



OPEN Precision fetal cardiology detects cyanotic congenital heart disease using maternal saliva metabolome and artificial intelligence

Ray Bahado-Singh¹, Nadia Ashrafi², Amin Ibrahim², Buket Aydas³, Ali Yilmaz², Perry Friedman¹, Stewart F. Graham^{1,2} & Onur Turkoglu⁴✉

Prenatal sonographic diagnosis of congenital heart disease (CHD) can lead to improved morbidity and mortality. However, the diagnostic accuracy of ultrasound, the sole prenatal screening tool, remains limited. Failed prenatal or early newborn detection of cyanotic CHD (CCHD) can have disastrous consequences. We therefore sought to use a Precision Fetal Cardiology based approach combining metabolomic profiling of maternal saliva and machine learning, a major branch of artificial intelligence (AI), for the prenatal detection of isolated, non-syndromic cyanotic CHD. Metabolomic analyses using Ultra-High Performance Liquid Chromatography/Mass Spectrometry identified 468 metabolites in the saliva. Six different AI platforms were utilized for the detection of CCHD and CHD overall. AI achieved excellent accuracy for the CCHD detection: Area Under the ROC curve: AUC (95% CI) = 0.819 (0.635-1.00) with a sensitivity and specificity of 92.5% and 87.0%, and for CHD overall: AUC (95% CI) = 0.828 (0.635-1.00) with a sensitivity of 90.5% and specificity of 88.0%. Similarly high accuracies were achieved for the detection of CHD overall: AUC (95% CI) = 0.8488 (0.635-1.00) with a sensitivity of 92.5% and specificity of 91.0%. Pathway analysis showed significant alterations in Arachidonic Acid, Alpha-linoleic acid, and Tryptophan metabolism indicating significant lipid dysfunction in cyanotic CHD. In summary, we report for the first time, the accurate detection of non-syndromic cyanotic CHD using maternal salivary metabolomics. Further, analysis revealed significant alteration of lipid metabolism.

Congenital heart disease (CHD) is the most common severe congenital abnormality and accounts for more than half of childhood mortality cases related to birth defects. The prevalence of CHD is around 8 to 9 per 1000 live births in large population studies, however, this number may be higher based on the population examined¹. Prenatal diagnosis of CHD improves perinatal management and perioperative mortality² hence the focus on detection of these disorders. Current prenatal detection is based entirely on mid-trimester ultrasounds with a fetal ECHO performed in those with suspicious ultrasound findings or based on the presence of clinical or historical risk factors. However, most of CHD cases are diagnosed in patients without known risk factors and only 50% of CHD cases overall are detected prenatally³. Late diagnosis can result in increased neonatal and pediatric morbidity and mortality. Cyanotic CHD (CCHD) constitutes a special subset of and accounts for approximately 25% of CHD cases. This group is characterized by significantly higher morbidity and mortality and requires urgent newborn care compared non-cyanotic CHD. The traditional definition of CCHD commonly included Tetralogy of Fallot (TOF), transposition of the great arteries (TGA), tricuspid atresia, truncus arteriosus, total anomalous pulmonary venous return (TAPVR), and hypoplastic left heart syndrome (HLHS). However, there are a variety of lesions that present with cyanosis that are not covered by this definition⁴. Although fetuses with CCHDs remain clinically stable, most neonates become critically ill with ductus arteriosus closure after birth. Delays in recognition of these ductal dependent lesions lead to neonatal metabolic acidosis which increases neonatal morbidity and mortality even after the surgical intervention⁵. The prenatal detection of CCHDs

¹Department of Obstetrics and Gynecology, Corewell Health William Beaumont University Hospital, Oakland University William Beaumont School of Medicine, Royal Oak, MI 48073, USA. ²Metabolomics Department, Corewell Health William Beaumont University Hospital, Beaumont Research Institute, Royal Oak, MI 48073, USA. ³Department of Care Management Analytics, Blue Cross Blue Shield of Michigan, Detroit, MI 48226, USA. ⁴Department of Obstetrics and Gynecology, Baylor College of Medicine, Texas Children's Hospital, Houston, TX, USA. ✉email: dronurturkoglu@gmail.com

permits meaningful alternations in perinatal care plans such as maternal transfer to tertiary centers capable of maintaining newborn ductus arteriosus patency and providing surgical and other interventions in unstable newborns⁶.

While genetic and chromosomal factors play an important role in some CHD, in the majority of cases no identifiable causes or risk factors are present⁷. Environmental factors such as drugs and alcohol exposure are known to play a significant role in the development of CHD. Epigenetics, the study of the interaction between the environment and genes, have recently been used for the accurate prediction of isolated, non-syndromic CHD^{8–10}, particularly when combined with machine learning an important branch of artificial intelligence (AI) analysis of circulating cell-free DNA⁸. Relatively little work has so far been performed on the metabolomics of CHD, in particular during the fetal period. Our group and others have demonstrated the feasibility of minimally and non-invasive prenatal detection of CHD based on metabolomic analysis of maternal blood or urine^{11–14}. In addition, recent studies in the newborn have reported significant metabolic alterations in CCHD cases^{15,16}. There is, however, a paucity of published data on the prenatal biomarker detection of CCHD.

Saliva is a complex fluid that consists of various compounds including proteins, metabolites, and lipids. Most compounds found in blood are also present in saliva¹⁷. Saliva sampling is non-invasive, convenient and potentially highly suited for population screening, and can provide insight into health and disease states¹⁸. For these reasons, there is interest in the study of salivary metabolomics and the development of accurate markers for the detection of diverse disorders examples of which include cancers and neurodegenerative disorders¹⁹. Recently, untargeted salivary metabolomics was shown to provide accurate detection of gestational diabetes²⁰. However, to our knowledge, there are no studies related to the use of maternal salivary metabolomics for the detection of fetal anomalies including CHDs.

Precision Medicine (PM) seeks to develop and tailor the optimal treatment to the individual patient or subgroup through the development of accurate biomarkers. “Omics” science has been identified as being pivotal for the development of PM^{21,22}. In addition, sophisticated computational tools such as artificial intelligence (AI) methods will be needed to interpret the ‘big’ data generated from omics experiments. AI techniques appear to be superior to conventional statistical approaches such as logistic regression analysis for identifying patterns in data that can be used for the accurate discrimination of disease affected from unaffected individuals²¹. Machine learning is a major branch of AI²³. With machine learning, computer systems are enabled to perform analytic tasks in the absence of explicit programming to the task. Rather, programmers develop algorithms or instructions by which the computers analyze data. Machine learning also demonstrates superior capacity for group classification such as separating ‘disease’ versus ‘normal’, based on otherwise unrecognized patterns in the data. The authors have previously reported on the potential use of multiple AI platforms for disease detection using metabolomic markers²⁴. Consistent with the principles of PM, we combined sophisticated analytic techniques in the form of AI with ‘omics’ technology. Specifically, our objectives were the accurate non-invasive prediction of CCHD based on maternal salivary metabolomics. Secondly, we sought to elucidate the biochemical mechanisms in fetal CCHD development using pathway analysis.

Results

There were a total of 40 CHD cases and 40 unaffected controls. Table 1 compares the clinical and demographic parameters between CHD cases (cyanotic and acyanotic) and healthy controls. In total, 20 cyanotic and 20 non-cyanotic CHD cases were included. There were no significant differences between the groups with respect to clinical and demographic characteristics. Given that the primary objective of our study was to predict cyanotic CHD cases, we also included the comparison of demographics and clinical factors between cyanotic CHDs vs. others (non-cyanotic + healthy controls) (Supplementary Table 1).

Parameter	CHD cases (non-cyanotic + cyanotic) (n = 40)	Controls (n = 40)	p-value
Age, mean (SD)	29.3 (5.3)	32.7 (5.2)	0.547 [†]
Gestational age (days), median (IQR)	22w2d (4w2d)	21w0d (2w2d)	0.118 [†]
BMI, median (IQR)	33.0 (8.8)	32.9 (7.93)	0.293 [†]
Race, n (%)			
• Caucasian/White	21	22	0.128 [#]
• AA/Black	12	14	
• Others	7	4	
Gravidity, nulliparous (%)	26%	34%	0.064 [†]
Newborn gender, male / female	18	22	0.08 [†]
Prior child with CHD, n (%)	0	0	N/A
Family history of CHD, n (%)	3	2	0.104 [†]
Preexisting medical disorders, n (%)	4	6	0.102 [†]
Preexisting hypertension, n (%)	2	3	0.09 [†]
Preexisting diabetes, n (%)	4	7	0.06 [†]

Table 1. Clinical demographic comparison between CHD cases (cyanotic + non-cyanotic) and controls. [†]T-test. [#]Kruskal-Wallis.

The categories of CHDs found in cyanotic and non-cyanotic groups are presented in Supplementary Table 2. Cyanotic CHD included cases with TOF, HLHS, truncus arteriosus (TA), double outlet right ventricle (DORV), and transposition of great arteries. Non-cyanotic CHD cases were those with generally less critical defects including ventricular septal defects (VSD), atrial septal defects, and aortic arch abnormalities.

In total 626 metabolites were identified and after the exclusion of metabolites with missing values i.e., in >40% of samples, 468 metabolites were included in the metabolomic analysis. When cyanotic CHD cases were compared to others (non-cyanotic + unaffected controls), 30 metabolites were found to be significantly altered ($p < 0.05$) (Supplementary Table 3). Figure 1 demonstrates the top 10 altered metabolites in CCHD compared to others. OPLS-DA analysis was performed comparing the cyanotic CHD group vs. all others. OPLS-DA plot showed some overlap. The permutation test with 2000 repeats yielded an $R^2Y = 0.735$, and $Q^2 = 0.171$. This did not achieve statistical significance, likely due to the relatively small sample sizes (Supplementary Fig. 1).

Regarding the predictive models, the predictive performance was determined in a training subgroup validated in a test subgroup for overall CHD detection. First, we used a 20-marker model that initially simultaneously evaluated metabolite, clinical and demographic predictors. Clinical and demographic predictors are those that are considered in clinical practice to be risk markers for a fetus developing CHD and as a basis for detailed cardiac imaging or ECHO in the fetus. The ‘ensemble’ (summary) performance of the AI algorithms achieved an AUC (95% CI) = 0.8372 (0.635–1.00), with a sensitivity of 91.5% and specificity of 88.0% in the training set using the cross-validation (CV) approach (Supplementary Table S4a). Comparable performance was achieved in the independent test or validation group: AUC (95% CI) = 0.8222 (0.635–1.00) with a sensitivity of 90.5% and specificity of 87.0% (Supplementary Table S4b). The performances of the 6 individual AI algorithms are also displayed. A bootstrapping analytic approach was also used for AI analysis based on the 20-metabolite marker algorithm. Combining metabolite, clinical, and demographic predictors for overall CHD detection evaluated. Predictive markers were ranked using AI in decreasing order of contribution. This approach achieved an AUC (95% CI) = 0.8533 (0.635–1.00), with a sensitivity of 92.5% and specificity of 89.0% in the training set (Supplementary Table S5a). A similar predictive performance was achieved in the separate test or validation group with AUC (95% CI) = 0.8288 (0.635–1.00) and a sensitivity of 90.5% and specificity of 88.0% (Supplementary Table S5b). Overall, the performance in the training and test sets were very close indicating that there was no significant overfitting using AI prediction. Of note, commonly used clinical and demographic risk predictors for CHD did not independently contribute to CHD prediction as only the metabolomic markers persisted in the predictive algorithms.

We further evaluated the performance of a 25-marker metabolite-only model (clinical and demographic potential predictors not considered) for the overall detection of CHD. Using a bootstrapping approach, the summary (‘ensemble’) performance of the AI platforms achieved an AUC (95% CI) = 0.8488 (0.635–1.00) with a sensitivity of 92.5% and specificity of 91.0% in the separate test group (Table 2). Similarly, robust predictive performance was also achieved using the 10X CV approach: AUC (95% CI) = 0.8388 (0.635–1.00) with a sensitivity of 92.5% and specificity of 90.0% (Supplementary Table S6). In summary, AI prediction based on maternal saliva metabolomics achieved robust performance in the prediction of fetal CHD overall.

As noted, the main objective of the study was for the detection of cyanotic CHD in the fetus. We therefore assessed the performance of maternal salivary markers for the detection of CCHD from others (non-cyanotic CHD and unaffected controls combined). Using a 50-marker metabolite-only model achieved excellent diagnostic performance. The ensemble overall AI performance for the test group was AUC (95% CI) = 0.8198 (0.635–1.00) with a sensitivity and specificity of 92.5% and 87.0%, respectively using the bootstrapping approach was achieved for the detection of cyanotic CHD in the test group (Table 3). This was essentially the same as the performance in the training set in which the model was developed namely: AUC (95% CI) = 0.831 (0.635–1.00) with a sensitivity and specificity of 92.5% and 87.0%, respectively (Supplementary Table S7). Ten-fold

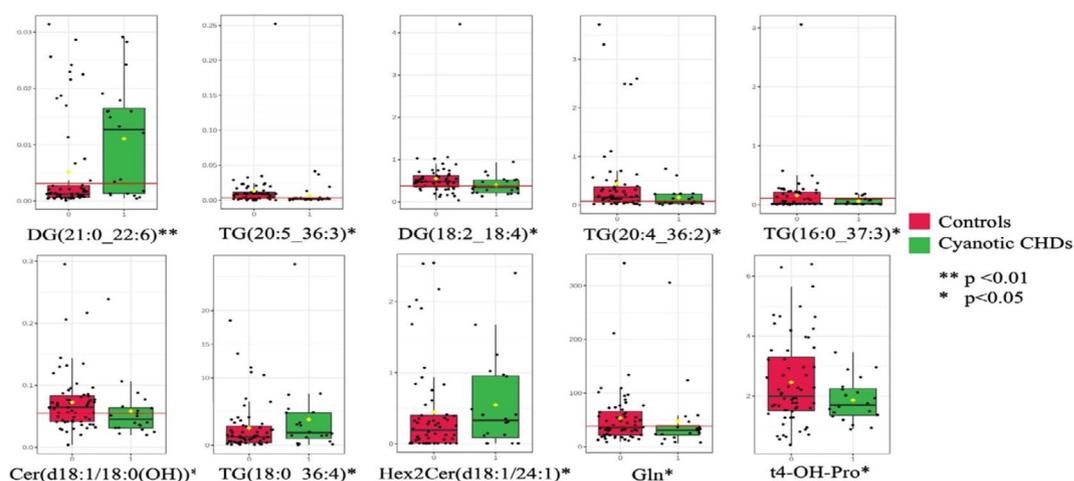


Fig. 1. Top 10 metabolites/lipids altered in cyanotic CHD when compared to others (acyanotic + unaffected controls).

	SVM	GLM	PAM	RF	LDA	DL	Ensemble
AUC	0.8311	0.8275	0.8240	0.8399	0.8122	0.8411	0.8488
95% CI	(0.6000-1)	(0.6000-1)	(0.6000-1)	(0.6000-1)	(0.6000-1)	(0.6300-1)	(0.6350-1)
Sensitivity	0.9050	0.9150	0.9000	0.9000	0.9050	0.9200	0.9250
Specificity	0.8700	0.8600	0.8750	0.8800	0.8680	0.9100	0.9100

Table 2. AI prediction of CHD overall. Maternal saliva metabolomics, clinical and demographic markers-Cross Validation approach (all CHD versus Controls) - Test Group (25 metabolite only model). **SVM:** C9, Trp, Suc, DG(21:0_22:6), C16:2-OH, PEA, Glu, Ile, Spermidine, DG(18:0_20:4), -Met-His, alpha-AAA, C16-OH, DG(14:0_14:0), C14:2, OH-GlutAcid, Cer(d18:2/23:0), TMAO, 5-AVA, C8, Spermine, Cer(d18:1/18:1), DG(17:0_17:1), Trigonelline, C0. **GLM:** GUDCA, Lac, CE(18:1), Arg, Cit, AconAcid, Cer(d18:0/24:0), Val, Asp, Orn, FA(12:0), Cer(d18:1/26:0), Sarcosine, Gln, HipAcid, 3-Met-His, alpha-AAA, C16-OH, DG(14:0_14:0), C14:2, OH-GlutAcid, Gly, C12, CE(18:3), Tyr, FA(14:0), GCDCA. **PAM:** Gly, C12, CE(18:3), Tyr, FA(14:0), GCDCA, C6 (C4:1-DC), CE(16:0), CA, Cer(d18:1/24:0), Cer(d18:1/23:0), Cer(d18:0/18:0(OH)), Creatinine, DG(18:1_20:1), Cer(d18:1/20:0), p-Cresol-SO4, DG(17:0_18:1), EPA, DG(18:2_18:2), Phe, HArg, Cer(d18:2/14:0), SDMA, DG(18:1_18:4), Met. **RF:** Spermine, Cer(d18:1/18:1), DG(17:0_17:1), Trigonelline, C0, PAG, DG(18:2_18:4), C16, GLCA, DG(14:0_20:0), C10:2, CE(20:1), Cys, CE(22:2), Cer(d18:1/18:0), Histamine, C14:2-OH, Cer(d16:1/20:0), C18:1, C7-DC, CE(22:5), DG(16:1_18:1), DG(18:2_20:0), DCA, Cer(d18:0/20:0). **LDA:** DG(18:1_20:1), Cer(d18:1/20:0), p-Cresol-SO4, DG(17:0_18:1), EPA, DG(18:2_18:2), Phe, HArg, Cer(d18:2/14:0), SDMA, DG(18:1_18:4), Cer(d18:0/22:0), C9, Trp, Suc, DG(21:0_22:6), C16:2-OH, PEA, Glu, Ile, Spermidine, Cys, CE(22:2), Cer(d18:1/18:0), Histamine, C14:2-OH. **DL:** Leu, Cer(d18:2/24:1), DG(16:0_18:2), DG(18:1_20:2), DG-O(14:0_18:2), AABA, GABA, Cer(d18:0/24:1), GLCAS, FA(18:1), Cer(d16:1/22:0), Cer(d18:1/26:0), Sarcosine, Gln, HipAcid, 3-Met-His, alpha-AAA, C16-OH, DG(14:0_14:0), C14:2, OH-GlutAcid, Cer(d18:2/23:0), TMAO, 5-AVA, C8. **Ensemble:** GUDCA, Lac, CE(18:1), Arg, Cit, AconAcid, Cer(d18:0/24:0), Val, Asp, Orn, FA(12:0), CE(18:2), Cer(d18:1/26:1), Cer(d18:1/16:0), Cer(d18:0/22:0), C9, Trp, Suc, DG(21:0_22:6), C16:2-OH, PEA, Glu, Ile, Spermidine, DG(18:0_20:4).

CV analytical approach based on 50-markers was used to predict CCHD. Potential clinical and demographic predictors were considered along with metabolite markers. In the training group used to develop the algorithm, an AUC (95% CI) = 0.826 (0.635-1.00), with a sensitivity and specificity of 91.5% and 86.0% were achieved. For the test group that model achieved a similar performance: AUC (95% CI) = 0.811 (0.635-1.00) with a sensitivity and specificity of 91.5% and 85.0%, respectively and (Supplementary Table 8).

For the sake of comparison, we further evaluated the performance of conventional logistic regression for the prediction of cyanotic CHD. Supplementary Table 9 presents the top 2 metabolite models in the prediction of cyanotic CHD. With 10-fold cross-validation (CV) analysis achieved an AUC (95%CI) = 0.743 (0.617-0.870) with a 90.0% sensitivity and 72.0% specificity for the detection of cyanotic CHD. Not surprisingly, AI appeared to improve the predictive performance over conventional logistic regression approaches.

Metabolite set enrichment analysis was performed to evaluate the underlying changes in the maternal metabolome with cyanotic CHD. Arachidonic acid, Alpha-linoleic acid, and Tryptophan metabolism were the top 3 pathways that were significantly dysregulated ($p < 0.05$ or $-\log_{10}(p) > 1.301$) in cyanotic CHD (Fig. 2).

Discussion

We report for the first time the use of maternal salivary metabolomics combined with machine learning, a major branch of AI, the accurate detection of fetal CCHD. In addition, using metabolomic pathway analysis, we identified significant dysregulation of lipid metabolism with the development of CCHD. These are discussed in more detail below.

Multiple reviews on Precision Cardiology have been comprehensively summarized by Sethi et al.²⁵. These have emphasized the promise of Precision Cardiology powered by “omics” and AI, for improved understanding of disease pathogenesis that will ultimately guide the development of novel therapies. AI, specifically neural networks, and machine learning have significantly improved the diagnostic accuracy of such tools as echocardiograms, MRI, computed tomography, and electrocardiography which are used in pediatric cardiology²⁶. These tools, with the exception of prenatal echocardiography, are not available for fetal cardiac evaluation². The field of Precision Fetal Cardiology is in its infancy but has the potential to deliver similar dividends to fetal and newborn care⁸.

Using DL, a powerful upgrade of machine learning which is based on artificial neural networks²⁷, and other AI platforms in this study, we achieved non-invasive prediction of CHD. We used multiple different AI platforms. For example, an AUC (95% CI) = 0.8198 (0.6350-1) with a sensitivity of 92.5% and specificity of 87.0% was achieved for the detection of CCHD using maternal salivary metabolites (Table 3). Similarly, high predictive accuracies were also achieved for the overall detection of CHD: AUC (95% CI) = 0.8488 (0.635-1.00) with a sensitivity of 92.5% and specificity of 91.0% (Table 2). Overall, therefore using a Precision Medicine approach in fetal cardiology the combination of maternal saliva metabolomics and AI achieved robust diagnostic performance for the detection of fetal CCHD and CHD overall.

An important objective of Precision Medicine is to understand the pathogenesis of complex disorders. We found evidence of extensive dysregulation of lipid metabolism in the cyanotic CHD group. A total of 23 of the 30 metabolites (76.7%) that were significantly altered were lipids (Supplementary Table 3). These included

	SVM	GLM	PAM	RF	LDA	DL	Ensemble
AUC	0.8021	0.7985	0.7950	0.8110	0.7832	0.8121	0.8198
95% CI	(0.6000-1)	(0.6000-1)	(0.6000-1)	(0.6000-1)	(0.6000-1)	(0.6300-1)	(0.6350-1)
Sensitivity	0.9050	0.8950	0.9000	0.9000	0.9150	0.9200	0.9250
Specificity	0.8300	0.8300	0.8450	0.8440	0.8580	0.8700	0.8700

Table 3. AI prediction of cyanotic CHD. Maternal saliva metabolomics, clinical and demographic markers- bootstrapping approach (cyanotic CHD versus non-cyanotic + unaffected controls) - Test Group (50 metabolite only model). **Important predictors in decreasing order of contribution:** SVM: Arg, Race, Cer(d16:1/22:0), CE(14:1), DG(18:3_20:2), Cer(d18:0/16:0), Cer(d18:1/18:0), Val, Cer(d18:1/22:0), GCDCA, DG(18:1_22:6), Cer(d18:1/24:0), Cer(d18:2/16:0), DG(18:1_20:2), DG(14:0_18:1), HArg, CA, Cer(d16:1/23:0), DG(18:1_18:4), DG(18:2_18:2), C2, Cit, SDMA, Asp, TDCA, Spermine, CE(18:3), C14:2, Taurine, Phe, Cer(d18:1/26:0), Betaine, CE(22:1), C14:2-OH, Cer(d18:1/18:1), C3, Glu, C10:2, Spermidine, C18:2, ADMA, Cer(d18:2/22:0), Cer(d18:0/18:0), GLCA, C5-DC (C6-OH), DG(18:2_20:4), OH-GlutAcid, C16, Cer(d18:2/18:1). **GLM:** C10:2, Spermidine, DG(16:0_18:1), C14:1-OH, Trp, DG(18:0_20:4), FA(16:0), Cer(d18:0/24:0), Cer(d16:1/24:0), C18:2, CE(14:0), ADMA, AA, Cer(d18:2/22:0), Histamine, Cer(d18:0/18:0), GLCA, C5-DC (C6-OH), DG(18:2_20:4), OH-GlutAcid, AconAcid, C16, Cer(d18:1/26:1), Cer(d18:2/18:1), GABA, Lac, CE(20:3), TMCA, PEA, Anserine, C16:2, Ac-Orn, DG(18:1_20:1), Cys, HipAcid, DG(21:0_22:6), C7-DC, Trigonelline, GUDCA, DG-O(18:2_18:2), FA(12:0), CE(18:1), C5-OH (C3-DC-M), 3-Met-His, FA(14:0), Pro, Cer(d18:0/22:0), C9, Creatinine, AABA. **PAM:** C10:2, Spermidine, C18:2, ADMA, Cer(d18:2/22:0), Cer(d18:0/18:0), GLCA, C5-DC (C6-OH), DG(18:2_20:4), OH-GlutAcid, AconAcid, C16, Cer(d18:1/26:1), Cer(d18:2/18:1), GABA, Lac, CE(20:3), TMCA, PEA, Anserine, C16:2, Ac-Orn, DG(18:1_20:1), Cys, HipAcid, DG(21:0_22:6), C7-DC, Trigonelline, GUDCA, DG-O(18:2_18:2), FA(12:0), CE(18:1), C5-OH (C3-DC-M), 3-Met-His, FA(14:0), Pro, Cer(d18:0/22:0), C9, Creatinine, AABA, DG(17:0_17:1), Cer(d18:2/20:0), DG(14:0_20:0), CE(20:1), Cer(d16:1/20:0), Suc, CE(17:1), DG(18:2_18:4), C5-M-DC, CE(20:0). **RF:** AconAcid, Cer(d18:1/26:1), GABA, Lac, CE(20:3), TMCA, PEA, Anserine, C16:2, Ac-Orn, DG(18:1_20:1), Cys, HipAcid, DG(21:0_22:6), C7-DC, Trigonelline, GUDCA, DG-O(18:2_18:2), FA(12:0), CE(18:1), C5-OH (C3-DC-M), 3-Met-His, FA(14:0), Pro, Cer(d18:0/22:0), C9, Creatinine, AABA, DG(17:0_17:1), Cer(d18:2/20:0), DG(14:0_20:0), CE(20:1), Cer(d16:1/20:0), Suc, CE(17:1), DG(18:2_18:4), C5-M-DC, CE(20:0), Cer(d18:1/23:0), ProBetaine, Asn, Cer(d18:1/20:0), Cer(d18:1/16:0), DG-O(16:0_20:4), Ser, DG(22:1_22:2), DG(17:0_18:1), DG(18:1_18:1), Gln, PAG, Cer(d18:2/14:0). **LDA:** DG(22:1_22:2), DG(17:0_18:1), DG(18:1_18:1), Gln, PAG, AconAcid, Cer(d18:1/26:1), GABA, Lac, CE(20:3), TMCA, PEA, Anserine, C16:2, Ac-Orn, DG(18:1_20:1), Cys, HipAcid, DG(21:0_22:6), C7-DC, Trigonelline, GUDCA, DG-O(18:2_18:2), FA(12:0), CE(18:1), C5-OH (C3-DC-M), 3-Met-His, FA(14:0), Pro, Cer(d18:0/22:0), C9, Creatinine, AABA, DG(17:0_17:1), Cer(d18:2/20:0), DG(14:0_20:0), CE(20:1), Cer(d16:1/20:0), Suc, CE(17:1), DG(18:2_18:4), C5-M-DC, CE(20:0), Cer(d18:1/23:0), ProBetaine, Asn, Cer(d18:1/20:0), Cer(d18:1/16:0), DG-O(16:0_20:4). **DL:** Ser, DG(22:1_22:2), DG(17:0_18:1), DG(18:1_18:1), Gln, PAG, Cer(d18:2/14:0), Arg, Race, Cer(d16:1/22:0), AconAcid, CE(14:1), Cer(d18:1/26:1), DG(18:3_20:2), GABA, Cer(d18:0/16:0), Lac, Cer(d18:1/18:0), CE(20:3), Val, TMCA, Cer(d18:1/22:0), PEA, Anserine, C16:2, Ac-Orn, DG(18:1_20:1), Cys, HipAcid, DG(21:0_22:6), C7-DC, Trigonelline, GUDCA, DG-O(18:2_18:2), FA(12:0), CE(18:1), C5-OH (C3-DC-M), 3-Met-His, FA(14:0), Pro, Cer(d18:0/22:0), C9, Creatinine, AABA, DG(17:0_17:1), Cer(d18:2/20:0), DG(14:0_20:0), CE(20:1), DG(18:2_18:4), C5-M-DC, CE(20:0), Cer(d18:1/23:0), ProBetaine, Asn, Cer(d18:1/20:0), Cer(d18:1/16:0), DG-O(16:0_20:4), Ser, DG(22:1_22:2), DG(17:0_18:1), DG(18:1_18:1), Gln, PAG.

glycerolipids such as diglycerides (DG) and triglycerides (TG), and ceramides. The DG molecule consists of a glyceride moiety covalently bound to two fatty acid chains through ester linkages. TGs consist of the glycerol molecule similarly linked to three fatty acid groups. The many functions of TGs are the storage and supply of energy. Ceramides also known as N-acylsphingosines, consist of a sphingoid base linked to a fatty acid chain via the amine group. DG (21:0_22:6) is diglyceride or diacylglycerol which consists of two fatty acid chains covalently bound to a glycerol molecule. This was the most significantly altered metabolite in our study. Similarly, there were two other diglycerides, DG (18:2_18:4) and DG (18:2_18:2) whose concentrations were statistically significantly altered in CCHD, and in both cases, their concentrations were reduced. Diacylglycerols are precursors to TGs. Overall, these changes highlight the fact that the underlying maternal metabolome alteration in cyanotic CHDs is related to lipid pathways, mainly fatty acids. We have previously reported significant alterations in the lipid profile in maternal urine in the fetal CHD group¹¹. A recent review that evaluated maternal metabolomic profiling during pregnancy in those with a fetal CHD demonstrated findings consistent with significant lipid changes in serum, and urine. These included alterations in levels of acylcarnitines, phospholipids, fatty acids,

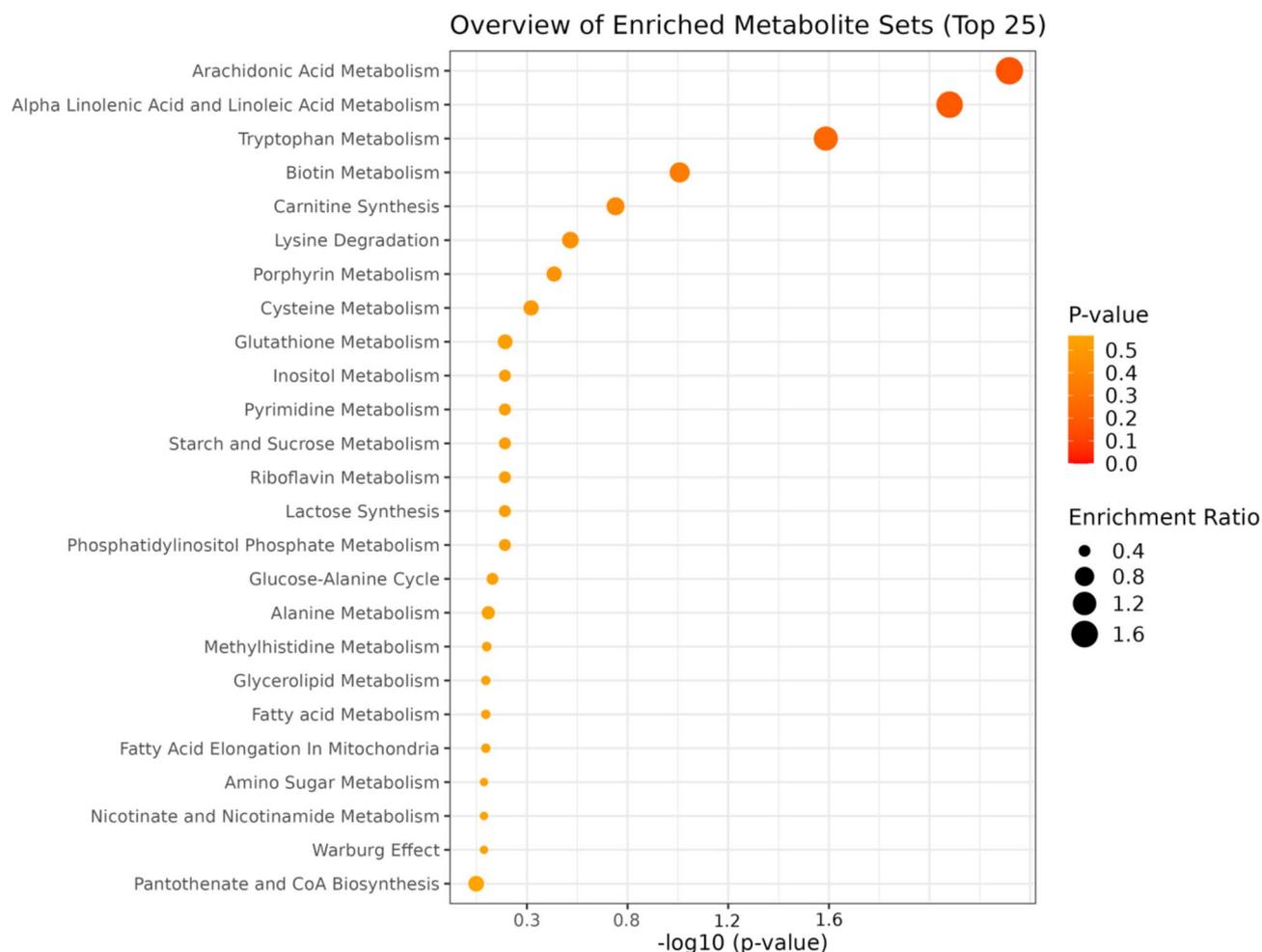


Fig. 2. Pathway analysis for Cyanotic CHDs vs. others (acyanotic CHD + healthy controls).

lysophospholipids and sphingolipids²⁸. Additionally, a study evaluating maternal serum metabolomics in cyanotic CHD subtype (HLHS and TOF) showed lipid dysregulation with an increase in ceramides in HLHS and a decrease in TG subclasses in TOF²⁹. In contrast, in our study, some of TG subtypes were significantly increased in CCHD, e.g. TG (14:0 _ 36:3), TG (18:0 _ 36:5), and others. These lipids and in particular TG perturbation appears consistent with a large case-control study which found elevated first-trimester maternal blood lipid levels in fetal CHD versus controls³⁰.

There are a few pediatric studies that have evaluated the metabolome of cardiac tissue in live individuals. A pediatric study using left atrial biopsy tissue obtained after placement on cardiopulmonary bypass compared CCHD and noncyanotic CHD. The concentration of approximately 25% of measured lipids increased while approximate levels of 10% of measured levels decreased¹⁶. On average CCHD cases were 14 months old. The increase in fat-related metabolites was interpreted to indicate a downregulation of mitochondrial beta-oxidation of fatty acids. Our findings of reduced triglycerol and diglyceride levels may suggest a shift towards fatty acid oxidation rather than glycolysis in the fetus. In another study comparing the tissue obtained from 10 TOF cases and 10 atrial septal defect cases, altered Butanoate and Purine metabolisms was shown to be associated with TOF³¹. Butanoate is a short-chain fatty acid and dysfunction in butanoate metabolism has been linked to postnatal cardiovascular conditions such as infarction, stroke, and coronary revascularization³². Overall, these pediatric studies of actual cardiac tissue support our findings of dysregulation of lipid metabolism in CCHD.

Consistent with the principles of Precision Cardiology, we also interrogated the underlying pathogenesis of CCHD cases using Metabolite Set Enrichment Analysis (MSEA). Arachidonic acid (AA), Alpha-linolenic acid (LA), and Tryptophan metabolism were significantly dysregulated in pregnant patients with fetal CCHD. Arachidonic acid and LA pathways are crucial in fatty acid metabolism and are essential for fetal development. Dysregulation of these pathways is consistent with our previous work utilizing maternal urine metabolomics for CHD prediction where alterations in fatty acid metabolism were identified¹¹. The derivatives of AA and LA metabolism, such as hydroxyeicosatetraenoic acids (HETE) and hydroxyoctadecadienoic acids (HODE), are proinflammatory mediators and play important role in organogenesis, and fetal growth³³. AA is a long-chain polyunsaturated fatty acid that is acquired by the fetus through placental transfer from the mother³⁴. AA is also an integral constituent of all cells including cardiac myocytes and regulates the expression of vascular endothelial

growth factor (VEGF), which itself plays a critical role in cardiac development³⁵. Normal levels of VEGF are required for normal cardiovascular development and alterations in VEGF expression have been reported to be associated with CHDs including cyanotic TOF cases^{36,37}.

Alpha-linoleic acid is another essential fatty acid that structures the intracellular ligands for gene expression and functions as signaling molecules and maintains cellular metabolic homeostasis³⁸. In a study comparing pediatric cyanotic vs. non-cyanotic CHD, a significant decrease in low-density lipoprotein (LDL) levels was shown in the cyanotic group. Linoleic acid is known to lower LDL levels and regulate the overall lipid profile³⁹. Alpha-linolenic acid intake has also been shown to correlate with higher birth weight⁴⁰, which may explain the observed higher risk of CHD in patients with pregestational diabetes and macrosomia i.e. large fetal weight. Additionally, in the small pilot study of newborn blood spots from CHD newborns⁴¹, dysregulation of Arachidonic acid and Linoleic acid metabolism was identified in TOF cases, an important category of CCHD. Further, tryptophan metabolism was significantly dysregulated in the fetal CCHD group in our study. Tryptophan is the precursor of Kynurenine pathway metabolites that promotes neonatal heart regeneration by stimulating cardiomyocyte proliferation and cardiac angiogenesis⁴². The Kynurenine pathway was shown to be perturbed in newborns who were diagnosed with transposition of great arteries and underwent surgical repair⁴³. These reports provide insights into the potential role of altered maternal tryptophan metabolism and its link to cyanotic CHD.

A limitation of our study was the relatively modest sample size. Consequently, we employed strategies detailed in the AI analysis section, to minimize overfitting. Also, using bioinformatic techniques we were able to determine the predictive accuracy in a training and separate test group. The diagnostic performances were similar between the two groups indicating that there was no significant overfitting with AI analysis. While dietary intake can confound or account for metabolites found in the saliva, this also applies to maternal blood. We were unable to correlate metabolites in the maternal saliva with their levels in the fetal blood. For practical and ethical reasons this was not possible.

Our study has multiple strengths. One such is its novelty. Another is its potential future clinical value to overcome the practical limitations of prenatal ultrasound detection⁴⁴ and also newborn pulse-oximetry screening for CHD detection⁴⁵ both of which have limitations in detecting CHD and critical CHD respectively. Given the robust diagnostic performance achieved, salivary metabolomics could potentially help to surmount the limitations imposed by the lack of access to skilled fetal cardiac diagnosticians among significant portions of the US population as well as globally. Our study focuses on the most critical group of CHD cases, those with cyanotic defects. However, there was good overall predictive performance in isolated CHD, the biggest subgroup of CHD. Missed prenatal diagnosis of CCHD and failure to deliver in tertiary centers are most likely to increase the risk of catastrophic outcomes. Finally, saliva samples on the other hand are easy to obtain, and completely without pain or discomfort and is available in abundance. These features, further enhance the attraction of this approach as a potential future screening tool that could be deployed in the general pregnant population.

In summary, we report for the first time the use of maternal salivary metabolomics for the accurate detection of fetal CCHD, the subgroup of CHD with the highest rates of pediatric morbidity and mortality. Similarly excellent diagnostic performance was achieved for the non-invasive detection of CHD overall. Based on the principles of Precision Fetal Cardiology, the pathogenesis of CCHD was interrogated. Our findings indicate that dysregulation of lipid metabolism is associated with the development of cyanotic CHD. Validation of these findings in a larger patient group is desirable.

Methods

Study population

This was a prospective case-control cohort study that was conducted at Corewell Health William Beaumont University Hospital, Royal Oak, MI. IRB approval (#2017–145) was obtained. Study participants provided written informed consent. All procedures performed in this study were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Patients with fetuses suspected or diagnosed prenatally with CHD were included in the study. Gestational age matched controls were recruited at the time of their prenatal anatomic survey. Inclusion criteria were (i) singleton gestation, (ii) maternal age between 18 and 50 years of age (iii) gestational age at recruitment between 14w 0d–37w 0d gestation, (iv) isolated, non-syndromic CHD that were subsequently confirmed postnatally. Exclusion criteria included pregnancies that are complicated with aneuploidy, syndromic (e.g., Di-George syndrome), extracardiac defects, and multifetal deletions. All the CHD cases underwent fetal echocardiograms prior to recruitment and controls underwent standard ultrasound screening of the fetal heart. Fetal echocardiograms and standard heart screening exams were performed in a single Fetal Imaging unit location by an experienced sonographer and a maternal fetal medicine specialist. The gold standard for diagnosis of CHD was based on a newborn or neonatal ECHO reviewed by a pediatric cardiologist.

Data collection

Maternal clinical and demographic data such as age, race, BMI, gravidity and parity, gestational age at sample collection, pregnancy complications, medical history including diabetes and obstetric history including a prior child/ pregnancy with CHD, social history including smoking, alcohol consumption, family history and medication use in pregnancy were recorded. The results of the fetal and neonatal echocardiogram were recorded including the type of cardiac malformation. The suspected diagnosis of CHD was confirmed via neonatal echocardiogram by a pediatric cardiologist prior to the allocation of cases and controls.

Sample collection

Maternal saliva samples were collected following approximately 3–4 h of fasting. Each participant provided a sample of saliva by spitting directly into a falcon tube after rinsing their mouth with distilled water 5 min prior to the collection. Protease inhibitor in a 1% solution was added to the saliva sample to inhibit salivary enzymes and the generation of new metabolites during sample preparation and storage. Samples were vortexed for 1 min at 4 °C and were subsequently stored in cryovials at -80 °C until analysis.

Metabolomic analysis using UPLC-MS/MS (ultra high-performance liquid chromatography-mass spectrometry)

Saliva extracts were analyzed using a Waters I-class ultra-performance liquid chromatography unit coupled with a Waters Xevo-TQ-S (Waters Corporation, Milford, MA, USA). For UPLC analysis, saliva sample extracts were separated using the MxP Quant 500 C18 column with attached guard and precolumn mixer (Biocrates Life Sciences, AG, Innsbruck, Austria). LC-MS grade Acetonitrile, Methanol, Isopropyl alcohol, Formic acid ($\geq 99.0\%$ purity) were obtained from Fisher Scientific (Hanover Park, IL, USA). LC-MS grade Ethanol, Pyridine, and Phenylisothiocyanate were purchased from Sigma Aldrich (St. Louis, MO, USA). Milli-Q Water was used for the aqua mobile phase (EMD Millipore, Billerica, MA, USA). Saliva samples and calibration standards were thawed on ice. Saliva samples were subsequently mixed for 10s and centrifuged at 10,000 g at 4 °C for 10 min. The saliva samples were prepared as directed in the manufacturer's instructions (Biocrates Life Sciences, AG, Innsbruck, Austria). An overview of the sample preparation process is in the supplementary section.

Exploratory data analysis

Prior to analysis, metabolites that are below the limits of quantification ($> 40\%$) were excluded. These missing values were replaced with a value that is half of the minimum positive values obtained in the original sample set. We made this assumption because most missing values are caused by low-abundance metabolites. Prior to performing Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA), all data were normalized to the median and auto-scaled⁴⁶. Models were cross validated using permutation testing (2000 iterations) to determine if the observed separation in the representative scores' plots achieved statistical significance ($p < 0.05$). Cyanotic CHD vs. others (non-cyanotic CHD and unaffected controls) and CHD cases overall versus unaffected controls were separately compared.

Artificial intelligence analysis

We have extensively described the methodologies involved in the use of AI for disease detection in multiple prior publications for the analysis of both metabolomic²⁴ and molecular data⁴⁷. In this study, comprehensive AI analyses were performed to identify the best predictive metabolite markers for distinguishing CCHDs from others (unaffected controls + non-cyanotic CHDs) and CHDs overall from unaffected controls. The combination of controls and non-cyanotic CHDs were done to reflect the real-world clinical setting. A total of seven AI algorithms or platforms: Deep Learning (DL), Support Vector Machine (SVM), Generalized Linear Model (GLM), Prediction Analysis for Microarrays (PAM), Random Forest (RF), and Linear Discriminant Analysis (LDA) and summary analysis of the combined AI platforms i.e. Ensemble Learning (EL) were used for the prediction of CHD⁴⁸. A description of each of the AI platforms was previously provided in our other publications⁸, is again briefly presented in the Supplementary Methods section. Details of the methods of training and validating the data for AI analyses are also provided in the Supplementary Methods section.

Clinical and demographic predictors

The following potential clinical and demographic predictors of CHD were considered in the predictive algorithms along with the metabolite markers: maternal age, race, body mass index, prior child with CHD, family history of CHD, preexisting diabetes, in vitro fertilization, first trimester alcohol use and tobacco use. These are generally considered risk factors for fetal development of CHD and are widely used to identify pregnancies for detailed fetal echocardiograms. Other factors such as race and age while not directly linked to increased risk of CHD can affect the human metabolome and as such were regarded as potential confounders in our predictive models.

Software Tools

The variable importance functions *varimp* in h2o and *varImp* in caret R packages were utilized to rank the model features in each of the AI predictive algorithms. We used pROC R package to compute the area under the receiver-operating characteristic (ROC) curve and 95% CI, specificity, and sensitivity to assess the overall performance of the model⁴⁹.

Modeling & evaluation

Several parameters were used to tune the AI models while implementing them: Number of trees for RF, classification cost for SVM, the threshold amount for shrinking toward the centroid for PAM, and for DL model: (a) Epochs (number of passes of the full training set), (b) l1 (penalty to converge the weights of the model to 0), (c) l2 (penalty to prevent the enlargement of the weights), (d) input dropout ratio (ratio of ignored neurons in the input layer during training), (e) number of hidden layers. In addition to, l1 and l2 parameters, *input_dropout_ratio* was used as the third parameter to avoid overfitting in the DL model. This controls the amount of input layer neurons that are randomly dropped (set to zero) and controls overfitting with respect to the input data (useful for high-dimensional noisy data). The objective was to randomly drop units (along with their connections) from the neural network during training⁵⁰. This prevents excessive co-adapting of units. Using these three strategies helped to avoid the major complication of DL model generation namely overfitting⁵⁰. *Overfitting* refers to the

situation where the AI model fits too closely (accurately) to the original data and performs significantly less effectively when presented with new data.

Dividing a dataset into training, validation, and test sets is crucial in machine learning and deep learning for effectively developing and validating models. The dataset was divided into two parts: a training/validation set and a test set. We used 75% of the data for training and validation, and the remaining 25% for testing. The next step was to split the initial set (75% of the entire dataset) into training and validation sets, 75% for training and 25% for validation. We used 10-fold cross-validation on the training/validation set to ensure robust validation. This helped in utilizing the available data more effectively. We kept the remaining 25% of the data as a separate test set. This data was not used during the model development or parameter tuning phases to ensure unbiased evaluation. We randomly shuffled the data before splitting to ensure that the training, validation, and test sets are representative of the overall dataset. To divide the data into training and test groups, Python's *train_test_split* method from the *scikit-learn* library was utilized.

Artificial intelligence: ranking important features and minimizing overfitting

The contribution of a feature (i.e., a predictor variable) to overall model performance was determined using a model-based approach. We ranked the importance of the features in each of the predictive algorithms by using the variable importance functions *varimp* in h2o and *varImp* in caret R packages. As noted, one of the risks of AI analysis, particularly with a relatively small data sets, is overfitting as defined previously. To minimize the risk of overfitting we employed several strategies. For the DL model, we applied L1 and L2 regularization parameters, causing some weights to become 0 and preventing weight enlargement. We further utilized the 'input dropout ratio' to control overfitting for high-dimensional noisy data. For predictive models based on the five other AI platforms, we tuned different parameters, such as the number of trees for RF, the threshold amount for PAM, and the classification cost for SVM, to further overcome the challenge of overfitting.

Logistic regression analysis

Studies have demonstrated the superiority of AI platforms such as Deep Learning compared to more widely employed prediction approaches such as Logistic Regression⁵¹. For the sake of comparison, we also evaluated model performance based on logistic regression. Logistic regression analysis was performed using a stepwise variable selection method to optimize all the model components for the sake of comparison. A k-fold cross-validation (CV) technique ensures the validity and generalizability of our logistic regression model by randomly dividing the entire sample data into k equal-sized subsets, of which only one is used as the validation data for the model, and the remaining subsets are used as training sets. Optimal and robust predictive algorithms were generated⁵². Model performance was also reported in the form of AUC (95% CI), sensitivity and specificity values.

Metabolite set enrichment analysis

Consistent with the primary objective of the study, we investigated the pathogenesis of CCHD. Biologically meaningful patterns in fully quantified metabolite concentrations comparing cyanotic CHD versus others i.e. combined acyanotic CHDs and controls were identified using the Pathway Analysis tool in MetaboAnalyst (v5.0)⁵³. The *Homo sapiens* (human) pathway library was utilized and all the compounds in the selected pathways were used when referencing the specific metabolome. Pathway analysis determines whether a group of functionally related metabolites in each biochemical pathway are significantly perturbed in a disorder. This eliminates the need to preselect compounds and determine significance based on arbitrary thresholds⁴⁶.

Data availability

The datasets generated during the current study are available from the corresponding author on reasonable request.

Received: 20 January 2024; Accepted: 1 January 2025

Published online: 15 January 2025

References

- Hoffman, J. I. & Christianson, R. Congenital heart disease in a cohort of 19,502 births with long-term follow-up. *Am. J. Cardiol.* **42**, 641–647. [https://doi.org/10.1016/0002-9149\(78\)90635-5](https://doi.org/10.1016/0002-9149(78)90635-5) (1978).
- Li, Y. F. et al. Efficacy of prenatal diagnosis of major congenital heart disease on perinatal management and perioperative mortality: a meta-analysis. *World J. Pediatr.* **12**, 298–307. <https://doi.org/10.1007/s12519-016-0016-z> (2016).
- Pinto, N. M. et al. Barriers to prenatal detection of congenital heart disease: a population-based study. *Ultrasound Obstet. Gynecol.* **40**, 418–425. <https://doi.org/10.1002/uog.10116> (2012).
- Desai, K., Rabinowitz, E. J. & Epstein, S. Physiologic diagnosis of congenital heart disease in cyanotic neonates. *Curr. Opin. Pediatr.* **31**, 274–283. <https://doi.org/10.1097/MOP.0000000000000742> (2019).
- Hobbes, B. et al. Determinants of adverse outcomes after Systemic-To-Pulmonary shunts in biventricular circulation. *Ann. Thorac. Surg.* **104**, 1365–1370. <https://doi.org/10.1016/j.athoracsur.2017.06.043> (2017).
- Singh, Y. & Mikrou, P. Use of prostaglandins in duct-dependent congenital heart conditions. *Arch. Dis. Child. Educ. Pract. Ed.* **103**, 137–140. <https://doi.org/10.1136/archdischild-2017-313654> (2018).
- Williams, K., Carson, J. & Lo, C. Genetics of congenital heart disease. *Biomolecules* **9** <https://doi.org/10.3390/biom9120879> (2019).
- Bahado-Singh, R. et al. Cell-free DNA in maternal blood and artificial intelligence: accurate prenatal detection of fetal congenital heart defects. *Am J Obstet Gynecol* **228**, 76 e71-76 e10 (2023). <https://doi.org/10.1016/j.ajog.2022.07.062>
- Bahado-Singh, R. O. et al. Epigenetic markers for newborn congenital heart defect (CHD). *J. Matern Fetal Neonatal Med.* **29**, 1881–1887. <https://doi.org/10.3109/14767058.2015.1069811> (2016).
- Radhakrishna, U. et al. Placental epigenetics for evaluation of fetal congenital heart defects: ventricular septal defect (VSD). *PLoS One.* **14**, e0200229. <https://doi.org/10.1371/journal.pone.0200229> (2019).

11. Friedman, P. et al. Urine metabolomic biomarkers for prediction of isolated fetal congenital heart defect. *J. Matern Fetal Neonatal Med.* **35**, 6380–6387. <https://doi.org/10.1080/14767058.2021.1914572> (2022).
12. Li, Y. et al. Analysis of biomarkers for congenital heart Disease based on maternal amniotic fluid metabolomics. *Front. Cardiovasc. Med.* **8**, 671191. <https://doi.org/10.3389/fcvm.2021.671191> (2021).
13. Troisi, J. et al. Noninvasive screening for congenital heart defects using a serum metabolomics approach. *Prenat Diagn.* **41**, 743–753. <https://doi.org/10.1002/pd.5893> (2021).
14. Yuan, X. et al. Biomarkers for isolated congenital heart disease based on maternal amniotic fluid metabolomics analysis. *BMC Cardiovasc. Disord.* **22**, 495. <https://doi.org/10.1186/s12872-022-02912-2> (2022).
15. Guvenc, O. et al. Early postnatal metabolic profile in neonates with critical CHDs. *Cardiol. Young.* **33**, 349–353. <https://doi.org/10.1017/S1047951122003134> (2023).
16. Dong, S. et al. Metabolic profile of heart tissue in cyanotic congenital heart disease. *Am. J. Transl Res.* **13**, 4224–4232 (2021).
17. Ferrari, E. et al. Human serum and salivary metabolomes: diversity and closeness. *Int. J. Mol. Sci.* **24** <https://doi.org/10.3390/ijms242316603> (2023).
18. Hyvarinen, E., Savolainen, M., Mikkonen, J. J. W. & Kullaa, A. M. Salivary metabolomics for diagnosis and monitoring diseases: challenges and possibilities. *Metabolites* **11** <https://doi.org/10.3390/metabo11090587> (2021).
19. Farah, R. et al. Salivary biomarkers for the diagnosis and monitoring of neurological diseases. *Biomed. J.* **41**, 63–87. <https://doi.org/10.1016/j.bj.2018.03.004> (2018).
20. Li, Y. et al. Untargeted metabolomics of saliva in pregnant women with and without gestational diabetes mellitus and healthy non-pregnant women. *Front. Cell. Infect. Microbiol.* **13**, 1206462. <https://doi.org/10.3389/fcimb.2023.1206462> (2023).
21. Chen, R. & Snyder, M. Promise of personalized omics to precision medicine. *Wiley Interdiscip Rev. Syst. Biol. Med.* **5**, 73–82. <https://doi.org/10.1002/wsbm.1198> (2013).
22. McColl, E. R., Asthana, R., Paine, M. F. & Piquette-Miller, M. The age of Omics-Driven Precision Medicine. *Clin. Pharmacol. Ther.* **106**, 477–481. <https://doi.org/10.1002/cpt.1532> (2019).
23. Ali, S. & Byrne, M. F. in *AI Clin. Med.* **1–12** (2023).
24. Bahado-Singh, R. O. et al. Artificial intelligence and amniotic fluid multiomics: prediction of perinatal outcome in asymptomatic women with short cervix. *Ultrasound Obstet. Gynecol.* **54**, 110–118. <https://doi.org/10.1002/uog.20168> (2019).
25. Sethi, Y. et al. Precision Medicine and the future of Cardiovascular diseases: a clinically oriented Comprehensive Review. *J. Clin. Med.* **12** <https://doi.org/10.3390/jcm12051799> (2023).
26. Sethi, Y. et al. Artificial Intelligence in Pediatric Cardiology: a scoping review. *J. Clin. Med.* **11** <https://doi.org/10.3390/jcm11230702> (2022).
27. Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C. & Collins, J. J. Next-generation machine learning for Biological Networks. *Cell* **173**, 1581–1592. <https://doi.org/10.1016/j.cell.2018.05.015> (2018).
28. Mires, S., Reddy, S., Skerritt, C., Caputo, M. & Eastwood, K. A. Maternal metabolomic profiling and congenital heart disease risk in offspring: a systematic review of observational studies. *Prenat Diagn.* **43**, 647–660. <https://doi.org/10.1002/pd.6301> (2023).
29. Hsu, P. C., Maity, S., Patel, J., Lupo, P. J. & Nembhard, W. N. Metabolomics Signatures and Subsequent Maternal Health among Mothers with a Congenital Heart Defect-Affected Pregnancy. *Metabolites* **12** (2022). <https://doi.org/10.3390/metabo12020100>
30. Cao, L. et al. High maternal blood lipid levels during early pregnancy are associated with increased risk of congenital heart disease in offspring. *Acta Obstet. Gynecol. Scand.* **100**, 1806–1813. <https://doi.org/10.1111/aogs.14225> (2021).
31. Liu, J. et al. Metabolic variation dictates cardiac pathogenesis in patients with tetralogy of Fallot. *Front. Pediatr.* **9**, 819195. <https://doi.org/10.3389/fped.2021.819195> (2021).
32. Pradhan, A. D. et al. Rationale and design of the Pemafibrate to Reduce Cardiovascular outcomes by reducing triglycerides in patients with diabetes (PROMINENT) study. *Am. Heart J.* **206**, 80–93. <https://doi.org/10.1016/j.ahj.2018.09.011> (2018).
33. Szczuko, M. et al. The role of arachidonic and linoleic acid derivatives in pathological pregnancies and the Human Reproduction process. *Int. J. Mol. Sci.* **21** <https://doi.org/10.3390/ijms21249628> (2020).
34. Crawford, M. & 275S-284S. Placental delivery of arachidonic and docosahexaenoic acids: implications for the lipid nutrition of preterm infants. *Am. J. Clin. Nutr.* **71** <https://doi.org/10.1093/ajcn/71.1.275S> (2000).
35. Yang, S., Wei, S., Pozzi, A. & Capdevila, J. H. The arachidonic acid epoxygenase is a component of the signaling mechanisms responsible for VEGF-stimulated angiogenesis. *Arch. Biochem. Biophys.* **489**, 82–91. <https://doi.org/10.1016/j.abb.2009.05.006> (2009).
36. Dor, Y. et al. A novel role for VEGF in endocardial cushion formation and its potential contribution to congenital heart defects. *Development* **128**, 1531–1538. <https://doi.org/10.1242/dev.128.9.1531> (2001).
37. van den Akker, N. M. et al. Tetralogy of fallot and alterations in vascular endothelial growth factor-A signaling and notch signaling in mouse embryos solely expressing the VEGF120 isoform. *Circ. Res.* **100**, 842–849. <https://doi.org/10.1161/01.RES.0000261656.04773.39> (2007).
38. Duttaroy, A. K. & Basak, S. Maternal fatty acid metabolism in pregnancy and its consequences in the fetoplacental development. *Front. Physiol.* **12**, 787848. <https://doi.org/10.3389/fphys.2021.787848> (2021).
39. Froyen, E. & Burns-Whitmore, B. The effects of Linoleic Acid Consumption on lipid risk markers for Cardiovascular Disease in healthy individuals: a review of human intervention trials. *Nutrients* **12** <https://doi.org/10.3390/nu12082329> (2020).
40. Phang, M. et al. Increased alpha-linolenic acid intake during pregnancy is Associated with higher offspring Birth Weight. *Curr. Dev. Nutr.* **3**, nzy081. <https://doi.org/10.1093/cdn/nzy081> (2019).
41. Ceresnac, S. R. et al. A Metabolic Biomarker Panel for Congenital Heart Disease Assessment with Newborn Dried Blood Spots. *medRxiv*, 2023.2008.2001.23293520 (2023). <https://doi.org/10.1101/2023.08.01.23293520>
42. Zhang, D. et al. Kynurenine promotes neonatal heart regeneration by stimulating cardiomyocyte proliferation and cardiac angiogenesis. *Nat. Commun.* **13**, 6371. <https://doi.org/10.1038/s41467-022-33734-7> (2022).
43. Simonato, M. et al. Urinary metabolomics reveals kynurenine pathway perturbation in newborns with transposition of great arteries after surgical repair. *Metabolomics* **15**, 145. <https://doi.org/10.1007/s11306-019-1605-3> (2019).
44. van Velzen, C. L. et al. Prenatal detection of congenital heart disease—results of a national screening programme. *BJOG* **123**, 400–407. <https://doi.org/10.1111/1471-0528.13274> (2016).
45. Jullien, S. Newborn pulse oximetry screening for critical congenital heart defects. *BMC Pediatr.* **21**, 305. <https://doi.org/10.1186/s12887-021-02520-7> (2021).
46. Xia, J., Mandal, R., Sinelnikov, I. V., Broadhurst, D. & Wishart, D. S. MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis. *Nucleic Acids Res.* **40**, W127–W133. <https://doi.org/10.1093/nar/gks374> (2012).
47. Bahado-Singh, R. O. et al. Precision gynecologic oncology: circulating cell free DNA epigenomic analysis, artificial intelligence and the accurate detection of ovarian cancer. *Sci. Rep.* **12**, 18625. <https://doi.org/10.1038/s41598-022-23149-1> (2022).
48. Alakwaa, F. M., Chaudhary, K. & Garmire, L. X. Deep learning accurately predicts estrogen receptor status in breast Cancer Metabolomics Data. *J. Proteome Res.* **17**, 337–347. <https://doi.org/10.1021/acs.jproteome.7b00595> (2018).
49. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **12**, 77. <https://doi.org/10.1186/1471-2105-12-77> (2011).
50. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
51. Kim, W. J. et al. Cox Proportional Hazard Regression Versus a Deep Learning Algorithm in the prediction of dementia: an analysis based on Periodic Health examination. *JMIR Med. Inf.* **7**, e13139. <https://doi.org/10.2196/13139> (2019).

52. Xia, J., Psychogios, N., Young, N. & Wishart, D. S. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* **37**, W652–W660. <https://doi.org/10.1093/nar/gkp356> (2009).
53. Pang, Z. et al. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **49**, W388–W396. <https://doi.org/10.1093/nar/gkab382> (2021).

Author contributions

RBS and OT had role in conceptualization, RBS and OT wrote the manuscript, SFG, NA, AY worked in editing, NA, AI and AY performed the metabolomics analysis and data acquisition; PF contributed to sample and data collection, BA performed the artificial intelligence analysis, OT and SFG completed validation and formal analysis, RBS and SFG funded resources and supervised project administration. All authors have read and agreed to the published version of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-85216-7>.

Correspondence and requests for materials should be addressed to O.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025