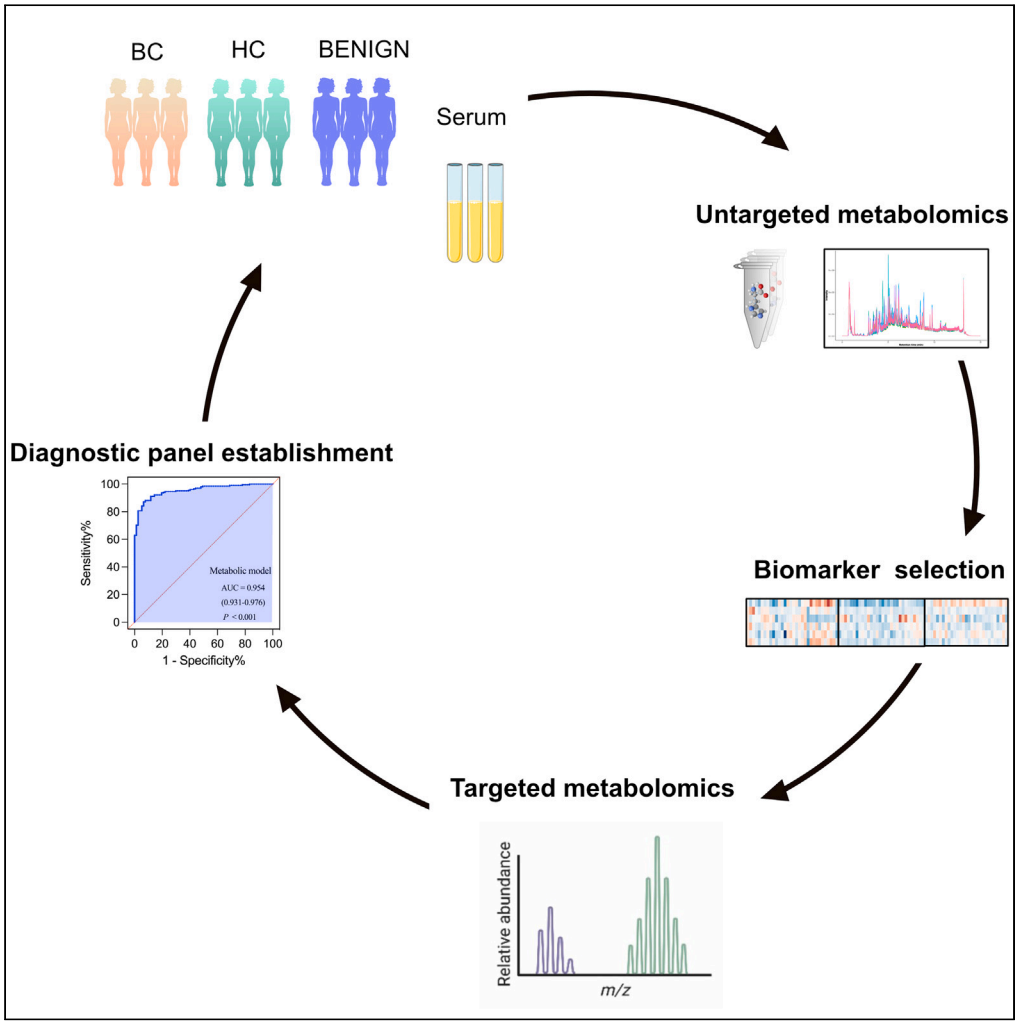
iScience

CellPress
OPEN ACCESS

Article

Largescale multicenter study of a serum metabolite biomarker panel for the diagnosis of breast cancer



Yanzhong Wang, Rui An, Haitao Yu, ..., Yuzhen Gao, Xinyou Xie, Jun Zhang

jameszhang2000@zju.edu.cn

Highlights

Metabolomic features of breast cancer patients are profiled

Metabolite biomarker panel is established for breast cancer detection

Shows potential as a supplement to current screening methods

Wang et al., iScience 27, 110345
July 19, 2024
https://doi.org/10.1016/j.isci.2024.110345

iScience

CellPress
OPEN ACCESS

Article

Largescale multicenter study of a serum metabolite biomarker panel for the diagnosis of breast cancer

Yanzhong Wang,[1,2,8] Rui An,[1,2,8] Haitao Yu,[1,2,8] Yuehong Dai,[1] Luping Lou,[1] Sheng Quan,[3] Rongchang Chen,[3] Yanjun Ding,[1,2] Hongcan Zhao,[4] Xuanlan Wu,[5] Zhen Liu,[6] Qinchuan Wang,[7] Yuzhen Gao,[1,2] Xinyou Xie,[1,2] and Jun Zhang[1,2,9,*]

SUMMARY

Breast cancer (BC) is currently the most prevalent malignancy worldwide, and finding effective non-invasive biomarkers for routine clinical detection of BC remains a significant challenge. Here, we performed non-targeted and targeted metabolomics analysis on the screening, training and validation cohorts of serum samples from 1,947 participants. A metabolite biomarker model including glutamate, erythronate, docosahexaenoate, propionylcarnitine, and patient's age was established for detecting BC. This model demonstrated better diagnostic performance than carbohydrate antigen 15-3 (CA15-3) and carcinoembryonic antigen (CEA) alone in discriminating BC from healthy controls both in the training and validation cohorts [area under the curve (AUC), 0.954; sensitivity, 87.1% and specificity, 93.5% for the training cohort and 0.834, 68.3%, and 85.2%, respectively, for the validation cohort 1]. This study has established a noninvasive approach for the detection of BC, which shows potential as a suitable supplement to the clinical screening methods currently employed for BC.

INTRODUCTION

Breast cancer (BC) is the most frequently diagnosed malignancy and the leading cause of cancer-related mortality in women worldwide.[1] China currently has the largest number of BC cases and related deaths in the world,[2] and new BC cases accounted for 15.59% (357,200) of all new female cancer cases in 2022.[3] Early stage BC is usually asymptomatic, approximately 20% of BC patients were diagnosed with advanced BC when diagnosed,[4] and China had more patients at the advanced stage compared to the U.S.[5] Tumor stage at diagnosis is closely associated with BC survival. Specifically, the 5-year survival rate for stage I tumor patients is almost 100%, but it drops dramatically for patients who have developed distant metastases.[6] No effective preventive strategies are clinically available for BC, making early diagnosis of BC crucial in allowing patients to undergo potentially curative treatments in time, thereby reducing BC mortality.

The National Comprehensive Cancer Network (NCCN) recommends breast mammography, breast ultrasonography, and breast MRI for early breast cancer diagnosis.[7] Since the sensitivity of mammography in the high-risk population is low at 55%,[8] breast MRI typically supplements mammography to increase screening sensitivity.[9] Additionally, interpretation and analysis of the images often rely heavily on the expertise of clinicians, leading to inherent deviations.[10] For noninvasive screening, carbohydrate antigen 15-3 (CA15-3) and carcinoembryonic antigen (CEA) are currently the most widely used serum clinical tumor markers.[11] However, both have low sensitivity, especially in early stage BC patients, increasing the urgency of finding more sensitive and noninvasive biomarkers for the screening and diagnosis of BC at the early stage.

New biomarkers of BC, such as circulating miRNAs, circulating tumor DNA, and exosomes, have been reported.[12] Many studies have confirmed that metabolites may be effective biomarkers for BC.[13] For example, high levels of histidine, glycerol, *N*-acetyl glycoproteins, and ethanol are reported to be associated with higher risk of BC in premenopausal woman.[14] Metabolic characteristics could also help in

[1]Department of Clinical Laboratory, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, 3 East Qingchun Road, Hangzhou, Zhejiang, People's Republic of China
[2]Key Laboratory of Precision Medicine in Diagnosis and Monitoring Research of Zhejiang Province, 3 East Qingchun Road, Hangzhou, Zhejiang, People's Republic of China
[3]Hangzhou Calibra Diagnostics Co., Ltd. (A Subsidiary of DIAN Diagnostics), 329 Jinpeng Street, Xihu Industrial Park, Hangzhou, Zhejiang, People's Republic of China
[4]Department of Clinical Laboratory, Affiliated Hangzhou First People's Hospital, Westlake University School of Medicine, 261 Huansha Road, Hangzhou, Zhejiang, People's Republic of China
[5]Department of Clinical Laboratory, Sir Run Run Shaw Hospital Xiasha Campus, Zhejiang University School of Medicine, 368 Xiasha Road, Hangzhou, Zhejiang, People's Republic of China
[6]Department of Clinical Laboratory, Ningbo Medical Centre Lihuili Hospital, Ningbo University, 1111 Jiangnan Street, Ningbo, Zhejiang, People's Republic of China
[7]Department of Surgical Oncology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, 3 East Qingchun Road, Hangzhou, Zhejiang, People's Republic of China
[8]These authors contributed equally
[9]Lead contact
*Correspondence: jameszhang2000@zju.edu.cn
https://doi.org/10.1016/j.isci.2024.110345

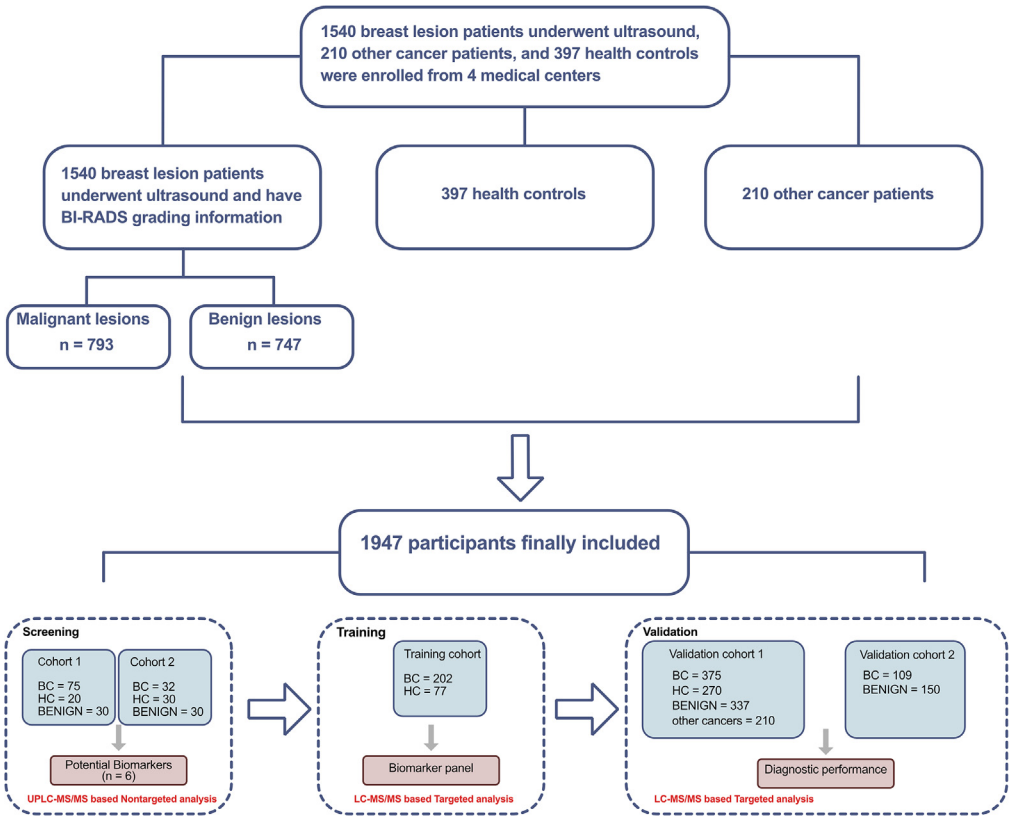


Figure 1. Study design of this study

identifying BC subtypes and guiding personalized treatment. Yue Gong et al. classified triple negative breast cancer into three different metabolic subtypes (MPS1, MPS2, and MPS3) and proposed corresponding treatment protocols for each metabolic subtype.[15] Metabolites, the downstream products of genes and proteins, are more intuitive predictors of phenotypic changes compared to genes and proteins.[16] An increasing number of studies have utilized metabolomics to explore diagnostic biomarkers for BC.[17,18] However, these studies have certain limitations, including, for example, (1) small sample size; (2) lack of validation of the discovered potential biomarkers; and (3) validated studies not conducted on a large scale or tested using samples from a single clinical location.

This study aimed to identify and validate new metabolite biomarkers of BC in a large-scale multicenter study and establish a diagnostic model incorporating the selected metabolite biomarkers to detect BC.

RESULTS

Demographic characteristics of study cohorts

The workflow of this study is shown in Figure 1. To define potential metabolite biomarkers, 217 participants in the screening set were recruited. 279 participants containing 202 BC patients and 77 healthy control (HC) individuals were recruited in the training cohort to determine the metabolite model to be used for the diagnosis of BC. The model was further evaluated for its diagnostic performance in two independent validation cohorts (375 BC, 337 BENIGN, 270 HC, and 210 other cancer patients in validation cohort 1; 109 BC and 150 BENIGN in validation cohort 2). Most BC patients (75.79%) were in the early stages of tumor [Tumor Node Metastasis (TNM) I + II]. Importantly, post-operative serum samples from 57 BC patients were also collected to evaluate the change in trend of candidate biomarkers before and after operation. Moreover, serum samples of seven other cancers, including lung cancer (LC), thyroid cancer (TC), gastric cancer (GC), colorectal cancer (CC), rectal cancer (RC), liver cancer (LIC), and pancreatic cancer (PC) were collected to evaluate the specificity of the metabolite biomarkers panel for BC. Detailed demographic and clinical characteristics of all participants were summarized in Table S1.

Selection of potential metabolite biomarkers for BC

Non-targeted ultrahigh performance liquid chromatography-tandem mass spectrometry analysis (UPLC-MS/MS) metabolomics analysis was performed in the screening phase of 217 serum samples to define potential metabolite biomarkers of BC. Principal component analysis (PCA) did not show significant separation among the three groups (Figures S1A and S1B). Samples from the BC, BENIGN, and HC groups showed trends of separation in the partial least squares discriminant analysis (PLS-DA) score plot (Figures 2A and 2B), and validation of the PLS-DA is

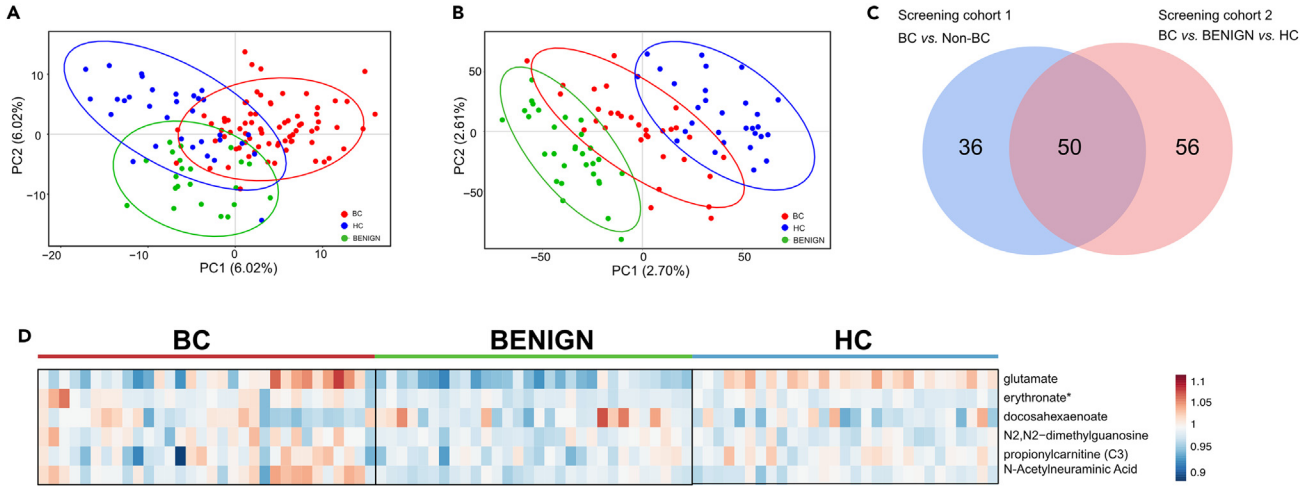


Figure 2. UPLC-MS/MS based nontargeted metabolomic analysis

(A and B) Score plots of PLS-DA models showing the separation among the BC, BENIGN, and HC groups in the screening cohort 1 and screening cohort 2, respectively. Each point represents a sample, red: BC patients, green: benign patients, blue: HC controls.

(C) Veen diagram showing the 50 differential metabolites that were selected as potential biomarkers from the group comparisons in the screening cohort 1 and screening cohort 2.

(D) Heatmap showing the relative levels of these six candidate metabolites in the screening cohort 2.

suggestive of an excellent model (Figures S1C and S1D). Model parameters for the explained ability (R2) and predictive ability (Q2) for the two models were acceptable (R2Y = 0.548, Q2 = 0.054 and R2Y = 0.865, Q2 = 0.048, respectively). Potential metabolite biomarkers were screened from the Welch's t test and Random forest (RF) analysis. 50 metabolites were identified in both group comparisons in the two screening cohorts (Figure 2C). After removing xenobiotics, small peptides, vitamins, and metabolites with more than 30% missing values, 41 metabolites remained. These 41 metabolites were then ranked according to their average effect sizes between BC vs. non-BC and BC vs. BENIGN vs. HC (Table S2). The final list of candidate BC-related metabolite biomarkers was selected based on (1) their ranking among the top 35 in the RF analysis (Tables S3 and S4) and (2) high mean effect size in each metabolite super-pathway. Ultimately, six metabolites were retained: glutamate, erythronate, docosahexaenoate, N^2,N^2-dimethylguanosine, propionylcarnitine, and N-acetylneuraminic acid. The detailed comparative fragmentation spectra for the six candidate metabolites were shown in Table S5. And the figures that juxtapose our experimentally obtained fragmentation spectra against the corresponding spectra from the reference library were shown in Figure S2. The relative levels of these six candidate metabolites in screening cohort 1 and screening cohort 2 were presented in heatmaps (Figure S1E; Figure 2D).

Development and validation of targeted LC-MS/MS assays of the six metabolites

After rounds of continuous method optimization, three LC-MS/MS detection methods (A, B, and C) for the six candidate metabolites were established. Docosahexaenoate, N-acetylneuraminic acid, propionylcarnitine, and N^2,N^2-dimethylguanosine were detected with Method A, while erythronate was detected with method B, and glutamate was detected with method C. The chromatograms of the standard solutions of the six target metabolites in LC-MS/MS were shown in Figure S3.

Seven levels of standards were used to generate calibration curves for the six metabolites, and the calibration curves showed good linearity and excellent coefficients ($r^2 > 0.999$) for all six metabolites. The limit of quantification (LOQ) for glutamate, erythronate, docosahexaenoate, N^2,N^2-dimethylguanosine, propionylcarnitine, and N-acetylneuraminic acid were 0.1 μg/mL, 50 ng/mL, 0.25 μg/mL, 1 ng/mL, 5 ng/mL, and 20 ng/mL, respectively (Table 1). As shown in Table 1, intra-batch precision and inter-batch precision were both less than 15% for all six metabolites at each concentration level. Specifically, intra-batch precision for the six metabolites ranged from 1.2% to 5.19%, and inter-batch precision ranged from 2.42% to 9.26%. Accuracy values (in relative bias, bias%) for the six metabolites ranged from −10.14% to 16.06%. Thus, both precision and accuracy were within acceptable limits, laying a foundation for accurate quantification of the six metabolites in human serum.

Stability was evaluated at different storage or processing conditions expressed as accuracy. As shown in Table S9, the short-term stability for 6 h at room temperature, 4°C for 1 day and 4°C for 3 days was recorded as −2%–6.13%, −0.72%–5.15%, and −1.40%–8.83%, respectively. Mid-term stability at −20°C for 1 week and −20°C for 2 weeks was recorded as −2.43%–9.18% and −2.35%–14.24%, respectively. Serum samples were stable over two freeze-thaw cycles at −20°C with an accuracy between −3.62% and 13.66% for the six metabolites. These bias values were acceptable and demonstrated that serum samples were stable under the tested storage or processing conditions.

Metabolite biomarker panel for the identification of BC patients

Binary logistic regression was conducted to explore the optimum serum metabolite combination for the detection of BC patients. In univariate binary logistic regression analysis, all six candidate metabolites and age, body mass index (BMI), menopause status, CA15-3, and CEA

Table 1. Targeted metabolomics methodology validation (*n* = 6)

Metabolites	LOQ	Quantification Range	Regression coefficient	Intra-batch precision (CV %)	Inter-batch precision (CV %)	Accuracy (Bias%)	Recovery	S/N
Docosahexaenoate	0.25 μg/mL	0.25–25 μg/mL	0.9999	1.20–1.84	4.76–9.26	−10.14∼-7.37	82.5–100.5	28.3
Erythronate	50 ng/mL	50-5000 ng/mL	0.9991	1.58–3.65	4.17–4.39	2.89–16.06	83.6–103.9	2816.2
N-Acetylneuraminic Acid	20 ng/mL	20-2000 ng/mL	0.9998	2.24–3.60	6.33–8.75	−5.54∼3.29	87.1–115.1	13.5
Propionylcarnitine	5 ng/mL	5-500 ng/mL	0.999	1.84–4.01	2.42–4.45	0.11–2.87	88.5–109.3	447.1
N^2,N^2-dimethylguanosine	1 ng/mL	1-100 ng/mL	0.9998	1.69–5.19	2.58–4.97	−3.09∼-0.56	92.1–110.9	69
Glutamate	0.1 μg/mL	0.1–10 μg/mL	0.9999	2.48–4.76	2.63–3.72	0.21–1.77	87.6–108.7	90.1

LOQ, limit of quantification; S/N, Signal/Noise.

were associated with BC (Figure 3A). Age, glutamate, erythronate, docosahexaenoate, and propionylcarnitine remained significantly correlated with BC in multivariate binary logistic regression analysis (Figure 3B). Thus, the combination of glutamate, erythronate, docosahexaenoate, and propionylcarnitine and age was finally determined to be the optimum panel for the identification of BC, and the model was constructed as:

$$[P = BC] = 1 \Big/ \left[1 + e^{(-10.982 + 0.274 \times \text{Glutamate} + 0.007 \times \text{Erythronate} + 0.839 \times \text{Docosahexaenoate} - 0.529 \times \text{Propionylcarnitine} + 0.121 \times \text{Age})}\right]$$

In this equation, [P = BC] was the predicted probability of BC using the developed panel. Glutamate, erythronate, docosahexaenoate, and propionylcarnitine were the measured serum concentrations of each metabolite. The serum levels of these four metabolites in the training cohort were presented in Figures 3C–3F. Compared to HC, glutamate, erythronate, and docosahexaenoate were upregulated in BC, whereas propionylcarnitine was downregulated. The cut-off value of [P = BC] was 0.768, and values greater than this cut-off were determined to be BC. The diagnostic performance of this metabolite panel was evaluated by examining AUC, specificity, and sensitivity in receiver-operating characteristics curve (ROC) analysis of the panel. The diagnostic accuracy of this metabolite panel was higher than that of individual metabolites alone (Figures 4A–4D). When comparing this model against CA15-3 and CEA for the ability to discriminate BC from HC in the training cohort, the panel had an AUC of 0.954, while the AUCs of CA15-3 and CEA were 0.604 and 0.721, respectively (Figures 4E–4G). The sensitivity of this metabolite panel in BC diagnosis was obviously higher than that of either CA15-3 or CEA (with a clinical cut-off of 25 U/mL and 5 ng/mL, respectively), which were 87.1%, 6.2%, and 7.7%, respectively. The specificity of the metabolite model for BC (93.5%) was slightly lower than that of CA15-3 (97.4%) and CEA (100.0%) (Table 2).

The panel was further evaluated in validation cohort 1, which included 1,192 participants from multiple clinical centers. Notably, 337 BENIGN patients were enrolled in this cohort to enrich the clinical background of the controls. Consistent with the results of the training cohort, the metabolite panel showed a better diagnostic performance than that of either CA15-3 or CEA not only for BC but also for early stage BC. The AUCs for BC detection using the metabolite panel, CA15-3, and CEA were 0.834, 0.641, and 0.733, respectively. The AUCs for early stage BC detection were 0.832, 0.609, and 0.713, respectively. Similarly, the sensitivity of the metabolite panel was also better than that of either CA15-3 or CEA for the detection of BC and early stage BC in validation cohort 1 (Table 2). Although the diagnostic ability of this metabolite model decreased when incorporating BENIGN patients, it still outperformed CA15-3 and CEA with AUCs of 0.741, 0.605, and 0.681, respectively). In addition, we managed to collect both pre-operative and post-operative serum samples from 57 BC patients. Compared with their pre-operative serum levels, post-operative levels of propionylcarnitine were significantly elevated, and post-operative levels of glutamate and docosahexaenoate were significantly decreased, while the levels of erythronate were comparable between pre-operative and post-operative (Figures S4A–S4D).

Comparison of diagnostic performance of metabolite model and ultrasound

Validation cohort 2, consisting of 259 participants, was used to further evaluate the constructed metabolite model. 259 breast lesion patients with breast imaging-reporting and data system (BI-RADS) grading information were prospectively recruited in validation cohort 2. All 259 patients underwent pathological examination, and 109 patients were breast cancer, while 150 patients were breast benign disease. Meanwhile, these 259 patients were also judged as breast cancer or breast benign disease according to the cut-off value of the metabolite model (0.768) and IVA classification of ultrasound results, showing category 1-IVA as benign, and category IVB-6 as malignant. These results were then compared with the pathology results. Specifically, based on the metabolite model, 143 and 116 patients were judged as breast cancer

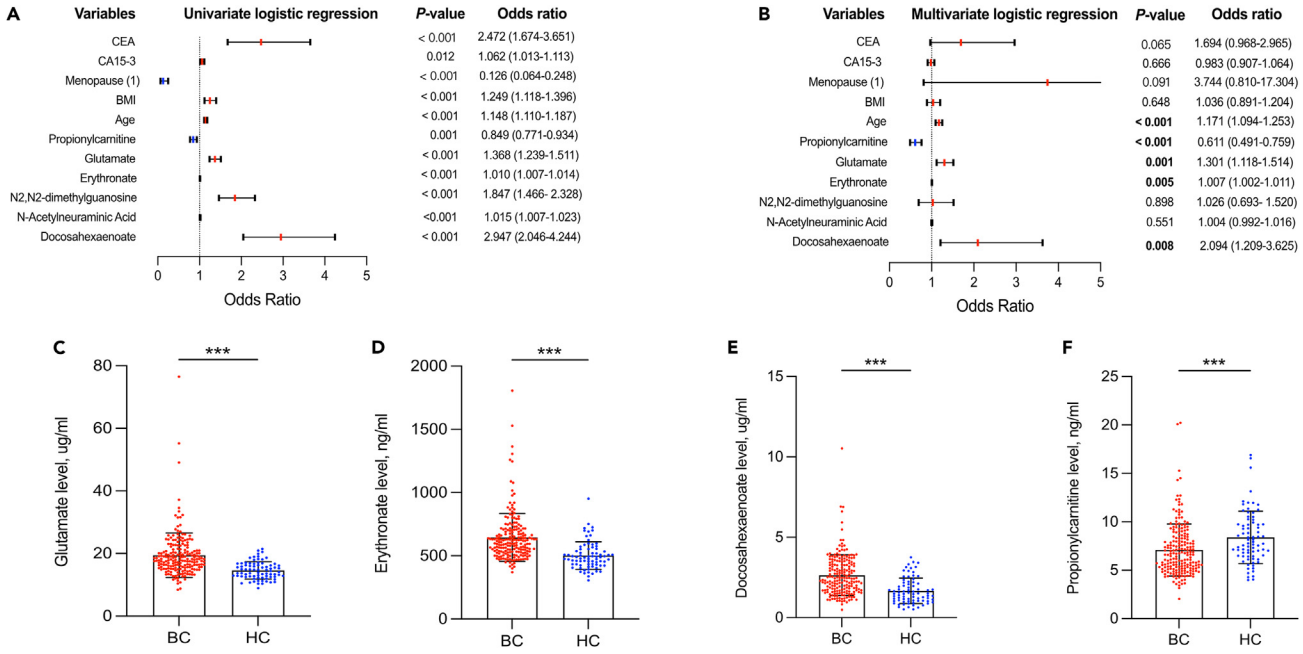


Figure 3. Identification of biomarker candidates for the diagnosis of BC

(A and B) Univariate and multivariate binary logistic regression analysis of the risk factors for the incidence of BC. "Menopause (1)" indicates that "post-menopause" was used as a reference.

(C–F) Serum concentrations of the glutamate, erythronate, docosahexaenoate, and propionylcarnitine in the training cohort. Data are presented as mean ± SD, comparisons were performed using Student's t test, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. BMI, body mass index; CA15-3, carbohydrate antigen 15-3; CEA, carcinoembryonic antigen; BC, breast cancer; HC, health controls.

and breast benign disease, respectively. 106 were breast cancer and 153 were breast benign disease according to ultrasound. Based on the combination of metabolite model and ultrasound, 79 were breast cancer and 180 were breast benign disease (Table S10).

As shown in Table 3, standalone ultrasound showed higher accuracy (85.7% vs. 76.8%), specificity (88.7% vs. 68.7%), and positive predictive value (PPV) (84.0% vs. 67.1%) than the metabolite model. However, metabolite model had better negative predictive value (NPV) and sensitivity than those of ultrasound (88.8% vs. 86.9% and 88.1% vs. 81.7%, respectively). Importantly, the metabolite model combined with ultrasound was superior to ultrasound alone in PPV (96.2% vs. 84.0%) and specificity (98.0% vs. 88.7%), but inferior in sensitivity (69.7% vs. 81.7%) and NPV (81.7% vs. 86.9%). Thus, the diagnostic performance of the metabolite model alone is inferior to that of ultrasound. Nonetheless, the combination of metabolite model and ultrasound can obviously improve PPV and specificity, which would reduce unnecessary breast biopsies.

Specificity of the metabolite panel for BC

The four candidate metabolites were also measured in seven other cancers, including LC, TC, GC, CC, RC, LIC, and PC, to evaluate their specificity for BC. Overall, the levels of glutamate, propionylcarnitine, and erythronate were significantly different between BC patients and other cancer patients (Figures 5A–5C). Compared to GC, CC, RC, LIC, and PC patients, the serum levels of glutamate and propionylcarnitine were significantly decreased in BC patients, whereas the levels of erythronate were increased in BC subjects (Figures 5E–5G). Docosahexaenoate levels were comparable in BC patients and other cancer patients (Figures 5D and 5H). Thus, the combination of these four metabolites discriminated patients with BC from patients with GC, CC, RC, LIC, and PC with AUC of 0.848, 0.823, 0.777, 0.857, and 0.807, respectively (Figures 5I–5M), indicating good specificity of these four metabolite biomarkers for BC.

DISCUSSION

BC has become one of the biggest health threats to women worldwide, and early detection and intervention are crucial for BC patients. The identification of effective biomarkers for the diagnosis of BC remains a daunting challenge, especially for early stage BC. Metabolomics, the comprehensive assessment of metabolites, has emerged as a powerful tool for cancer diagnosis, including BC. Metabolomics could detect subtle changes in metabolite levels that may be indicative of early stage disease, which offers a promise for earlier detection than current imaging techniques. In addition, new metabolomics technology like nanomaterial-based metabolomics provides enhanced selectivity and sensitivity,[19,20] Huang et al. used fast nanoparticle-enhanced laser desorption/ionization mass spectrometry (NPELDI-MS) to record serum metabolic fingerprints of BC in seconds, achieving high reproducibility and low consumption of direct serum detection without treatment.[21]

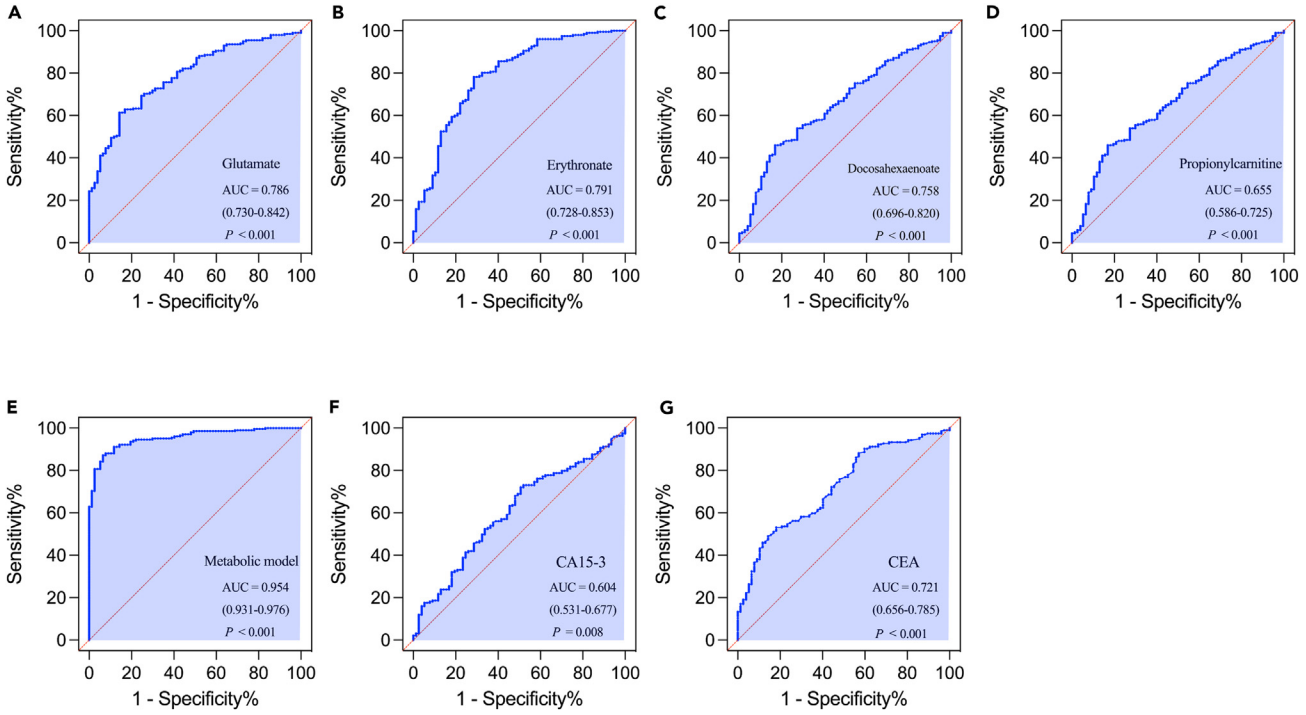


Figure 4. Diagnostic performance of the metabolic panel

(A–D) ROC curves of the four metabolic biomarkers in the training cohort.

(E) ROC curves of the metabolic model in the training cohort.

(F and G) ROC curves of CA15-3 and CEA in the training cohort. AUC, area under the curve; CA15-3, carbohydrate antigen 15-3; CEA, carcinoembryonic antigen.

Moreover, metabolite biomarkers can be monitored non-invasively through blood tests, providing a convenient and patient-friendly approach to disease surveillance.

In this study, we developed and validated a serum metabolite model for BC detection in a total of 1,947 participants. Six metabolite biomarkers potentially related to BC were first selected by Welch's t test and RF from a non-targeted UPLC-MS/MS analysis in the screening phase. Then, three targeted LC-MS/MS methods for the absolute quantification of these six metabolites were established. The developed targeted LC-MS/MS methods had good test linearity, precision, and accuracy. After further univariate and multivariate binary logistic regression analysis incorporating the six metabolites and other clinical factors, a panel consisting of glutamate, erythronate, docosahexaenoate, propionylcarnitine, and age was finally determined. Compared to CA15-3 and CEA, this panel performed better in discriminating BC from HC with higher sensitivity in both training cohort and validation cohort 1 (Table 2). When applied in the identification of early stage BC, this panel also performed better than CA15-3 and CEA (Table 2). However, the diagnostic performance of this metabolite model decreased when incorporating BENIGN patients, indicating that the current model cannot effectively distinguish BC patients from BENIGN patients. The diagnostic performance of this metabolite model alone is inferior to that of ultrasound; however, the combination of metabolite model and ultrasound obviously improved PPV and specificity, which would reduce unnecessary breast biopsies. Furthermore, the serum levels of glutamate, erythronate, and propionylcarnitine in BC patients were significantly different from other cancer patients, and the panel of four candidate metabolites showed good performance in discriminating BC patients and other cancer patients, indicating good specificity of these four metabolite biomarkers for BC. Thus, this metabolite panel has high sensitivity and accuracy for the diagnosis of BC, and when it is combined with ultrasound, PPV and specificity are even improved, making it a promising panel for noninvasive screening of BC and a potential supplement to the currently available clinical screening methods for BC. Besides, the serum levels of glutamate and docosahexaenoate were significantly downregulated, propionylcarnitine were significantly upregulated, after effective surgical treatment, indicating that these three metabolites may also be a valuable biomarker for the assessment of therapeutic effect on BC patients.

Metabolic reprogramming is a hallmark of tumors,[22] suggesting that the dysregulation of metabolic pathways and metabolites may play an important role in the development of BC. Among the four metabolites in the panel, glutamate is widely reported to be closely associated with cancer through the glutamine-glutamate metabolic pathway.[23,24] In addition to our study, many metabolomic studies have also identified glutamate as a metabolic biomarker of BC.[25–27] Rapidly proliferating cancer cells have a high demand for energy, and in addition to the Warburg effect,[28] "glutamine addiction" is another important metabolic feature of tumor cells.[23] Glutamine metabolism refers to the process whereby cells convert glutamine into tricarboxylic acid cycle (TCA cycle) metabolites through various enzymes. Specifically, glutamine is first converted into glutamate by cytosol glutaminase (GLS/GLS2), followed by the conversion of glutamate into α-ketoglutarate (α-KG) by

Table 2. Results of measurement of serum metabolites panel, CEA, and CA15-3 in the diagnosis of BC and early stage BC

	Training cohort				Validation cohort 1			
	AUC (95% CI)	Sensitivity (%)	Specificity (%)	*p* values[#]	AUC (95% CI)	Sensitivity (%)	Specificity (%)	*p* values[#]
BC vs. HC								
Metabolic model	0.954 (0.931–0.976)	87.1 (80.2–94.6)	93.5 (90.9–100.0)		0.834 (0.803–0.864)	68.3 (66.1–77.1)	85.2 (73.7–89.6)	
CA15-3	0.604 (0.531–0.677)	6.2 (1.0–21.8)	97.4 (93.5–100.0)	<0.001	0.641 (0.598–0.684)	8.6 (3.5–13.7)	98.9 (97.0–100.0)	<0.001
CEA	0.721 (0.656–0.786)	7.7 (3.3–24.2)	100.0 (100.0–100.0)	<0.001	0.733 (0.695–0.772)	8.0 (5.1–22.0)	99.6 (98.9–100.0)	<0.001
Early stage BC vs. HC								
Metabolic model	–	–	–		0.832 (0.796–0.868)	71.4 (63.4–77.2)	80.8 (67.4–88.5)	
CA15-3	–	–	–		0.609 (0.546–0.672)	6.2 (1.2–14.2)	98.6 (93.8–100.0)	<0.001
CEA	–	–	–		0.713 (0.657–0.770)	5.6 (2.5–22.2)	99.3 (98.0–100.0)	0.001
BC vs. NonBC								
Metabolic model	–	–	–		0.741 (0.709–0.772)	70.9 (65.3–76.0)	66.6 (59.1–73.3)	
CA15-3	–	–	–		0.605 (0.569–0.641)	8.6 (5.6–13.1)	97.3 (94.8–99.0)	<0.001
CEA	–	–	–		0.681 (0.647–0.714)	8.0 (4.3–13.4)	98.8 (97.2–99.7)	0.003

BC, breast cancer; HC, health controls; AUC, Area Under Curve; CI, Confidence interval; CA15-3, carbohydrate antigen 153; CEA, carcinoembryonic antigen.[#]: *p* values for the comparison of diagnostic performance between the metabolic model and CA153/CEA.

glutamate dehydrogenase (GLUD) or amino-transaminase, finally, α-KG enters the TCA cycle as a replenishing substrate to maintain the normal mitochondrial function.[23] This process can rapidly produce a large amount of ATP to meet the energy needs of tumor cells. Proliferating tumor cells, including pancreatic cancer, ovarian cancer, and breast cancer, rely on glutamine as the main energy source to meet the requirements of TCA cycle. Once these cells are deprived of glutamine, cell growth arrest and even cell death soon follow.[29–31] Under normal physiological conditions, glutamate can convert to other non-essential amino acids through transamination.[23,32] Glutamate is also the precursor for glutathione, an antioxidant that acts as a free radical scavenger and antidote in cells.[33] Glutamate is an excitatory neurotransmitter that transmits signals in the central nervous system by binding to different types of receptors.[34] In fact, glutamate signaling was initially thought to be specific to the nervous system, but growing evidence supports the idea that glutamate signaling, through its receptors, plays an important role in modulating cancer cell development, proliferation, and metastasis.[33,35,36] For example, dysfunction of excitatory amino acid transport proteins and the cystine-glutamate exchanger (X_C-system) caused excessive glutamate concentrations in extracellular space, which, in turn, activated glutamate receptors on cancer cells, leading to malignant cell growth.[37] These features make glutamate an attractive therapeutic target for anticancer drug development.

Propionylcarnitine is a compound composed of carnitine and propionyl groups. Propionylcarnitine is transported across the cell membranes by the carnitine system and serves as mitochondrial fuel for fatty acid oxidation, providing energy and metabolic intermediates.[38,39] Besides, carnitine plays an important role in modulating carbohydrate and lipid metabolism by regulating acyl-CoA/CoA balance.[38] Carnitine supply is necessary for fatty acid β-oxidation via mitochondria, which is one of the most productive energy-producing pathways in cells.[40] Tumor cells need a large amount of energy for rapid proliferation and progression to malignancy.[22] Therefore, we speculate that most carnitine is mobilized to tumor tissues for energy production, which decreases the serum carnitine pool. Docosahexaenoate is an omega-3 polyunsaturated fatty acid. In the present study, docosahexaenoate was found to be elevated in the serum of BC patients and positively associated with the risk of BC. Besides, Park et al. also identified docosahexaenoate as a candidate biomarker for BC.[41] Erythronate is an erythronic acid and a breakdown product of threonine metabolism, hyaluronic acid metabolism, and oxidative stress.[42,43] In the present study, erythronate was higher in the serum of BC patients, and the levels showed a downward trend after surgery. Studies have found that erythronate was strongly associated with hepatorenal dysfunction.[44,45] Unfortunately, there are few reports on the association between docosahexaenoate and erythronate with cancer, and the exact molecular mechanism of the involvement of docosahexaenoate and erythronate in cancers is unclear. However, these two indicators are indispensable for the diagnostic metabolite model established by the current research, and more investigations in this area are needed.

Table 3. Diagnostic performance of metabolic model, ultrasound, and metabolic model combined ultrasound

	Accuracy%	PPV%	NPV%	Sensitivity%	Specificity%	FPR%	FNR%
Metabolic model	76.8	67.1	88.8	88.1	68.7	31.3	11.9
Ultrasound	85.7	84.0	86.9	81.7	88.7	11.3	18.4
Metabolic model+Ultrasound	86.1	96.2	81.7	69.7	98.0	2.0	30.3

PPV, positive predictive value; NPV, negative predictive value; FPR, false positive rate; FNR, false negative rate.

In summary, a metabolite panel consisting of glutamate, erythronate, docosahexaenoate, propionylcarnitine, and age was developed and validated for its ability to detect BC in a large-scale and multicenter setting. The results indicated that this panel could discriminate BC and early stage BC from HC with better performance and higher sensitivity when compared to traditional markers such as CA15-3 and CEA. Additionally, by integrating our metabolite panel with ultrasound, we propose an effective diagnostic approach that could enhance the specificity of breast cancer screening and reduce the rate of unnecessary biopsies. Therefore, the panel offers a practical BC diagnostic method by using only small amounts of serum samples, making it a promising suitable supplement to current clinical screening and diagnostic methods for BC.

Limitations of the study

The current study has several limitations. First, the subjects of this study were all from the Zhejiang region of China and did not include participants from other regions, such as northern China and western China; so, it is impossible to determine the adaptability of this panel to

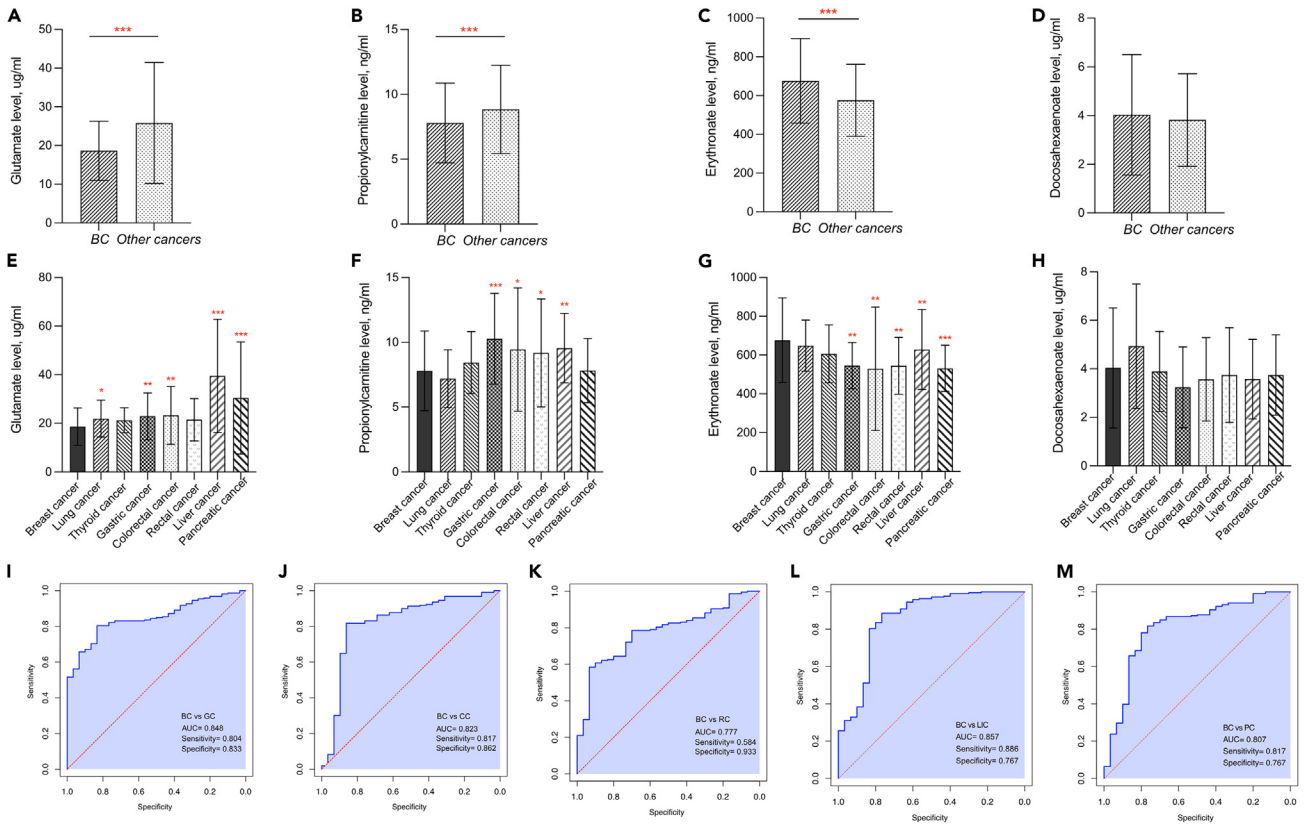


Figure 5. Specificity of the metabolite panel for BC

(A–D) Serum concentrations of glutamate, propionylcarnitine, erythronate, and docosahexaenoate in serum of BC patients and other cancer patients.

(E–H) Serum concentrations of glutamate, propionylcarnitine, erythronate, and docosahexaenoate in the BC, LC, TC, GC, CC, RC, LIC, and PC groups, respectively.

(I–M) ROC curves of the four metabolites combined for BC patients versus patients with GC, CC, RC, LIC, and PC, respectively. Data are presented as mean ± SD, comparisons were performed using Student's t test, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. BC, breast cancer; GC, gastric cancer; CC, colorectal cancer; RC, rectal cancer; LIC, liver cancer; PC, pancreatic cancer.

regional differences. Therefore, before applying this metabolite model to clinical practice, populations from other regions should also be included for comprehensive evaluation. Second, this study did not include a prospective cohort to evaluate the ability of this metabolite model to screen preclinical BC from high-risk populations. Nested case-control studies in prospective cohorts are needed in the future to further validate this metabolite model. Finally, the molecular mechanism of these four metabolite biomarkers relative to BC needs to be further investigated.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Patients and cohorts
- METHOD DETAILS
  - Serum sample collection and pretreatment
  - Non-targeted UPLC-MS/MS metabolomics analysis
  - Targeted LC-MS/MS methods development and validation
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110345.

ACKNOWLEDGMENTS

We gratefully acknowledge Ziqing Kong and Xue Zhang for their efforts in LC-MS/MS analysis. And we also would like to acknowledge Feng Zhao for his efforts in data processing and image mapping.

This study was founded by Key Research and Development Program of Zhejiang Province (no. 2019C03021), National Natural Science Foundation of China (no. 82172362), Key Research and Development Program of Zhejiang Province (no. 2022C03037), Natural Science Foundation of Zhejiang Province (LY22H200002), and Key Laboratory of Precision Medicine in Diagnosis and Monitoring Research of Zhejiang Province (2022E10018).

AUTHOR CONTRIBUTIONS

J.Z.: conceptualization, project administration, supervision, funding acquisition, and writing – review and editing. Y.W. and R.A.: methodology, data curation, formal analysis, software, and writing – original draft. H.Y.: methodology and data curation. D.Y. and L.L.: methodology and writing – review and editing. S.Q.: writing – review and editing. Y.D., H.Z., X.W., Z.L., and Q.W.: sample collection. Y.G. and R.C.: software. X.X.: writing – review and editing. The authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 25, 2024
Revised: May 23, 2024
Accepted: June 19, 2024
Published: June 21, 2024

REFERENCES

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA. Cancer J. Clin. *71*, 209–249. https://doi.org/10.3322/caac.21660.
2. Lei, S., Zheng, R., Zhang, S., Wang, S., Chen, R., Sun, K., Zeng, H., Zhou, J., and Wei, W. (2021). Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020. Cancer Commun. *41*, 1183–1194. https://doi.org/10.1002/cac2.12207.
3. Han, B., Zheng, R., Zeng, H., Wang, S., Sun, K., Chen, R., Li, L., Wei, W., and He, J. (2024). Cancer incidence and mortality in China, 2022. J. Natl Cancer Center *4*, 47–53. https://doi.org/10.1016/j.jncc.2024.01.006.
4. Li, J., Zhang, B.N., Fan, J.H., Pang, Y., Zhang, P., Wang, S.L., Zheng, S., Zhang, B., Yang, H.J., Xie, X.M., et al. (2011). A nationwide multicenter 10-year (1999-2008) retrospective clinical epidemiological study of female breast cancer in China. BMC

Cancer *11*, 364. https://doi.org/10.1186/1471-2407-11-364.
5. Qian, X., Zou, X., Xiu, M., Liu, Y., Chen, X., Xiao, M., and Zhang, P. (2023). Epidemiology and clinicopathologic features of breast cancer in China and the United States. Transl. Cancer Res. *12*, 1826–1835. https://doi.org/10.21037/tcr-22-2799.
6. Kadys, A., Gremke, N., Schnetter, L., Kostev, K., and Kalder, M. (2023). Intercontinental comparison of women with breast cancer treated by oncologists in Europe, Asia, and Latin America: a retrospective study of 99,571 patients. J. Cancer Res. Clin. Oncol. *149*, 7319–7326. https://doi.org/10.1007/s00432-023-04681-7.
7. Bevers, T.B., Helvie, M., Bonaccio, E., Calhoun, K.E., Daly, M.B., Farrar, W.B., Garber, J.E., Gray, R., Greenberg, C.C., Greenup, R., et al. (2018). Breast Cancer Screening and Diagnosis, Version 3.2018, NCCN Clinical Practice Guidelines in Oncology. J. Natl. Compr. Canc. Netw. *16*, 1362–1389. https://doi.org/10.6004/jnccn.2018.0083.
8. Phi, X.A., Houssami, N., Hooning, M.J., Riedl, C.C., Leach, M.O., Sardanelli, F., Warner, E., Trop, I., Saadatmand, S., Tilanus-Linthorst, M.M.A., et al. (2017). Accuracy of screening women at familial risk of breast cancer without a known gene mutation: Individual patient data meta-analysis. Eur. J. Cancer *85*, 31–38. https://doi.org/10.1016/j.ejca.2017.07.055.
9. Chiarelli, A.M., Blackmore, K.M., Muradali, D., Done, S.J., Majpruz, V., Weerasinghe, A., Mirea, L., Eisen, A., Rabeneck, L., and Warner, E. (2020). Performance Measures of Magnetic Resonance Imaging Plus Mammography in the High Risk Ontario Breast Screening Program. J. Natl. Cancer Inst. *112*, 136–144. https://doi.org/10.1093/jnci/djz079.
10. Gao, Y., Lin, J., Zhou, Y., and Lin, R. (2023). The application of traditional machine learning and deep learning techniques in mammography: a review. Front. Oncol. *13*, 1213045. https://doi.org/10.3389/fonc.2023.1213045.
11. Tarighati, E., Keivan, H., and Mahani, H. (2023). A review of prognostic and predictive biomarkers in breast cancer. Clin. Exp. Med. *23*, 1–16. https://doi.org/10.1007/s10238-021-00781-1.
12. Loke, S.Y., and Lee, A.S.G. (2018). The future of blood-based biomarkers for the early detection of breast cancer. Eur. J. Cancer *92*, 54–68. https://doi.org/10.1016/j.ejca.2017.12.025.
13. Silva, C., Perestrelo, R., Silva, P., Tomás, H., and Câmara, J.S. (2019). Breast Cancer Metabolomics: From Analytical Platforms to Multivariate Data Analysis. A Review. Metabolites *9*, 102. https://doi.org/10.3390/metabo9050102.
14. Jobard, E., Dossus, L., Baglietto, L., Fornili, M., Lécuyer, L., Mancini, F.R., Gunter, M.J., Trédan, O., Boutron-Ruault, M.C., Elena-Herrmann, B., et al. (2021). Investigation of circulating metabolites associated with breast cancer risk by untargeted metabolomics: a case-control study nested within the French E3N cohort. Br. J. Cancer *124*, 1734–1743. https://doi.org/10.1038/s41416-021-01304-1.
15. Gong, Y., Ji, P., Yang, Y.S., Xie, S., Yu, T.J., Xiao, Y., Jin, M.L., Ma, D., Guo, L.W., Pei, Y.C., et al. (2021). Metabolic-Pathway-Based Subtyping of Triple-Negative Breast Cancer Reveals Potential Therapeutic Targets. Cell Metab. *33*, 51–64.e9. https://doi.org/10.1016/j.cmet.2020.10.012.
16. Rinschen, M.M., Ivanisevic, J., Giera, M., and Siuzdak, G. (2019). Identification of bioactive metabolites using activity metabolomics. Nat. Rev. Mol. Cell Biol. *20*, 353–367. https://doi.org/10.1038/s41580-019-0108-4.
17. Suman, S., Sharma, R.K., Kumar, V., Sinha, N., and Shukla, Y. (2018). Metabolic fingerprinting in breast cancer stages through (1)H NMR spectroscopy-based metabolomic analysis of plasma. J. Pharm. Biomed. Anal. *160*, 38–45. https://doi.org/10.1016/j.jpba.2018.07.024.
18. Jasbi, P., Wang, D., Cheng, S.L., Fei, Q., Cui, J.Y., Liu, L., Wei, Y., Raftery, D., and Gu, H. (2019). Breast cancer detection using targeted plasma metabolomics. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. *1105*, 26–37. https://doi.org/10.1016/j.jchromb.2018.11.029.
19. Yang, J., Huang, L., and Qian, K. (2022). Nanomaterials-assisted metabolic analysis toward in vitro diagnostics. Exploration (Beijing) *2*, 20210222. https://doi.org/10.1002/exp.20210222.
20. Chen, X., Shu, W., Zhao, L., and Wan, J. (2023). Advanced mass spectrometric and spectroscopic methods coupled with machine learning for in vitro diagnosis. VIEW *4*, 20220038. https://doi.org/10.1002/VIW.20220038.
21. Huang, Y., Du, S., Liu, J., Huang, W., Liu, W., Zhang, M., Li, N., Wang, R., Wu, J., Chen, W., et al. (2022). Diagnosis and prognosis of breast cancer by high-performance serum metabolic fingerprints. Proc. Natl. Acad. Sci. USA *119*, e2122245119. https://doi.org/10.1073/pnas.2122245119.
22. Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. Cancer Discov. *12*, 31–46. https://doi.org/10.1158/2159-8290.Cd-21-1059.
23. Yang, L., Venneti, S., and Nagrath, D. (2017). Glutaminolysis: A Hallmark of Cancer Metabolism. Annu. Rev. Biomed. Eng. *19*, 163–194. https://doi.org/10.1146/annurev-bioeng-071516-044546.
24. Sharma, S., Agnihotri, N., and Kumar, S. (2022). Targeting fuel pocket of cancer cell metabolism: A focus on glutaminolysis. Biochem. Pharmacol. *198*, 114943. https://doi.org/10.1016/j.bcp.2022.114943.
25. Yuan, B., Schafferer, S., Tang, Q., Scheffler, M., Nees, J., Heil, J., Schott, S., Golatta, M., Wallwiener, M., Sohn, C., et al. (2019). A plasma metabolite panel as biomarkers for early primary breast cancer detection. Int. J. Cancer *144*, 2833–2842. https://doi.org/10.1002/ijc.31996.
26. Wang, X., Zhao, X., Chou, J., Yu, J., Yang, T., Liu, L., and Zhang, F. (2018). Taurine, glutamic acid and ethylmalonic acid as important metabolites for detecting human breast cancer based on the targeted metabolomics. Cancer Biomark. *23*, 255–268. https://doi.org/10.3233/cbm-181500.
27. Liao, H.W., Chen, G.Y., Wu, M.S., Liao, W.C., Lin, C.H., and Kuo, C.H. (2017). Development of a Postcolumn Infused-Internal Standard Liquid Chromatography Mass Spectrometry Method for Quantitative Metabolomics Studies. J. Proteome Res. *16*, 1097–1104. https://doi.org/10.1021/acs.jproteome.6b01011.
28. Pascale, R.M., Calvisi, D.F., Simile, M.M., Feo, C.F., and Feo, F. (2020). The Warburg Effect 97 Years after Its Discovery. Cancers *12*, 2819. https://doi.org/10.3390/cancers12102819.
29. Tian, Y., Du, W., Cao, S., Wu, Y., Dong, N., Wang, Y., and Xu, Y. (2017). Systematic analyses of glutamine and glutamate metabolisms across different cancer types. Chin. J. Cancer *36*, 88. https://doi.org/10.1186/s40880-017-0255-y.
30. Yang, L., Moss, T., Mangala, L.S., Marini, J., Zhao, H., Wahlig, S., Armaiz-Pena, G., Jiang, D., Achreja, A., Win, J., et al. (2014). Metabolic shifts toward glutamine regulate tumor growth, invasion and bioenergetics in ovarian cancer. Mol. Syst. Biol. *10*, 728. https://doi.org/10.1002/msb.20134892.
31. van Geldermalsen, M., Wang, Q., Nagarajah, R., Marshall, A.D., Thoeng, A., Gao, D., Ritchie, W., Feng, Y., Bailey, C.G., Deng, N., et al. (2016). ASCT2/SLC1A5 controls glutamine uptake and tumour growth in triple-negative basal-like breast cancer. Oncogene *35*, 3201–3208. https://doi.org/10.1038/onc.2015.381.
32. Wetzel, T.J., Erfan, S.C., Figueroa, L.D., Wheeler, L.M., and Ananieva, E.A. (2023). Crosstalk between arginine, glutamine, and the branched chain amino acid metabolism in the tumor microenvironment. Front. Oncol. *13*, 1186539. https://doi.org/10.3389/fonc.2023.1186539.
33. Yi, H., Talmon, G., and Wang, J. (2019). Glutamate in cancers: from metabolism to signaling. J. Biomed. Res. *34*, 260–270. https://doi.org/10.7555/jbr.34.20190037.
34. Peng, S., Zhang, Y., Zhang, J., Wang, H., and Ren, B. (2011). Glutamate receptors and signal transduction in learning and memory. Mol. Biol. Rep. *38*, 453–460. https://doi.org/10.1007/s11033-010-0128-9.
35. Stepulak, A., Rola, R., Polberg, K., and Ikonomidou, C. (2014). Glutamate and its receptors in cancer. J. Neural. Transm. *121*, 933–944. https://doi.org/10.1007/s00702-014-1182-6.
36. Luksch, H., Uckermann, O., Stepulak, A., Hendruschk, S., Marzahn, J., Bastian, S., Staufner, C., Temme, A., and Ikonomidou, C. (2011). Silencing of selected glutamate receptor subunits modulates cancer growth. Anticancer Res. *31*, 3181–3192.
37. de Groot, J., and Sontheimer, H. (2011). Glutamate and the biology of gliomas. Glia *59*, 1181–1189. https://doi.org/10.1002/glia.21113.
38. Console, L., Scalise, M., Mazza, T., Pochini, L., Galluccio, M., Giangregorio, N., Tonazzi, A., and Indiveri, C. (2020). Carnitine Traffic in Cells. Link With Cancer. Front. Cell Dev. Biol. *8*, 583850. https://doi.org/10.3389/fcell.2020.583850.
39. Melone, M.A.B., Valentino, A., Margarucci, S., Galderisi, U., Giordano, A., and Peluso, G. (2018). The carnitine system and cancer metabolic plasticity. Cell Death Dis. *9*, 228. https://doi.org/10.1038/s41419-018-0313-7.
40. Console, L., Scalise, M., Giangregorio, N., Tonazzi, A., Barile, M., and Indiveri, C. (2020). The Link Between the Mitochondrial Fatty Acid Oxidation Derangement and Kidney Injury. Front. Physiol. *11*, 794. https://doi.org/10.3389/fphys.2020.00794.
41. Park, J., Shin, Y., Kim, T.H., Kim, D.H., and Lee, A. (2019). Plasma metabolites as possible biomarkers for diagnosis of breast cancer. PLoS One *14*, e0225129. https://doi.org/10.1371/journal.pone.0225129.
42. Ellis, J.K., Athersuch, T.J., Thomas, L.D.K., Teichert, F., Pérez-Trujillo, M., Svendsen, C., Spurgeon, D.J., Singh, R., Järup, L., Bundy, J.G., and Keun, H.C. (2012). Metabolic profiling detects early effects of

environmental and lifestyle exposure to cadmium in a human population. BMC Med. *10*, 61. https://doi.org/10.1186/1741-7015-10-61.

43. Sinha, R., Ahn, J., Sampson, J.N., Shi, J., Yu, G., Xiong, X., Hayes, R.B., and Goedert, J.J. (2016). Fecal Microbiota, Fecal Metabolome, and Colorectal Cancer Interrelations. PLoS One *11*, e0152126. https://doi.org/10.1371/journal.pone.0152126.
44. Mindikoglu, A.L., Opekun, A.R., Putluri, N., Devaraj, S., Sheikh-Hamad, D., Vierling, J.M., Goss, J.A., Rana, A., Sood, G.K., Jalal, P.K., et al. (2018). Unique metabolomic signature associated with hepatorenal dysfunction and mortality in cirrhosis. Transl. Res. *195*, 25–47. https://doi.org/10.1016/j.trsl.2017.12.002.
45. Mindikoglu, A.L., Coarfa, C., Opekun, A.R., Shah, V.H., Arab, J.P., Lazaridis, K.N., Putluri, N., Ambati, C.R., Robertson, M.J., Devaraj, S., et al. (2019). Metabolomic biomarkers are associated with mortality in patients with cirrhosis caused by primary biliary cholangitis or primary sclerosing cholangitis. Future Sci. OA *6*, Fso441. https://doi.org/10.2144/fsoa-2019-0124.
46. Wayne, PA: CLSI (2014). Liquid Chromatography-Mass Spectrometry Methods; Approved Guideline. CLSI document C62-A.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
serum samples	This paper	N/A
Chemicals, peptides, and recombinant proteins		
Methanol	Sigma	Cat#106035
Acetonitrile	Sigma	Cat#100029
Formic acid	Sigma	Cat#533002
Ammonia	Sigma	Cat#543830
Ammonium formate	Sigma	Cat#70221
Perfluoropentanoic acid	Sigma	Cat#68542
Ammonium bicarbonate	Sigma	Cat#533005
Deposited data		
Results for mass spectrometry of metabolomics	This paper	MetaboLights: MTBLS10205
Software and algorithms		
SPSS 26.0	IBM	N/A
R language, 3.5.2	The Comprehensive R Archive Network	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jun Zhang (jameszhang2000@zju.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The polar metabolomic data of our cohorts are provided in Supplemental information, Tables S11–S15. Raw metabolomics data derived from human samples have been deposited at Metabolights (www.ebi.ac.uk/metabolights/MTBLS10205), and accession numbers are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Patients and cohorts

A total of 1947 participants, including patients with breast cancer (BC), breast benign diseases (BENIGN), lung cancer (LC), thyroid cancer (TC), gastric cancer (GC), colorectal cancer (CC), rectal cancer (RC), liver cancer (LIC), and pancreatic cancer (PC), as well as healthy controls (HC), were recruited from four hospitals from January 2018 to December 2022. In the screening and training sets, 496 serum samples were collected from Sir Run Run Shaw Hospital, Zhejiang University School of Medicine (Hangzhou, China). Validation cohort 1 included 1192 participants enrolled from Sir Run Run Shaw Hospital, Zhejiang University School of Medicine (Hangzhou, China), Ningbo Medical Center Lihuili Hospital (Ningbo, China), and Sir Run Run Shaw Hospital Xiasha Campus (Hangzhou, China). Independent validation cohort 2 comprised 259 participants enrolled from the Affiliated Hangzhou First People's Hospital, Westlake University School of Medicine (Hangzhou, China) to further compare the diagnostic performance between metabolite model and ultrasound (Figure 1). Clinical characteristics, including age, body mass index (BMI), menstrual history, and tumor markers (CA15-3 and CEA), were collected from each participant. Tumor pathology, such as molecular type and tumor node metastasis (TNM) stage, were also recorded for BC patients. All BC, BENIGN, and HC subjects were

female. Detailed demographic and clinical characteristics of all participants were summarized in Table S1. All participants provided signed informed consent. The study was conducted in compliance with the Helsinki Declaration (revised in 2013). The study protocol was approved by the Ethics Committee of each clinical center (20180601-006).

Physical examination and laboratory test results of HC were normal. All BC and BENIGN patients were enrolled based on the diagnosis of "breast lesions", as determined by ultrasound or mammography, and confirmed by pathology. Samples were collected on the day of admission or the morning of the second day. The median period between sample acquisition and final diagnosis based on pathological examination was seven days. No treatment (surgery, chemotherapy, or radiotherapy) was administered before sample acquisition. Exclusion criteria included the following: (1) history of breast disease or tumor for HC; (2) history of other tumor besides breast cancer for BC; (3) history of systemic chronic diseases, such as hypertension or diabetes, metabolism-related diseases, such as phenylketonuria or hepatic encephalopathy, and mental illness; (4) female individual in menstruation, pregnancy, or lactation.

METHOD DETAILS

Serum sample collection and pretreatment

Peripheral blood samples were collected from each participant in the morning after overnight fasting. The samples were centrifuged at 3000 rpm for 10 min. The supernatant (serum) was collected and immediately stored at −80°C until further analysis. For non-targeted ultrahigh performance liquid chromatography-tandem mass spectroscopy (UPLC-MS/MS) analysis, samples were prepared using the automated MicroLab STAR system from Hamilton Company. Proteins were precipitated with methanol under vigorous shaking for 2 min followed by centrifugation. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase (RP)-UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by RP-UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. Samples were placed briefly on a TurboVap (Zymark) to remove the organic solvent. The sample extracts were stored overnight under nitrogen before preparation for analysis. For targeted LC-MS/MS, 100 μL of each thawed serum sample were extracted by adding 300 μL methanol, vortexed and mixed for 5 min, and then centrifuged at 4000 rpm for 10 min. The supernatant was stored at 4°C for further use.

Non-targeted UPLC-MS/MS metabolomics analysis

UPLC-MS/MS analysis was performed by the Metabolon-Dian Joint Metabolomics Laboratory (Hangzhou, China). Briefly, the UPLC-MS/MS system consisted of the Waters ACQUITY UPLC and Thermo Scientific Q-Exactive high-resolution mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. UPLC was performed on a C18 column (Waters UPLC BEH C18 2.1 × 100 mm, 1.7 μm) and hydrophilic interaction liquid chromatography column (Waters UPLC BEH Amide 2.1 × 150 mm, 1.7 μm), respectively, and linked to the mass spectrometer operated under positive or negative ESI mode. Quality control (QC) process was applied in the non-targeted metabolomic analysis which included using multiple internal standards and a pooled quality control (PQC) sample. The PQC sample was prepared by mixing together a small volume of each experimental sample. The PQC sample were inserted into the experimental samples periodically throughout the metabolomics run. The signals of the internal standards in both PQC samples and experimental samples were monitored to check overall process and platform variability. The MS analysis alternated between MS and data-dependent MS^n scans using dynamic exclusion. The scan range varied slighted between methods but covered 70–1000 m/z.

Raw data was extracted, peak-identified and QC processed using Metabolon's hardware and software. Compounds were identified by comparison to a self-built library. Metabolon maintains a library based on authenticated standards that contains the retention time/index (RI), mass to charge ratio (m/z), and chromatographic data (including MS/MS spectral data) on all molecules present in the library. Furthermore, biochemical identifications are based on three criteria: retention index within a narrow RI window of the proposed identification, accurate mass match to the library +/− 10 ppm, and the MS/MS forward and reverse scores between the experimental data and authentic standards.

Peaks were quantified using area under the curve. Each compound was corrected in run-day blocks by setting the median equal to 1 and normalizing each data point proportionately. Then, the data were log transformed for statistical analysis. In the univariate analysis phase, fold change threshold (FC) of each metabolite was calculated, and Welch's t test or one-way ANOVA analysis was applied to measure the differences in metabolites between groups. In the multivariate analysis stage, partial least squares discriminant analysis (PLS-DA) score plot was used to show the overall differences among groups.

Targeted LC-MS/MS methods development and validation

Targeted LC-MS/MS analysis was established on the AB SCIEX 6500+ Triple Quadrupole mass spectrometer coupled to an ultrahigh performance liquid chromatography system (AB SCIEX, USA) to obtain the absolute quantities of candidate metabolite biomarkers. Glutamate, N^2,N^2-dimethylguanosine, and propionylcarnitine were quantified by the internal standard method. Erythronate, docosahexaenoate, and N-Acetylneuraminic Acid were quantified by the external standard method. Docosahexaenoate, N-Acetylneuraminic Acid, propionylcarnitine, and N^2,N^2-dimethylguanosine were assayed in Method A using an Agilent Eclipse XDB 4.6 × 150 mm C18 column, 5 μm. Erythronate was assayed in Method B using the Waters ACQUITY UPLC BEH 2.1 × 150 mm Amide column, 1.7 μm. Glutamate was assayed in Method C using an Agilent Eclipse XDB 4.6 × 150 mm C18 column, 5 μm, which used a separate internal standard and required a 10-fold dilution of the

serum samples before running. The chromatography and mass spectrometer programs, along with instrument parameters corresponding to the three methods, were specified in Tables S6 and S7.

According to the Clinical and Laboratory Standards Institute (CLSI) C62-A document,[46] the developed targeted LC-MS/MS methods were further validated for linearity, limit of quantification, precision, accuracy, and stability.

Linearity and limit of quantification

A mixed standard solution of five metabolites (docosahexaenoate, N-Acetylneuraminic Acid, propionylcarnitine, erythronate, and N^2,N^2-dimethylguanosine) was serially diluted 5×, 10×, 20×, 50×, 100×, 200×, and 500× to construct calibration curve A. For the glutamate assay, a 50 μg/mL standard solution was further diluted to 0.1, 0.2, 0.5, 1, 2, 5, and 10 μg/mL to construct calibration curve B. Each concentration of the standard curve was measured three times, and linearity was evaluated by calculating the regression coefficient r^2 of the linear regression curve and deviation between measured and theoretical values (Bias%). The limit of quantification (LOQ) was determined based on a signal/noise (S/N) ratio of 10:1, and the LOQ was also determined with acceptance criteria of ±20% precision and accuracy.

Precision and accuracy

To determine precision and accuracy, quality control (QC) samples at three concentration levels (LQC, MQC, HQC, Table S8) were analyzed. QC samples at these three levels were analyzed in six replicates on the same day and over three consecutive days, respectively. Accuracy is expressed as relative bias (Bias%), which represents the deviation between the average of measured values and theoretical values. Precision was expressed in RSD (CV%) of experimentally measured concentrations. Intra-batch precision was the RSD of six replicates of each concentration on the same day. Inter-batch precision was the RSD of the daily average of six replicates over three days.

Stability

Stability of the six target metabolites was evaluated using pooled serum samples under different storage or processing conditions. For short-term stability, pooled serum samples were stored at room temperature for 6 h, 4°C for 1 day and 4°C for 3 days. For mid-term stability, pooled serum samples were stored at −20°C for 1 week and −20°C for 2 weeks. Freeze–thaw stability of the pooled serum samples was assessed for two cycles. For each freeze-thaw cycle, samples were frozen at −20°C overnight and subsequently thawed at room temperature. The average values of 6 repeated tests of fresh pooled serum samples (stored at room temperature for 0 h) were used as the theoretical values. The measured values of pooled serum samples were compared with the corresponding theoretical values, and the CVs% under each storage or processing condition should be less than 15%.

QUANTIFICATION AND STATISTICAL ANALYSIS

For demographic characteristics, numerical variables were expressed as mean ± standard deviation (mean ± SD), and categorical variables were expressed as percentages. Comparisons between numerical variables were performed using Student's t test, Mann–Whitney U test, or one-way ANOVA analysis. Categorical variables were analyzed using chi-square test. For non-targeted metabolomics analysis, Welch's t test or Mann–Whitney U test was applied to measure the differences in metabolite level among groups. Random forest analysis (RF) was conducted to identify important metabolites according to the Mean Decrease Accuracy plot. Binary logistic regression analysis was used to further select biomarkers and build the metabolite model. The receiver-operating characteristic (ROC) curve was constructed to determine the area under the curve (AUC) of the model. These statistical analyses were performed using SPSS 26.0 (Statistical Product and Service Solutions, IBM, USA) and R language, version 3.5.2 (R Foundation for Statistical Computing, Austria). The original *p*-values of multiple tests were adjusted by FDR (Benjamin–Hochberg). P-values were two-sided, and statistical significance was achieved when $p < 0.05$.