selected markers. We did the same analysis in the validation set, and we got the similar results that the AUC scores of different models were 0.65, 0.79 and 0.82. Besides, the sensitivity and specificity of the model including demographic characteristics and selected markers was 67.74% and 77.42% (Table 3). Figure 5 shows the ROC curves of different models, which were consistent with the AUC scores in Table 3.

Discussion

In this study, the characteristics of maternal lipid metabolites in early pregnancy of CVOTMs and control groups were obtained through targeted lipid metabolomics detection. Lipid biomarkers for CVOTMs occurrence or development were screened and validated. The possible mechanisms for differential metabolites in the occurrence of CHD were also explored. Our results can provide a wealth of data for CVOTMs prediction and basic research from a new perspective.

Fatty acids play an important role in the occurrence of CHD, especially for the development of blood vessels [29, 30]. Therefore, this study focused on the determination and analysis of lipid metabolome. As a traditional metabolomics technique, non-targeted metabolomics mainly detects water-soluble metabolites. Instead, the targeted lipidomics analysis of the metabolomes can provide more information on lipid metabolites. Compared with the previous semi-quantitative methods and non-targeted metabolomics methods, the accurate content determination of subclasses of molecules has been greatly improved [21–24]. Generally, non-targeted metabolomics can only be qualitatively based on the second-order daughter ions of metabolites and quantitatively based on the first-order parent ions of metabolites. There is signal interference and it is difficult to distinguish accurately [31]. The characterization of non-target metabolomics can only be performed using public libraries, based on the m/z information of metabolites. The targeted metabolomics detection method can obtain high-throughput quantitative lipid metabolomics data by comparing the characteristic peak ion with the internal standard curve [32], with high sensitivity and specificity [33], which can provide new clues for prenatal diagnosis and help to discover affected metabolic pathways and reveal the pathogenesis of diseases [18, 34].

The results found that the content of DG-O, DG and MG were higher in case group, and the unsaturated bond content of LPC, SPH, and Cer was higher than those of the control group. These suggest that the changes of subclasses DG-O, DG, MG, LPC, SPH, and Cer in the case group may be related to the development of the cardiovascular system. Other related studies also found differential lipid metabolites, but the subtypes were slightly different. Bahado et al. [21]. detected metabolomics markers in the first trimester serum of pregnant women and found that there were lipid metabolism abnormalities in the case group of fetal CHD, including phosphatidylcholine, carnitine, sphingolipids and choline. A urine metabolomics study on mothers of fetuses with congenital heart disease and 20 potential biomarkers, including

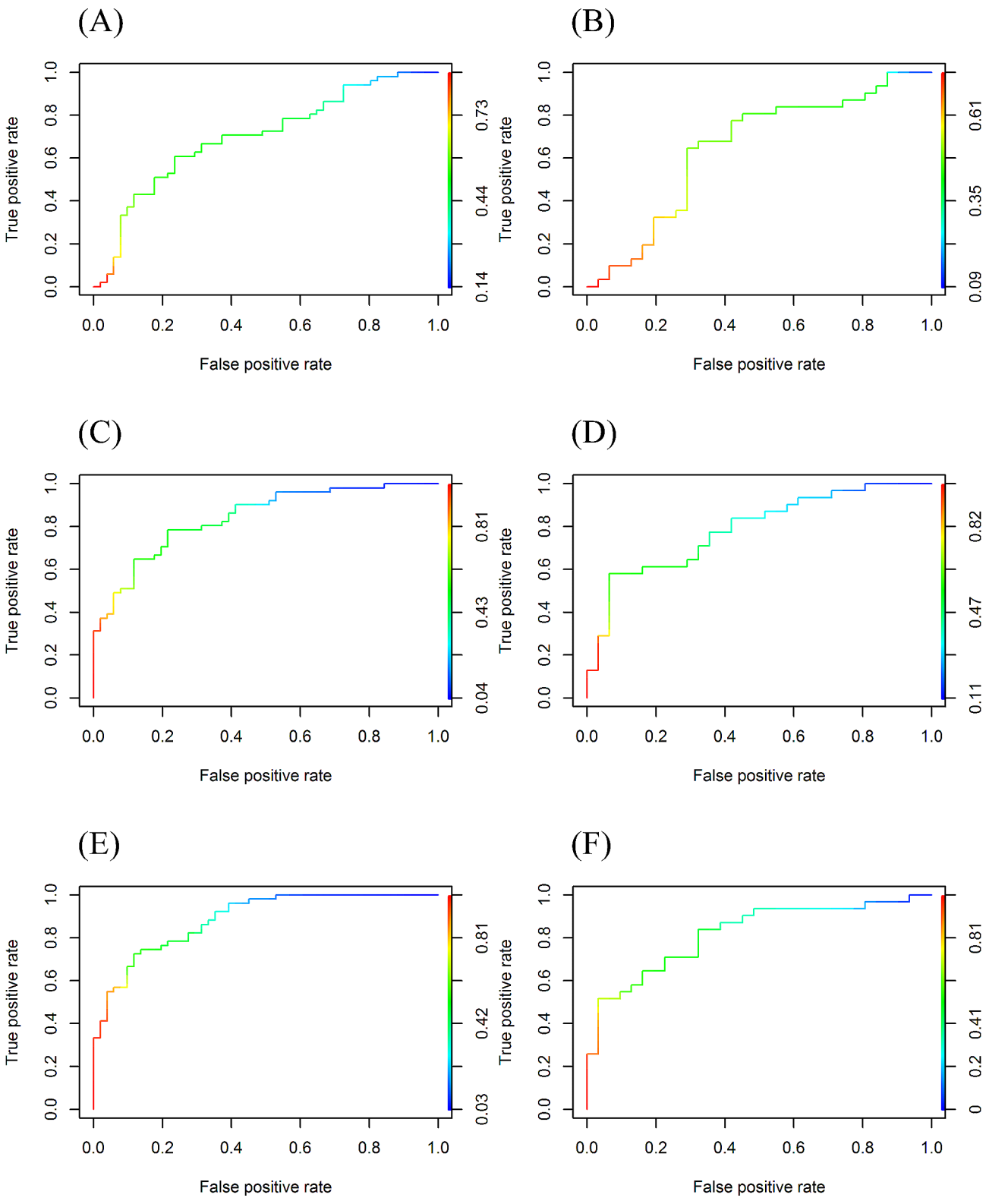

Fig. 5 The ROC curves for classifying the subjects in the logistic regression models. (**A**) and (**B**) show the ROC curve of the logistic regression model based on demographic characteristics for classifying the subjects in discovery set and validation set, respectively; (**C**) and (**D**) show the ROC curve of the logistic regression model based on 5 markers for classifying the subjects in discovery set and validation set, respectively; (**E**) and (**F**) show the ROC curve of the logistic regression model based on the combination of 5 markers and demographic characteristics for classifying the subjects in discovery set and validation set, respectively

short-chain fatty acids (SCFA), were screened out [23]. Metabolomics analysis of bicuspid aortic valve (BAV) cases and controls with congenital heart defects detected a total of 41 differential metabolites, of which three metabolites glycerophosphole-N-oleoyl ethanolamine, glycerophospholide monoester (18:2) and phosphatidylethanolamine (18:2) were most closely related to BAV [19]. Metabolomics studies on children with CHD also found that the metabolic pathways of serum amino acids and choline were closely related to CHD [20]. Although the above-mentioned studies did not screen the same biomarker as this project, the results all suggest that lipid metabolism is closely related to the occurrence of CHD.

Multiple machine learning including the RF, and LASSO were exploited to screen the biomarkers. The RF is an ensemble learning method that operates by constructing a collection of decision trees [35, 36]. For variable selection, RF performs well in dealing with complex nonlinear relationships and high-dimensional data, but it may be not ideal to classify small sample data. LASSO is good at feature selection and handling collinearity problems, and is suitable for situations where there are few features and the prediction model needs to be simple and explanatory. Considering the accuracy of model prediction and the interpretability of feature selection, the intersection of biomarkers selected by RF and LASSO is taken as the selected biomarkers. Besides, we further selected the final biomarkers with VIP value > 1.0 from the biomarkers selected by the above methods of machine learning.

Through machine learning, 5 biomarkers in maternal serum perform well in distinguishing cases from controls. They are two Diacylglycerol (DG), two Ceramide (Cer), and one Lysophosphatidylcholine (LPC). DG is a critical intermediate in lipid metabolism, which is not reabsorbed in the small intestine but is broken down into monoglycosides (MAG) and free fatty acids (FFA) by pancreatic lipase. It has physiological functions such as inhibiting the increase of postprandial blood lipids, reducing the accumulation of body fat, reducing body weight and regulating blood glucose [37]. Cer may induce vasodilator or vasoconstrictor effects by interacting with several signaling pathways in endothelial and smooth muscle cells, but is known to increase ROS production [38]. Ceramides play an important role in the development of atherosclerotic and valvular heart disease. Lowering cellular and tissue levels of ceramide by inhibiting the Cer-producing enzymes counteracts atherosclerotic and valvular heart disease development in animal models [39]. In humans, elevated blood Cer levels are associated with cardiovascular events. Important cardiovascular risk factors, such as obesity and diabetes, have also been linked to Cer accumulation [40]. Lysophosphatidylcholine (LPC) is increasingly recognized as a key factor positively associated with cardiovascular diseases like atherosclerosis. Lysophosphatidylcholine can induce the proliferation and migration of vascular smooth muscle cells by promoting the expression of chemokines in vascular smooth muscle cells [41]. The above three types of metabolites may all play a role in the occurrence of cardiovascular disease. However, the effects of these substances on embryonic heart development may be first be reported.

This study found that the differential metabolites were mainly concentrated in several metabolic pathways, such as inositol phosphate metabolism, phosphatidylinositol signaling and glycerolipid metabolism etc. pathways. It was inferred that glycerolipid and glycerophospholipids metabolism, insulin resistance and lipid & atherosclerosis played important roles in the occurrence and development of CHD.

The study creatively proposed maternal serum lipid metabolism markers that are closely related to the occurrence of fetal CVOTMs, which has certain advantages. Firstly, cohort samples were effectively used, and the serum samples in early pregnancy could better reflect maternal metabolic status during fetal heart development. Secondly, we focused on CVOTMs, a subtype of CHD, whose occurrence and development are closely related to lipid metabolism, so that the interpretation of the results is more persuasive. Thirdly, the quantitative targeted lipid metabolomics detection method was used to obtain more accurate concentration for metabolites distribution characteristics, which makes the results more accurate and reliable.

Of course, this study also has some shortcomings. Firstly, the sample size is relatively small, and the characteristics the of groups are somewhat inconsistent, but the key indicators were basically consistent between groups. Then, though the predictive performance of these selected markers in validation set was not outstanding enough, the validation of this study is an external validation and the results were more favorable for extrapolation. Finally, more detailed information on maternal and fetus (e.g., biological samples of fetuses) were not available in the study, which may affect the results in the study. Anyway, the study is only a preliminary screening of biomarkers in maternal blood for the detection of fetal COVTMs, and these markers need to be further validated based on large samples and consideration of more detailed fetal information (e.g., biological samples of fetuses). Future studies should focus on larger sample sizes with fetuses related samples included for in-depth analysis and validation.

Conclusions

These results would provide additional new metabolite data source for the research on CHDs, especially for COVTMs. The differential metabolites and involved metabolic pathways may suggest new ideas for further mechanistic exploration of CHDs, and the selected biomarkers may provide some new clues for detection of COVTMs.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12884-024-06738-y.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

The authors are indebted to the obstetricians, sonographer, prenatal diagnosis specialist, investigator, and experimental technicians who were involved in the project for recruiting the case and control participants and collecting the data. We thank all participating families for their cooperation and for providing personal information and biosamples.

Author contributions

Zhen Liu and Yanping Wang developed the study design and conducted the study; Xuelian Yuan and Hong Kang assisted in organizing and implementing the project and drafting the manuscript; Lu Li, Meixian Wang, Yuqin Qin, and Haibo Li contributed to recruiting participants and diagnosing cases; Xuelian Yuan, Hong Kang and Xiaohong Li performed data analysis and interpretation; Lu Li, Yuting Li, Meixian Wang, Nana Li, and Ping Yu participated in experimental detection and summarizing the data. Zhen Liu and Yanping Wang checked and revised the manuscript. All authors read and approved the final manuscript.

Funding

This study has received financial support from the National Natural Science Foundation of China (No. 82103858), the National Key Research and Development Program of China (No. 2016YFC1000102), the Applied Basic Research Program of Sichuan Province (No.2021YJ0212) and the Popularization Application Project (No.21PJ057).

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The program was approved by the Ethics Committee of Sichuan University (K2017045) and West China Second University Hospital (2022(113)). All subjects provided informed consent to participate. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not Applicable.

Competing interests

The authors declare no competing interests.

Received: 29 February 2024 / Accepted: 5 August 2024
Published online: 20 August 2024

References

1. Liu Y, Chen S, Zuhlke L, Black GC, Choy MK, Li N, Keavney BD. Global birth prevalence of congenital heart defects 1970–2017: updated systematic review and meta-analysis of 260 studies. Int J Epidemiol. 2019;48(2):455–63.
2. Collaborators GBDU-M. Global, regional, and national progress towards sustainable development goal 3.2 for neonatal and child health: all-cause and cause-specific mortality findings from the global burden of Disease Study 2019. Lancet. 2021;398(10303):870–905.
3. Giang KW, Mandalenakis Z, Fedchenko M, Eriksson P, Rosengren A, Norman M, Hanseus K, Dellborg M. Congenital heart disease: changes in recorded birth prevalence and cardiac interventions over the past half-century in Sweden. Eur J Prev Cardiol. 2023;30(2):169–76.
4. Anderson RH, Mori S, Spicer DE, Brown NA, Mohun TJ. Development and morphology of the ventricular outflow tracts. World J Pediatr Congenit Heart Surg. 2016;7(5):561–77.
5. Parikh LI, Grantz KL, Iqbal SN, Huang CC, Landy HJ, Fries MH, Reddy UM. Neonatal outcomes in fetuses with cardiac anomalies and the impact of delivery route. Am J Obstet Gynecol. 2017;217(4):469. e461-469 e412.
6. Jerves T, Beaton A, Kruszka P. The genetic workup for structural congenital heart disease. Am J Med Genet C Semin Med Genet. 2020;184(1):178–86.
7. Mat Bah MN, Sapian MH, Jamil MT, Abdullah N, Alias EY, Zahari N. The birth prevalence, severity, and temporal trends of congenital heart disease in the middle-income country: a population-based study. Congenit Heart Dis. 2018;13(6):1012–27.
8. Chen L, Guan J, Wei Q, Yuan Z, Zhang M. Potential role of omics technique in prenatal diagnosis of congenital heart defects. Clin Chim Acta. 2018;482:185–90.
9. Chitra N, Vijayalakshmi IB. Fetal echocardiography for early detection of congenital heart diseases. J Echocardiogr. 2017;15(1):13–7.
10. Garcia Delgado R, Garcia Rodriguez R, Ortega Cardenes I, Gonzalez Martin JM, De Luis Alvarado M, Segura Gonzalez J, Medina Castellano M, Garcia Hernandez JA. Feasibility and accuracy of early fetal Echocardiography Performed at 13(+0)-13(+6) weeks in a Population with Low and High Body Mass Index: a prospective study. Reprod Sci. 2021;28(8):2270–7.
11. McBrien A, Hornberger LK. Early fetal echocardiography. Birth Defects Res. 2019;111(8):370–9.
12. Pacileo G, Di Salvo G, Limongelli G, Miele T, Calabro R. Echocardiography in congenital heart disease: usefulness, limits and new techniques. J Cardiovasc Med (Hagerstown). 2007;8(1):17–22.
13. Burugupalli S, Smith AAT, Oshlensky G, Huynh K, Giles C, Wang T, George A, Paul S, Nguyen A, Duong T et al. Ontogeny of circulating lipid metabolism in pregnancy and early childhood - a longitudinal population study. Elife 2022, 11.
14. Lopaschuk GD, Jaswal JS. Energy metabolic phenotype of the cardiomyocyte during development, differentiation, and postnatal maturation. J Cardiovasc Pharmacol. 2010;56(2):130–40.
15. Iruretagoyena JI, Davis W, Bird C, Olsen J, Radue R, Teo Broman A, Kendziorski C, Splinter BonDurant S, Golos T, Bird I, et al. Metabolic gene profile in early human fetal heart development. Mol Hum Reprod. 2014;20(7):690–700.
16. Vora N, Kalagiri R, Mallett LH, Oh JH, Wajid U, Munir S, Colon N, Raju VN, Beeram MR, Uddin MN. Proteomics and metabolomics in Pregnancy-An overview. Obstet Gynecol Surv. 2019;74(2):111–25.
17. Liang L, Rasmussen MH, Piening B, Shen X, Chen S, Rost H, Snyder JK, Tibshirani R, Skotte L, Lee NC, et al. Metabolic Dynamics and Prediction of gestational age and Time to delivery in pregnant women. Cell. 2020;181(7):1680–e16921615.
18. Parfieniuk E, Zbucka-Kretowska M, Ciborowski M, Kretowski A, Barbas C. Untargeted metabolomics: an overview of its usefulness and future potential in prenatal diagnosis. Expert Rev Proteom. 2018;15(10):809–16.
19. Wang W, Maimaiti A, Zhao Y, Zhang L, Tao H, Nian H, Xia L, Kong B, Wang C, Liu M, et al. Analysis of serum metabolites to diagnose bicuspid aortic valve. Sci Rep. 2016;6:37023.
20. Yu M, Sun S, Yu J, Du F, Zhang S, Yang W, Xiao J, Xie B. Discovery and Validation of potential serum biomarkers for Pediatric patients with congenital Heart diseases by Metabolomics. J Proteome Res. 2018;17(10):3517–25.
21. Bahado-Singh RO, Ertl R, Mandal R, Bjorndahl TC, Syngelaki A, Han B, Dong E, Liu PB, Alpay-Savasan Z, Wishart DS et al. Metabolomic prediction of fetal congenital heart defect in the first trimester. *Am J Obstet Gynecol* 2014, 211(3):240 e241-240 e214.
22. Friedman P, Yilmaz A, Ugur Z, Jafar F, Whitten A, Ustun I, Turkoglu O, Graham S, Bahado Singh R. Urine metabolomic biomarkers for prediction of isolated fetal congenital heart defect. J Matern Fetal Neonatal Med 2021:1–8.

23. Xie D, Luo Y, Xiong X, Lou M, Liu Z, Wang A, Xiong L, Kong F, Wang Y, Wang H. Study on the Potential Biomarkers of Maternal Urine Metabolomics for Fetus with Congenital Heart Diseases Based on Modified Gas Chromatograph-Mass Spectrometer. *Biomed Res Int* 2019, 2019:1905416.
24. Li Y, Sun Y, Yang L, Huang M, Zhang X, Wang X, Guan X, Yang P, Wang Y, Meng L, et al. Analysis of biomarkers for congenital heart Disease based on maternal amniotic fluid metabolomics. Front Cardiovasc Med. 2021;8:671191.
25. Yuan X, Li L, Kang H, Wang M, Zeng J, Lei Y, Li N, Yu P, Li X, Liu Z. Biomarkers for isolated congenital heart disease based on maternal amniotic fluid metabolomics analysis. BMC Cardiovasc Disord. 2022;22(1):495.
26. Zhou Y, Tao J, Wang K, Deng K, Wang Y, Zhao J, Chen C, Wu T, Zhou J, Zhu J, et al. Protocol of a prospective and multicentre China Teratology Birth Cohort (CTBC): association of maternal drug exposure during pregnancy with adverse pregnancy outcomes. BMC Pregnancy Childbirth. 2021;21(1):593.
27. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random forest: a classification and regression tool for compound classification and QSAR modeling. J Chem Inf Comput Sci. 2003;43(6):1947–58.
28. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci. 2019;28(11):1947–51.
29. Cerf ME. High Fat Programming and Cardiovascular Disease. Med (Kaunas) 2018, 54(5).
30. Nasioudis D, Doulaveris G, Kanninen TT. Dyslipidemia in pregnancy and maternal-fetal outcome. Minerva Ginecol. 2019;71(2):155–62.
31. Chen S, Kong H, Lu X, Li Y, Yin P, Zeng Z, Xu G. Pseudotargeted metabolomics method and its application in serum biomarker discovery for hepatocellular carcinoma based on ultra high-performance liquid chromatography/triple quadrupole mass spectrometry. Anal Chem. 2013;85(17):8326–33.
32. Gika H, Virgiliou C, Theodoridis G, Plumb RS, Wilson ID. Untargeted LC/MS-based metabolic phenotyping (metabonomics/metabolomics): the state of the art. J Chromatogr B Analyt Technol Biomed Life Sci. 2019;1117:136–47.
33. Albreht A, Hussain H, Jimenez B, Yuen AHY, Whiley L, Witt M, Lewis MR, Chekmeneva E. Structure elucidation and mitigation of endogenous interferences in LC-MS-Based metabolic profiling of urine. Anal Chem. 2022;94(3):1760–8.
34. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol. 2016;17(7):451–9.
35. Olshen RA, Breiman L, Friedman J, Stone CJ. Classification and regression trees. Chapman and Hall; 1984.
36. Michel M, Laser KT, Dubowy KO, Scholl-Burgi S, Michel E. Metabolomics and random forests in patients with complex congenital heart disease. Front Cardiovasc Med. 2022;9:994068.
37. Ando Y, Saito S, Yamanaka N, Suzuki C, Ono T, Osaki N, Katsuragi Y. Alpha linolenic acid-enriched Diacylglycerol Consumption enhances Dietary Fat Oxidation in healthy subjects: a Randomized double-blind controlled trial. J Oleo Sci. 2017;66(2):181–5.
38. Cogolludo A, Villamor E, Perez-Vizcaino F, Moreno L. Ceramide and Regulation of Vascular Tone. Int J Mol Sci 2019, 20(2).
39. Li Y, Talbot CL, Chaurasia B. Ceramides in Adipose tissue. Front Endocrinol (Lausanne). 2020;11:407.
40. Carvalho LSF, Chaves-Filho AB, Yoshinaga MY. Orchestrating a ceramide-phosphatidylcholine cardiovascular risk score: it ain't over 'til the fat layer sings. Eur J Prev Cardiol. 2022;29(6):892–4.
41. Liu P, Zhu W, Chen C, Yan B, Zhu L, Chen X, Peng C. The mechanisms of lysophosphatidylcholine in the development of diseases. Life Sci. 2020;247:117443.

Publisher's Note