Novel Metabolic Biomarker for Early Detection and Prognosis to the Patients with Gastric Cardia Adnocarcinoma

Mengxia Wei
Zhengzhou University First Affiliated Hospital
Xueke Zhao
Zhengzhou University First Affiliated Hospital
Panpan Wang
Zhengzhou University First Affiliated Hospital
Xin Song
Zhengzhou University First Affiliated Hospital
Jingfeng Hu
Zhengzhou University First Affiliated Hospital
Kan Zhong
Zhengzhou University First Affiliated Hospital
Lingling Lei
Zhengzhou University First Affiliated Hospital
Ruihua Xu
Zhengzhou University First Affiliated Hospital
Wenli Han
Zhengzhou University First Affiliated Hospital
Miaomiao Yang
Zhengzhou University First Affiliated Hospital
Xuena Han
Zhengzhou University First Affiliated Hospital
Zongmin Fan
Zhengzhou University First Affiliated Hospital
Ran Wang
Zhengzhou University First Affiliated Hospital
Bei Li
Zhengzhou University First Affiliated Hospital
Jiajia Ji
Zhengzhou University First Affiliated Hospital
Liuyu Li

Zhengzhou University First Affiliated Hospital
Yao Chen
Zhengzhou University First Affiliated Hospital
Yuanze Yang
Zhengzhou University First Affiliated Hospital
Keke Zhao
Zhengzhou University First Affiliated Hospital
Fuyou Zhou
Anyang Tumor Hospital
Aili Li
Linzhou Tumor Hostipal
Zhiqiang Li
Huixian Mengtang Hostipal
Qide Bao
Anyang District Hospital
Li Dong Wang (ldwang@zzu.edu.cn)
Zhengzhou University First Affiliated Hospital https://orcid.org/0000-0002-7933-0410

Research

Keywords: Gastric cardia adenocarcinoma (GCA), Nontargeted metabolomics, Prognosis, Arginine biosynthesis

Posted Date: October 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-960379/v1

Novel metabolic biomarker for early detection and prognosis to the patients with gastric cardia adnocarcinoma

Running tittle: Novel metabolic biomarker in gastric cardia adnocarcinoma

Mengxia Wei[1], Xueke Zhao[1], Panpan Wang[1], Xin Song[1], Jingfeng Hu[1], Kan Zhong[1], Lingling Lei[1], Ruihua Xu[1], Wenli Han[1], Miaomiao Yang[1], Xuena Han[1], Zongmin Fan[1], Ran Wang[1], Bei Li[1], Jiajia Ji[1], Liuyu Li[1], Yao Chen[1], Yuanze Yang[1], Keke Zhao[1], Fuyou Zhou[2], Aili Li[3], Zhiqiang Li[4], Qide Bao[5], and Lidong Wang[1]

*Correspondence : ldwang@zzu.edu.cn

Author contributions: Conception and design: Li-Dong Wang; Collection and analysis of data: Meng-Xia Wei, Pan-Pan Wang; Manuscript drafting: Li-Dong Wang, Meng-Xia Wei; Final improvement: All authors.

Abstract

Background: Gastric cardia adenocarcinoma (GCA), which has been normalized as type II of adenocarcinoma at esophagogastric junction in western countries. In clinical, most of the GCA patients are lack of early alarming symptoms, more than 90% of GCA patients were diagnosed at advanced stage, resulted in a very poor prognosis, with less than 20% of 5-year survival. Obviously, early detection for GCA plays crucial role in decreasing the high mortality. Metabolomics allows for appraisal of small molecular mass compounds in a biofluid, which may provide a way for finding biomarkers for GCA.

Methods: The serum metabolic features of 276 curatively resected GCA patients and 588 healthy control participates were collected from the database of State Key Laboratory of Esophageal Cancer Prevention & Treatment and Henan Key Laboratory for Esophageal Cancer Research of The First Affiliated Hospital of Zhengzhou University to discover the metabolic dysregulation by using the ultraperformance liquid chromatography-mass spectrometry (UPLC-MS). Joint pathway analysis with metabolites identified, survival analysis and auxiliary diagnosis metabolites were discussed in present work.

Results: A sum of 200 known differential metabolites were obtained with $p < 0.05$ and fold change $FC \geq 1.25$ or $FC \leq 0.8$ by comparison GCA and healthy control participates. 12 metabolites significant correlated with survival ($p < 0.05$) and 17 metabolites for potential auxiliary diagnosis($FC > 1.5$ or $FC < 0.67$) for GCA. Dysregulated arginine biosynthesis was an important pathway of GCA. 9 differential metabolites of 12-ketolithocholic acid, 2-Hydroxybutanoic acid, Aldosterone, All-trans-13,14-dihydroretinol, Hododeoxycholic acid, L-histidine, Malonic acid, Prostaglandin E2 and Sphingosine were identified as potential metabolic markers for distinguishing the GCA and healthy control (AUC=0.976, sensitivity =0.913, specificity =0.027, optimal cut off value=0.470).

Conclusions: This work was first identified 12 metabolites significant correlated with survival and 17 metabolites for potential auxiliary diagnosis for GCA. In addition, arginine biosynthesis pathway metabolism showed important roles in GCA. Results provide the understanding of the molecular difference between GCA and healthy control. The novel plasma biomarkers panel could clearly distinguish GCA patients from the healthy control group. This finding may form the basis for the development of a minimally invasive method for GCA detection.

Keywords: Gastric cardia adenocarcinoma (GCA), Nontargeted metabolomics, Prognosis, Arginine biosynthesis

Introduction

Endoscopy with iodine staining was widely used for gastric cardia cancer (GCA) and esophageal cancer (EC) screening in high-incidence area. Most endoscopy screening-positive population was found to develop esophageal epithelium lesion, and therefore endured higher risk for developing gastric cardia cancer (GCA) and EC than normal population. However, endoscopic screening may be too costly and invasive for large-scale population, and non-invasive biomarkers may be more applicable and cost effective for population-based screening. In this population-based screening study, we aim to identify potential metabolic biomarkers for early screening of GCA, and establish the optimal early GCA screening model.

Gastric cardia adenocarcinoma (GCA) is a cancer which occurs in the gastric cardia area (gastric-esophageal boundary), that originates or mainly occupies within 2 cm of the esophagus and gastric mucosa junction line[1], and it is one of the common malignant tumor of the digestive tract in China. GCA morbidity and mortality have increased in recent years[2].GCA and esophageal squamous cell carcinoma (ESCC) are two common gastrointestinal tumors,and have been called sister cancer owing to their similar characteristics, including to the adjacent anatomical locations, and simultaneously occurrence in clinical practice[3]. And it suggests that they may have similar prognostic molecules mechanism. Early diagnosis, prevention and treatment are the keymeans to reducing the incidence and mortality of GCA, and it is particularly important to find non-injury (serum) early detection indicators.

The abnormal metabolism of cancer has been considered an important characteristic of tumors, which could clarify the pathogenesis and provide potential therapeutic targets for clinical treatments (3). According to the Warburg effect, the deregulated energy metabolism of cancer cells may also modify many related metabolic pathways that influence various biological processes, such as cell proliferation and apoptosis. As a common characteristic of cancer cells (4, 5), modified metabolism has been the focus of cancer research.

Metabolomics was first proposed by Nicholson Lindon and Holmes in 1999. Metabolomics is a newly developed discipline after genomics, and it is an important part of systems biology[4, 5]. Metabolomics is a new discipline that simultaneously conducts qualitative and quantitative analysis of all low molecular weight (relative molecular weight less than 1000 Da) metabolites of a certain organism or cell in a specific physiological and

pathological period. The object of metabolomics research is the endogenous metabolites of small molecules in the body. Through high-throughput qualitative and quantitative analysis, we can understand the changes in metabolites in the body. Changes in any physiological, pathological or other factors in the body will cause its metabolic level has changed. Through metabolomics research, the impact of various factors on the body can be analyzed at the overall level, which truly reflects the impact of the body[6]. The uniqueness of metabolomics is that it can reflect the changes currently taking place in the body in real time. Therefore, metabolomics is a powerful tool to reveal the dynamic changes closest to the phenotype[7]. Small changes in genes and proteins can be amplified at the metabolic level. Potential biomarkers related to the course of the disease can be screened based on differences in substance types and different concentrations of the same substance. In the past few decades, metabolomics has become the study of metabolic processes, identifying potential biomarkers and revealing A powerful tool for metabolic changes in various diseases [8-11].The screening biomarkers play an important role in measuring disease status[12], monitoring drug sensitivity[13], and physiological status[14].

Metabolic changes in tumors are usually manifested as: (1)Enhancement of aerobic glycolysis process, leading to glucose carbon being directed to lactic acid and nuclmeotides; (2)Anabolism enhancement of amino acids and lipids, which related to the interruption of the tricarboxylic acid cycle and the increased use of glutamine as a carbon source (increased glutamine uptake and catabolism); (3)Enhancement of the tumor-induced pentose phosphate pathway which increased the cycle of reduced coenzyme II to protect cell from oxidative stress; (4)Consumption and uptake of glucose.

In this present work, we characterized the metabolic features of GCA using a nontargeted metabolic profiling strategy based on liquid chromatography-mass spectrometry metabolomics analysis, and a two-phase biomarker development strategy (discovery set and validation set) was applied in 864 subjects including to clinically relevant controls, and univariate statistical analysis and multivariate analysis (MVA) methods were used to identify differential metabolites. The serum of 276 curatively resected GCA patients and 588 healthy people were collected to discover metabolic dysregulation, and a technique was used to establish a novel diagnostic tool. Joint pathway analysis with metabolites identified relevant metabolic pathways and detection biomarkers for GCA.

Materials and methods

Study participates

In this study, 276 GCA patients and 588 non-GCA participants (control) were collected. The GCA patients were diagnosed between 1987 to 2020 and followed up to 2020. We divided the study cohort into training set (138 GCA participates and 363 control participates) and validation set (138 GCA participates and 225 control participates), respectively.Entry criteria for the experimental participates: ①The pathological type was clearly diagnosed by histopathology; ②The pathological type was cardia adenocarcinoma; ③The age of GCA participates were from18 to90 years old; ④ None of the participates had received chemotherapy, radiotherapy and other anti-tumor treatments before the examination; ⑤ Clinical stage T1-4N1M0/T4N0M0 (stage IIIA or IIIB), no distant metastasis; ⑥No other malignant tumors; ⑦No other metabolic diseases; ⑧No symptoms of co-infection, such as fever, elevated blood picture, etc. Conditions for entry into the control group: ①18 to 90 years old; ②Without malignant tumors or other metabolic diseases.

The serum metabolites of GCA and the non-GCA (control) participants were explored through metabolomics strategy based on ultra-performance high liquid chromatography-mass spectrometry (UPLC-MS)(QTRAP®, https://sciex.com/). Differential metabolites, metabolic pathways and correlation networks (KEGG) were investigated and their potential for use in GCA early detection were investigated.

For the metabolites qualitative and relative quantitative analysis, UPLC-MS was used, and the chromatographic grade acetonitrile and formic acid were purchased from Merk during the measurement process. Ultra-water was filtered by Mili-Q system. Analytical metabolites were purchased from Sigma-Aldrich.

Metabolomics preparation and analysis

Blood sample

Serum collection in the morning on an empty stomach, peripheral venous blood was collected at 7:00 am; blood samples were collected in a centrifuge tube, and left standing at 37°C or room temperature for 1 hour for stratification; centrifuged at 3,000 r/min (r=8 cm) at room temperature for 10 min, and taken Transfer the supernatant to a clean centrifuge tube; centrifuge at 12,000 r/min (r=8 cm) at 4 ℃ for 10 min, transfer the supernatant to 1.5 mL centrifuge tubes, 0.2 mL per tube. Store in the refrigerator at -80 ℃, and transport in dry ice.

The GCA and non-GCA (control) blood samples were prepared on the ice in subsequent

follow-up treatment. Firstly, vortexed for 10s for sufficient mixing, after taking out from -80°C refrigerator and thoroughly dissolving out. Secondly, 300 μL chromatographic grade methanol was added into 50 μL blood sample in numbered centrifuge tube, and then vortexed for 3 min for adequately dissolution of metabolites. The extraction blood samples were centrifuged 10 min with the rate of 12,000 r/min. Thirdly, the 200 μL supernatant was separated into another numbered centrifuge tube, and then centrifuged 3 min with the rate of 12,000 r/min after 30 min 4°C after stored 30 min -20°C. Finally, 150 μL supernatant was analyzed by UPLC-MS.

UPLC-MS condition

The serum metabolites compounds were analyzed by ultraperformance liquid chromatography (ExionLC AD, https://sciex.com.cn/)-mass spectrometry(https://sciex.com/) (SCIEX, USA), which were included section 1 of HPLC conditions and section 2 of ESI-QTRAP-MS/MS.

Section 1: Mobile phase A was ultra-water (0.1% chromatographic formic acid), and mobile phase B was chromatographic acetonitrile (0.1% chromatographic formic acid). 2 μL of the 15 μL supernatant mentioned above was injected into the detector and separated by Water ACQUTTY UPLC HSS T3 C18 chromatographic column (1.8 μm, 2.1 mm ×100 mm) in 40 °C with the flow rate of 0.4 mL/min. The phase A elution condition decreased from 95% to 10% from 0 min to 10 min, and then changed to 95% in 11.1 min immediately and maintaining 3 min.

Section 2: The analysis was performed in positive and negative ion mode. ESI source parameters were as follows: capillary voltage-4,500 V (negative) and 5,500 V (positive); capillary temperature 500 °C; Sheath and Auxiliary Gas flow (N2), 55 psi (GS I) and 60 psi (GS II); Sweep gas 25 psi; Instrument tuning and mass calibration were performed with 10 and 100 μmol/L polypropylene glycol solutions in QQQ and LIT modes, respectively. A specific set of MRM transitions were monitored for each period according to the metabolites eluted within this period.

Xcalibur software version 2.1 was used for instrument control, data acquisition and data analysis.

Analysis method

Metabolomics analysis

Unsupervised PCA (principal component analysis) was performed by statistics function prcomp within R(www.r-project.org). The data was unit variance scaled before unsupervised PCA.The data set was analysed with pattern-recognition methodsusing the software package Simca version 14.1 (UmetricsAB, Umeå, Sweden). The response variables were centred and scaled to Pareto variance, and the base weight was computed as 1/square root (standard deviation of the response variables). Moreover, to normalize the skewed distributions,log transformations were used for non-linear conversions of the data. To eliminate the effect of inter-subject variability and to identify endogenous metabolites that contributed significantly to the classification, linear combinations of X variables orthogonal to the Y vector were removed by orthogonal projections to latent structures (OPLS)-discriminant analysis (DA)[15, 16].

Hierarchical cluster analysis with TNM stage

The HCA (hierarchical cluster analysis) results of samples and metabolites were presented as heatmaps with dendrograms and carried out by R package Complex Heatmap. For HCA, normalized signal intensities of metabolites (unit variance scaling) are visualized as a color spectrum.

The differential metabolites of interest were screened using the following rules: features that meet the following conditions are removed[17]. (1) The t-test results, fold change (FC) and p value of the QC samples of all features were $0.67 < FC < 1.5$, respectively; (2) The isotopic ions of the features; (3) Some different adduct ions which represent the same metabolite, and the adduct ion with the highest intensity compared with other adduct ions should be retained; (4) The features with such a low intensity that lack the MS/MS fragments information supplied by the information dependent acquisition (IDA) explorer function in the PeakView 2.2 software. (5) The feature with such a poor peak shape that led to inaccurate quantification. In the analysis results of the lower phase, the differential metabolites of interest screened in the negative mode could also be found in the positive mode with relatively high intensity. The metabolites of the upper phase in the negative mode were mainly some fatty acids without differential results. From the above description, 17 differential metabolites of interest were screened out including to 1,7-Dimethylxanthine, 18β-Glycyrrhetinic acid, 3-Indolepropionic Acid, 4-Hydroxy-3-methoxybenzaldehyde, 4-Hydroxytryptamine, 5(S),15(S)-DiHETE [5S,15S-dihydroxy-6E,8Z,10Z,13E-eicosatetraenoic acid], Bis(1-inositol) -3,1'-phosphate 1-phosphate, Coniferyl acetate,

Ginkgoic acid, Glycine deoxycholic acid, Leu-Val, L-Valyl-L-phenylalanine, Prostaglandin E2, Pyridoxal, Theophylline, Trans-3-Hydroxycotinine, Valyl-leucine.

Differential metabolites selected

Significantly regulated metabolites between groups were determined by absolute Log2FC (fold change) ≥ 1. The data was log transform (log2) and meancentering before OPLS-DA. In order to avoid overfitting, a permutation test (200 permutations) was performed.Establishing a diagnostic model to predict the presence of GCA through the distinct metabolic profile was difficult because of the high-dimensional dataset. To improve the prediction of cachexia, random forest analysis was performed to determine whether the metabolic data could successfully differentiate the two groups. Then, a forward stepwise logistic regression model was constructed on the training sample set to design the best metabolite combination. Receiver Operating characteristic (ROC) curves were used to evaluate the accuracy of this model in the validation sets following the DeLong method. The global performance of each biomarker model was evaluated using the Area under the curve (AUC) and the determination of sensitivity and specificity at the optimal cut-off point defined by the minimum distance to the top-left corner.

KEGG annotation and enrichment analysis

Identified metabolites were annotated using KEGG compound database (http://www.kegg.jp/kegg/compound/), annotated metabolites were then mapped to KEGG pathway database (http://www.kegg.jp/kegg/pathway.html). Pathways with significantly regulated metabolites mapped to were then fed into MSEA (metabolite sets enrichment analysis), their significance was determined by hypergeometrict test's p-values.

Follow-up

The follow-up of this present study wasmainly carried out through telephone, home visit and direct contact between village doctors and patients or patients' families, or through system query methods such as new cooperative medical database, Medical Security Bureau database and citizen death information registration management. The follow-up is carried out once a year to record the time and main cause of death. This study was follow-up until January 25, 2021.

Data analysis

The raw data were aligned, deconvoluted, and normalized (sum of total area) using the MarkerView™ Software 1.3 (SCIEX). The retention time (RT) was from 0.5 min to 14 min.

The mass and RT tolerance values were set to 10 ppm and 0.15 minutes, respectively. After the 80% rule was used to treat the missing values for each sample group, a list of the intensities for each detected peak was generated, using retention time and the mass-to-charge (m/z) ratio data pairs as the parameters for each ion. Thus, each spectral feature was represented by a unique m/z, retention time, and peak area. After the data preprocessing, the resulting 2-dimensional data matrix (pareto scaled) was subjected to MVA using the SIMCA-P software (version 14.1, Umetrics AB, Umea, Sweden). Principal component analysis (PCA) was used to visualize system stability of the system and sample distribution. The orthogonal partial least squares discriminant analysis (OPLS-DA) was used to identify the variables responsible for the discrimination. The "goodness of fit" and predictive power of the model were evaluated using R^2Y (sum) and Q^2 (sum), respectively. A 200-times permutation test was performed to evaluate the risk of model overfitting. Additionally, the Student's t-test was applied to measure the significance of each variable. To remove any p-values (up to a 95 % confidence) that could have been false positives, the resultant p values for each metabolite were corrected by Bonferroni correction. Volcano plot, S-plot, and Venn diagram depictions were used to filter important variables that displayed statistical significance (adjusted $p < 0.05$), significant fold changes (FC $\geqslant 1.25$ or $\leqslant 0.8$) between the two groups. The Formula Finder algorithm was used to identify potential differential metabolites and generate a group of probable formulas on an unknown ion based on the secondary fragment information, mass error, and isotope distribution patterns. Subsequently, the HMDB, METLIN, MoNA, and KEGG databases were browsed for these candidates, and the final decisions about possible structures and final biomarkers were based on the obtained MS/MS spectra. Statistical analyses were performed using the SPSS software version 21.0 (IBM Corp., Armonk, New York). By step wise regression analysis, the factors with significant influence were selected as independent variables, and an “optimal” regression equation was established to find potential biomarkers for distinguishing between the GCA group and control groups. Logistic regression analysis and receiver operating characteristic (ROC) analysis were used for the diagnosis of ESCC and HC. The area under the receiver-operating characteristic curves (AUROC) was calculated by SPSS to evaluate the predictive performance of the constructed signatures in both the training and validation sets. The Youden index (J) was used as the best threshold to select the optimal cut-point that maximized its value[22]. A heat map of the identified key metabolites was drawn by the pheatmap package (R version 3.3.0). Open

database sources, including the KEGG and MetaboAnalyst, were used to identify metabolic pathways. By this analysis, several dysregulated metabolic pathways associated with the development of ESCC were uncovered.

Results

Demographic characteristics of the study population

A total of 864 participants, consisting of 588 healthy volunteers, 276 patients with GCA from 2 independent cohorts were recruited (Fig. 1A). The discovery set included 363 healthy volunteers, 138 patients with GCA, and the validation set included 255 healthy volunteers and 138 patients with GCA. The GCA participants were enrolled without any chemoradiotherapy record, and healthy participates without any history of esophagus and cardia diseases.More than half of the discovery set GCA participates were collected from the Anyang cancer hospital, and most of the others were collected from the Henan regional hospitals, and the validation cohort of 138 GCA patients were collected from the Anyang cancer hospital, Linzhou people's hospital and Linzhou cancer hospital, with the detail information showing in supplementary TableS1. Written informed consent was obtained from each participant. The study was approved by the ethics committee of each cohort. The diagnosis was confirmed with histopathology.

The univariate analysis was carried out with the categorical and continuous variables of gender, age, family history, city/village, low/high incidence area, smoking, drinking, family history, BMI stage, blood group, degree of differentiation, lymph node metastasis, T stage, TNM stage, outcome variables of GCA and healthy controls subjects. And the clinical characteristics of the participants with the univariate analysis results are presented in Table 1. As we observed in the discovery group that the BMI stage is the important factor for the cancerous of cardia, and the age, high/low incidence area, city/village, smoking, TNM stage in the validation group.

Differential metabolites between GCA and DNT

The principal component analysis of this phase showed differences among samples from normal healthy controls and GCA participants (Fig. 1B), and no significant outliers were observed, which indicated that the stability of the analysis data is good.A PCA was performed on all plasma sample data. The best separation of groups was obtained in the principal components (PC) 1 and 2, which accounted for 15.3% and 6.0% of the whole variance of the data set, respectively.The two sets showed a major overlap but samples from GCA patients had a tendency towards lower scores in PC1, which was remarkable for a heterogeneous cohort with high inter individual variability due to diverse lifestyles, medications and comorbidities. In the PCA obtained in the second validation study, GCA

patients were added as another diagnosis group in addition to the test set (Fig. 3B). The best separation between the GCA and control groups was again observed in PC1 and PC2 (25.3% and 8.1% of the observed variance), which means that the special metabolites could clearly separate GCA from normal. Remarkably, an almost complete separation of the control group from the other two could be observed. The GCA patients tended to have higher scores in the PC2 than the liver cirrhosis patients, resulting in a visible separation between these groups.

We conducted test set and validation set simultaneously, and they were done this to show that the metabolomic profiles and distribution of the GCA versus control group patients in the test and validation set. And in conclusion, we observed that the similar separation between the two sets, which demonstrated that the two sets were actually comparable, although conducted independently. The two PCA of the test set and validation set were a good overview on how the metabolic profile of the GCA patients relate the metabolic profiles of GCA patients and controls.

Subsequently, the orthogonal partial least squares discriminant analysis shows significant separations of healthy from GCA participants, GCA participates were distributed in negative guardant, and the healthy control were distributed in positive guardant, which showed obvious difference between GCA and healthy control.The R2Y and Q2 were significant fitting parameters which showed the fitting effect of OPLS-DA model.R2Y was indicated the ratio of the explainable variation to the total variation in the OPLS-DA model, and the Q2 refers to the ratio of the predictable variation in the OPLS-DA model to the total variation.And the R2Y and Q2 of the OPLS-DA model were 0.349 and -0.303, which indicated that 34.9% of variation for the OPLS-DA model is explainable, and 30.3% of variation for the OPLS-DA model can be explained. In conclusion, the metabolic profile existed significant difference between GCA and healthy control.

The representative total chromatograms for healthy and GCA participants are shown in Fig. 1C. A total of 903 metabolites were detected in both positive and negative electrospray ionizaiton (ESI+ and ESI-) modes with 788 and 801 metabolites for discovery set and validation set, respectively. The differential metabolites biomarker candidates were satisfied with the criterion of $p < 0.05$ and (FC) ≥ 1.25 or FC ≤ 0.8. Hence, 501, 609 differential metabolites were identified from the sets of healthy vs GCA for discovery set and validation set, respectively, with 34 and 166 differential metabolites for upregulated and down regulated, respectively. Importantly, 25differential metabolites (showed in Table 2) including to 11up-

regulated and 14down-regulated metabolites were significantly altered in the two sets. 11up-regulated metabolites were 4-Guanidinobutyric Acid, N1-Acetylspermine, Acetylcholine, All-Trans-13,14-Dihydroretinol, L-2-Aminobutyric acid, 2-Hydroxybutanoic Acid, Capric Acid(C10:0), 19(S)-HETE, 4-Pyridoxic Acid, L-Fucose, Malonic acid, 3-Hydroxybutyrate. And 14down-regulated metabolites were 12,13-DiHOME, Aldosterone, 3-Indolepropionic Acid, Glycochenodeoxycholic Acid, L-Proline, Glycine deoxycholic acid, Trans-3-Hydroxycotinine, Prostaglandin E2, 7-ketolithocholic acid, 12-ketolithocholic acid, Hododeoxycholic acid, L-Histidine, Sphingosine, 4-Hydroxytryptamine.

Metabolomic profile of matched early and middle-late stage of GCA

Heatmapof serum metabolomic profilesshowed in Fig. 2,which indicated that each metabolite was normalized by dividing by the average of control samples. Samples colored in red to white to blue scheme indicate relatively higher, average, and lower concentrations, respectively. Only metabolites showing a significant difference in FC>1.5or FC<0.67were used. These metabolites were clustered using Pearson correlation. 17 metabolites showed significantly higher average concentrations consistently in CCA serum. (Raw data showed in Supplementary Tables S2 for discovery set and validation set). Moreover, there were significant difference between the early and middle-late stage group in different pathological stage.

Established the biomarker diagnostic model

From the 200 metabolites, 12 down-regulated differential metabolites significantly related with survival were selected, including to 9,10-DiHOME [(±)9,10-dihydroxy-12Z-octadecenoic acid], 12,13-DiHOME, Aldosterone, 9-HpODE, 12,13-EpOME [(±)12(13)epoxy-9Z-octadecenoic acid], 9,10-EpOME [(±)9,10-epoxy-12Z-octadecenoic acid], Lysope 18:2 (2N Isomer), 2-Hydroxycaprylic acid, 23-deoxydeoxycholic acid, 7-ketolithocholic acid, 12-ketolithocholic acid, RvE1 [5S,12R,18R-trihydroxy-6Z,8E,10E,14Z,16E-eicosapentaenoic acid]. And we found that the 6/12 were oxidized lipids, including to the 9,10-DiHOME [(±)9,10-dihydroxy-12Z-octadecenoic acid], 12,13-DiHOME, 9-HpODE, 12,13-EpOME [(±)12(13)epoxy-9Z-octadecenoic acid], 9,10-EpOME [(±)9,10-epoxy-12Z-octadecenoic acid], RvE1 [5S,12R,18R-trihydroxy-6Z,8E,10E,14Z,16E-eicosapentaenoic acid]. 3/12 differential metabolites were bile acids, and 1/12 differential metabolite (aldosterone) was ketones, and 1/12 differential metabolite (Lysope 18:2 (2N Isomer) was lysophatidylethanolamine, and 1/12 differential metabolite (2-Hydroxycaprylic

acid) was organic acid and its derivatives.

Based on the binary logistic regression model, 9 different metabolites of 12-ketolithocholic acid, 2-Hydroxybutanoic acid, Aldosterone, All-trans-13,14-dihydroretinol, Hododeoxycholic acid, L-histidine, Malonic acid, Prostaglandin E2, and Sphingosine were identified as potential metabolic markers for distinguishing the GCA and healthy control. The panel including to the up-regulated metabolites of 2-Hydroxybutanoic acid (AUC=0.852; P=0.000; 95%CI 0.823-0.831), All-trans-3,14-dihydroretinol (AUC=0.642; P=0.000;95%CI 0.604-0.681), and Malonic acid (AUC=0.593; P=0.000;95%CI 0.549-0.636), and the down-regulated metabolites of12-ketolithocholic acid (AUC=0.288; P=0.000; 95%CI 0.253-0.322), Aldosterone (AUC=0.263; P=0.000; 95%CI 0.227-0.298), Hododeoxycholic acid (AUC=0.139; P=0.000;95%CI 0.111-0.167), L-histidine (AUC=0.358; P=0.000; 95%CI 0.318-0.398), Prostagl and in E2 (AUC=0.224; P=0.000; 95%CI 0.193-0.256), and Sphingosine (AUC=0.271; P=0.000; 95%CI 0.235-0.307), showed the AUC of 0.976, sensitivity of 0.913, and specificity of 0.027, with the optimal cut off value of 0.470 showed in Fig. 2A.

As we mentioned above, we found 3 potential metabolites (2-Hydroxybutanoic acid, All-trans-13,14-dihydroretinol, and Malonic acid) with AUC above 0.5 with P-value below 0.05, which indicated the good diagnostic character of the 3 metabolites. The panel including to the up-regulated metabolites of 2-Hydroxybutanoic acid (AUC=0.852; P=0.000; 95%CI 0.823-0.831), All-trans-3,14-dihydroretinol (AUC=0.642; P=0.000; 95%CI 0.604-0.681), and Malonic acid (AUC=0.593; P=0.000; 95%CI 0.549-0.636) showed the AUC of 0.866, sensitivity of 0.913, 95% CI of 0.839-0.893, and specificity of 0.027, with the optimal cut off value of 0.470 showed in Fig. 2B.

Differential metabolites for survival analysis

We obtained 12 metabolites including to 9,10-DiHOME [(±)9,10-dihydroxy-12Z-octadecenoic acid], 12,13-DiHOME, Aldosterone, 9-HpODE, 12,13-EpOME [(±)12(13)epoxy-9Z-octadecenoic acid], 9,10-EpOME [(±)9,10-Epoxy-12Z-octadecenoic acid], Lysope 18:2 (2N Isomer), 2-Hydroxycaprylic acid, 23-deoxydeoxycholic acid, 7-ketolithocholic acid, 12-ketolithocholic acid, and RvE1 [5S,12R,18R-trihydroxy-6Z,8E,10E,14Z,16E-eicosapentaenoic acid] were significant associated with survival, and excretion of exogenous substances, we finally obtained 7 biomarkers of 12,13-DiHOME, Aldosterone, 12,13-EpOME [(±)12(13)epoxy-9Z-octadecenoic acid], Lysope 18:2 (2N

Isomer), 23-Deoxydeoxycholic acid, 7-Ketolithocholic acid, and 12-Ketolithocholic acid, which were showed significant association with survival, and the results were showed in Fig. 4.

Metabolic pathway analysis

To gain a deeper understanding of the biological significance of the potential biomarkers identified in the present study, KEGG database was used for metabolic pathway analysis. Perturbed metabolic pathways were mainly related to: ①alanine, aspartate and glutamate metabolism, ②arginine biosynthesis, ③phenylalanine, tyrosine and tryptophan biosynthesis, and ④ Vitamin B6 metabolism. These pathways are mainly associated with energy metabolism, inflammatory reactions and immune responses.

It is worth noting that eight metabolites of N-Acetylaspartate (up-regulated), L-Aspartic Acid (down-regulated), L-Alanine (down-regulated), L-Glutamic Acid (down-regulated), L-Asparagine Anhydrous (down-regulated), Argininosuccinic acid (up-regulated), N-Carbamoyl-L-aspartate (up-regulated), and L-Glutamine (down-regulated) were collected in①alanine, aspartate and glutamate metabolism, and three metabolites of L-Glutamine (up-regulated), Argininosuccinic acid (up-regulated), and N-Acetylornithine (down-regulated) were collected in ②arginine biosynthesis, in addition, the L-Tyrosine (down-regulated), 4-Pyridoxic Acid (up-regulated) were collected in ③phenylalanine, tyrosine and tryptophan biosynthesis, and ④Vitamin B6 metabolism, respectively.

For the ①Alanine, aspartate and glutamate metabolism pathway, GCA patients and healthy control serum both collected 8 metabolites with the same regulated trend for discovery set and validation set, and 5 of 8 metabolites were down-regulated in GCA patients compared with healthy controls, which indicated in detailly that the 5 amino acids of L-Aspartic Acid, L-Alanine, L-Glutamic Acid, L-Asparagine Anhydrous, and L-Glutamine biosynthesis pathway were restrained, and 3 of 8 metabolites were up-regulated in GCA patients compared with healthy controls, which indicated that the 3 metabolites biosynthesis pathway of N-Acetylaspartate, Argininosuccinic acid, and N-Carbamoyl-L-aspartate were activated.

For the ②Arginine biosynthesis pathway, GCA patients and healthy control serum both collected 3 metabolites with the same regulated trend for discovery set and validation set,and 2 of 3 metabolites of were up-regulated in GCA patients compared with healthy controls,

which indicated in detailly that the amino acid of L-Glutamine and organic acid of Argininosuccinic acid biosynthesis pathway were activated, and 1 of 3 metabolites were down-regulated in GCA patients compared with healthy controls, which indicated that the 1 metabolites biosynthesis pathway of N-Acetylornithine were restrained.Citrulline, one kind of α-amino acid, which is abundant in watermelon. It is the precursor of arginine and nitric oxide. It mainly participates in the ornithine cycle in the body. The current research on citrulline mainly focuses on immune diseases such as rheumatoid arthritis[18]. Barza et al. [19]found that plasma citrulline level is one of the potential non-invasive biomarkers of whether the gastrointestinal mucosa is damaged. Citrulline in the body comes from the glutamate pathway and the proline pathway. The synthesis of proline in the body includes the glutamate pathway and the ornithine pathway, in which glutamate is mainly derived from the citric acid cycle. Proline is mainly involved in collagen synthesis in the body[20]. In the study of colon cancer, proline can induce the production of hypoxia inducible factor-1α (HIF-1α), thereby promoting angiogenesis[21]. Arginine is an essential amino acid for the human body and participates in a variety of life activities. It is the precursor of nitric oxide, polyamines, proline, creatinine and glutamic acid. It is also a key amino acid in the ornithine cycle. Studies have shown that tumor cells lose the ability to independently synthesize arginine, so arginine depletion may be a target for the treatment of tumors[22]. There are currently 5 methods for depleting arginase. Considering the effectiveness, immunogenicity, stability and potential by-products of arginine depletion, only modified arginase and arginine deiminase are used. Applied to the treatment of related tumors [23]. Lohavanichbutr et al. [24] analyzed 101oral squamous cell carcinoma (OSCC) patients and 35 normal human saliva samples by nuclear magnetic resonance spectroscopy, liquid chromatography-tandem mass spectrometry, etc. The results showed that glycine, proline, and ornithine in the saliva of early OSCC patients The levels of four potential biomarkers such as citrulline and citrulline are lower than those of normal people, and the decrease of citrulline content is contrary to the results of this experimental study. This study found that the citrulline content in OSCC tissues was significantly increased, suggesting that the catabolism of proline and arginine in OSCC is enhanced, so citrulline, glutamate cycle, arginine and proline metabolism are in the development of OSCC. The role of citrulline is for further study.

For ③ phenylalanine, tyrosine and tryptophan biosynthesis pathway, GCA patients and healthy control serum both collected 1 metabolites of L-Tyrosine with the same regulated

trend for discovery set and validation set, which was down-regulated in GCA patients compared with healthy controls, and it was indicated that the L-Tyrosine biosynthesis pathway was restrained.

④ Vitamin B6 metabolism is one important metabolic pathway for human beings, in this present work, GCA patients and healthy control serum both collected 1 metabolites of 4-Pyridoxic Acid with the same regulated trend for discovery set, which was up-regulated in GCA patients compared with healthy controls, and it was indicated that the 4-Pyridoxic Acid biosynthesis pathway was activated.

We totally collected 53 metabolites for the metabolic pathway analysis, and 12 metabolites were collected with the same regulated trend, and 9 amino acids of 12 metabolites with 3 up-regulated and 6 down-regulated. And the other 3 metabolites were organic acid with 2 up-regulated and 3 down-regulated.

Discussion

The nature of malignant tumors is a genetic disease, and also a metabolic disease. Its occurrence, development, sensitivity to radiotherapy and chemotherapy, and toxic side effects are also caused by genetic abnormalities, but from genetic abnormalities to differences in functional performance, experience. The extremely complex biochemical process is mainly the result of the joint action of the individual patient, anti-tumor therapy and the tumor itself. To study the occurrence and development of tumors based on metabolomics has the advantages of objectiveness, accuracy, efficiency and directness. At present, early diagnosis of breast cancer, liver cancer, colorectal cancer, pancreatic cancer and other applied metabolomics has been studied. From this point of view, GCA should also have its own unique metabolic characteristics, which can be used to reveal the internal physiological and biochemical processes of its tumors, and may even further push back to the abnormality of gene regulation.

In this study, a cohort of 864 participants were enrolled that included healthy controls (n=588) and GCA participants (n=276) from 2 cohorts (discovery set and validation set). This study attempts to compare the differences in serum metabolomics between patients with cardia cancer and normal people, and to look for specific serum metabolites of cardia cancer, to provide clues for understanding the occurrence and development and early diagnosis of cardia cancer, as well as the sensitivity of radiotherapy and chemotherapy for cardia cancer in the next step. The evaluation provides the experimental basis. At present, there have been

many metabolomics researches on gastric cardia cancer, but the research design is very different, the specimens used are different, the detection methods and conditions used are not uniform, so the results are very different and the experiment reproducibility is poor.

We find no reports have been analyzed the serum metabolomics about GCA, however, owing to the sister cancer relationship between GCA and EAC, we could learn a lot from the EAC reports. Some previous metabolomics studies based on nuclear magnetic resonance spectroscopy [25, 26] showed that the metabolite composition in the urine of esophageal cancer patients is different from that in healthy controls, but the consistency of the different metabolic components screened is poor. Some studies have tried to compare esophageal cancer The correlation between tumor tissue and urine metabolism spectrum, to evaluate the reliability of using urine as a test sample, the results show that there is a certain correlation between the obvious metabolic characteristics and metabolic pathway disorders between the patient's tumor tissue and urine, and urine metabolism. How does the change of characteristics reflect the metabolism in esophageal tumor tissue, which may also undergo a series of biochemical processes, still needs further research[27]. Due to the complexity of biological samples encountered in metabolomics, in order to detect as many metabolites as possible, mass spectrometer (MS) analysis is usually combined with separation techniques. Common separation methods include gas chromatography (GC), liquid chromatography (liqiud chromatography, LC). Gas chromatography is suitable for the separation of compounds with good thermal stability or volatility. Liquid chromatography-mass spectrometry technology combines the high efficiency and rapid separation performance of liquid chromatography with the high sensitivity and high specificity of mass spectrometry. For complex biological samples such as polar, hard-to-volatile or macromolecular compounds, It is an ideal choice for determination and analysis[28]. Therefore, this study still uses LC-MS as the main detection technology. Sample preparation is a key step in metabolomics. Therefore, a comprehensive and systematic screening strategy for tissue preparation programs is very desirable. An author in my country has developed an optimization and evaluation strategy based on LC-MS to screen different types of endogenous metabolites (amino acids, carnitine, choline, etc.) with high extraction efficiency and reproducible esophageal tissue preparation programs , And pay special attention to low-level metabolites, and established a program with good stability, sensitivity and reproducibility, called "stepwise addition of solvents and a homogenized wet tissue program"

(stepwise addition of solvents and a homogenized wet tissue). protocol, SWHW)[29]. It can be seen that the sample selection and preparation process in the experiment is very important.

In this present work, we found 200 metabolites selected from the discovery set and validation set from the statistical analysis based on metabolomics (Table 2S), we found 35 CAR (carnitines) metabolites with decreased metabolism during the process of canceration of the cardiac. CAR compounds play an important role in energy metabolism, they mainly participate in fatty acid metabolism in body to provide energy. Many functions of carnitine have not been clearly elucidated, and many mechanisms regulating carnitine metabolism are still unclear. Especially for the down-regulated metabolite of dodecylcarnitine, which may be strongly related to the metabolic energy demand of tumors.

19 lysophosphatidylcholine (LPC) were selected from the 2 sets. The LPC is mainly involved in the glycerophospholipid metabolism pathway. It is the product obtained by phospholipid losing a molecular fatty acid bond under the action of phospholipase A1 or A2. It has pro-inflammatory and promoting apoptosis and the role of cell membrane lysis. As an endogenous metabolite with surface activity, LPC can transfer cytochrome C to the cytoplasm by regulating the mitochondrial membrane potential, which can destroy the cell membrane of various tissues and cause organ damage. Studies have shown that LPC can be used as a diagnostic marker for alcoholic liver disease[30]and liver cancer[31]. Taylor et al. [32]found that the concentration of LPC in cancer patients decreased, and the plasma concentration of LPC was related to weight loss and inflammatory parameters, so it may be an indicator of the severity of malignant diseases. However, the content of LPC decreases in patients with ovarian cancer[33] and increases in patients with colon cancer[34]. The specific mechanism needs to be further studied. The decrease in LPC content in this experiment may be related to the active proliferation of OSCC cells and the increased anti-apoptotic activity.

Acknowledge

The authors thank to the National Natural Science Foundation of China (Grant number: 81872032), and thanks to the 60,000 graduate students who have made great contributions to our laboratory database construction, for their collection of samples and the construction of clinical database,cannot be reflected in the article, but they make a great contribution to this article. The control groups mentioned in this work were selected from our laboratory database by the long term follow-up population in high incidence areas, and we thanks to them because most of them were normal people during follow-up, less number of real cancer was obtained finally.

Author contributions:

Lidong Wang conceived the original idea and designed the study. Mengxia Wei prepared the draft of the study. Mengxia Wei and Panpan Wang contributed to the experiments, data analysis and data interpretation. Lidong Wang revised the manuscript. The other authors made great jobs for the data statistics and manuscript revising. All authors agreed with the conclusion and approved the final version of manuscript.

Funding:

This study was funded by the National Natural Science Foundation of China (Grant number: 81872032).

Declarations

Ethics approval and consent to participate

Ethical approval for the study was obtained from the Zhengzhou University Life Science Ethics Review Committee

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]State Key Laboratory of Esophageal Cancer Prevention & Treatment and Henan Key Laboratory for Esophageal Cancer Research of The First Affiliated Hospital, Zhengzhou University, Zhengzhou, Henan Province, 450052, PR China. [2]Department of Thoracic Surgery, Anyang Tumor Hospital, Anyang Henan Province, 455000, PR China. [3]Department of Pathology, Linzhou Tumor Hospital, Linzhou 456500, Henan Province, China;

Department of Pathology, Linzhou Tumor Hospital, Linzhou 456500, Henan Province, China. [4]Department of Surgery, Hui Xian Meng Tang Hospital, Xinxiang 453000, Henan Province, China. [5]Department of Oncology, Anyang District Hospital, Anyang 455000, Henan Province, China.

References

1. Lai S, Wang G. Rule of carcinogenetic development of gastric cardia observed by endoscopy. Chinese Journal of Cancer Research. 2006;18:74-76. https://doi.org/10.1007/s10330-005-0447-9.
2. Song G, Ma Q, Ma S, Chen, C, Wei W. Analysis of the incidence and age characteristics of upper gastrointestinal cancer among 2003-2012 in the high incidence area of esophageal cancer, Cixian County, in Hebei Province. Zhonghua Yu Fang Yi Xue Za Zhi. 2017;51:398-402. https://doi.org/10.3760/cma.j.issn.0253-9624.2017.05.006.
3. Zhang G, Su M, Wang D, Hu S, Liu M, Li J, et al. Genetic heterogeneity of oesophageal cancer in high-incidence areas of southern and northern China. PloS one. 2010;5:e9668-e9668. https://doi.org/10.1371/journal.pone.0009668.
4. Bano K, Al-Shoha M, Raghavapuram S, George N, Perisetti A, Mallinger W, et al. Gastric cardia cancer masquerading as a fundic subepithelial tumor. Am. J. Gastroenterol. 2018;113:1428. https://doi.org/10.1038/s41395-018-0086-8.
5. Christofk H, Vander Heiden M, Harris M, Ramanathan A, Gerszten R, Wei R, et al. The M2 splice isoform of pyruvate kinase is import ant for cancer metabolism and tumour growth. Nature. 2008;452:230-233. https://doi.org/10.1038/nature06734.
6. Hanaha D, Weinberg R. Hallmarks of cancer: the next generation. Cell. 2011; 144:646-674. https://doi.org/10.1016/j.cell.2011.02.013.
7. Nicholson J, Lindon J, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica. 1999;29: 1181-1189. https://doi.org/10.1080/004982599238047.
8. Dunn W, Ellis D. Metabolomics: Current analytical platforms and methodologies. TrAC Trends Anal. Chem. 2005;24:285-294. https://doi.org/10.1016/j.trac.2004.11.021
9. Cheung P, Ma M, Tse H, Yeung K, Tsang H, Chu M, et al. The applications of metabolomics in the molecular diagnostics of cancer, Expert Rev. Mol. Diagn., 2019;19:785-793. https://doi.org/10.1080/14737159.2019.1656530.
10. Smilde A, Westerhuis J, Hoefsloot H, Bijlsma S, Rubingh C, Vis D, et al. Dynamic metabolomic data analysis: a tutorial review. Metabolomics. 2010;6:3-17. https://doi.org/10.1007/s11306-009-0191-1.
11. Trezzi J, Galozzi S, Jaeger C, Barkovits K, Brockmann K, Maetzler W, et al. Distinct

metabolomic signature in cerebrospinal fluid in early parkinson's disease. Mov. Disord. 2017;32:1401-1408. https://doi.org/10.1002/mds.27132.

12. Chang K, Cheng M, Tang H, Huang C, Wu Y, Chen C. Alternations of metabolic profile and kynurenine metabolism in the plasma of parkinson's disease. Mol. Neurobiol. 2018;55:6319-6328. https://doi.org/10.1007/s12035-017-0845-3.
13. Phelan M, Caamaño-Gutiérrez E, Gant M, Grosman R, Madine J. Using an NMR metabolomics approach to investigate the pathogenicity of amyloid-beta and alpha-synuclein. Metabolomics. 2017;13:151. https://doi.org/10.1007/s11306-017-1289-5.
14. Shukla A, Ratnasekhar C, Pragya P, Chaouhan H, Patel D, Chowdhuri D, et al. Metabolomic analysis provides insights on paraquat-induced parkinson-like symptoms in Drosophila melanogaster. Mol. Neurobiol. 2016;53:254-269. https://doi.org/10.1007/s12035-014-9003-3.
15. Blennow K, Zetterberg H. Biomarkers for Alzheimer's disease: current status and prospects for the future. J. Intern. Med. 2018;284:643-663. https://doi.org/10.1111/joim.12816.
16. McGuire A, Brown J, Kerin M. Metastatic breast cancer: the potential of miRNA for diagnosis and treatment monitoring. Cancer Metastasis Rev. 2015:34:145-155. https://doi.org/10.1007/s10555-015-9551-7.
17. Pitkänen A, Ndode-Ekane X, Lapinlampi N, Puhakka N. Epilepsy biomarkers- Toward etiology and pathology specificity. Neurobiol. Dis. 2019;123:42-58. https://doi.org/10.1016/j.nbd.2018.05.007.
18. Yang Q, Yang G, Wan L, Han Y, Huo Y, Li J, et al. Protective effects of dexrazoxane against Doxorubicin-induced cardiotoxicity: A metabolomic study. PLoS One, 2017;12:e0169567. https://doi.org/10.1371/journal.pone.0169567.
19. Yang Q, Zhang J, Zhang J, Han Y, Xin B, Zhang J, et al. Distinct metabolic profile of inhaled budesonide and salbutamol in asthmatic children during acute exacerbation. Basic Clin. Pharmacol. Toxicol. 2017;120:303-311. https://doi.org/10.1111/bcpt.12686.
20. Cao M, Han Q, Zhang J, Zhang R, Wang J, Gu W, et al. An untargeted and pseudotargeted metabolomic combination approach to identify differential markers to distinguish live from dead pork meat by liquid chromatography–mass spectrometry. J. Chromatogr. A. 2020;1610:460553. https://doi.org/10.1016/j.chroma.2019.460553.
21. Scherer H, Huizinga T, Krönke G, Schett G, Toes R. The B cell response to citrullinated

antigens in the development of rheumatoid arthritis. Nat. Rev. Rheumatol. 2018;14:157-169. https://doi.org/10.1038/nrrheum.2018.10.

22. Barzał J, Szczylik C, Rzepecki P, Jaworska M, Anuszewska E. Plasma citrulline level as a biomarker for cancer therapy-induced small bowel mucosal damage. Acta Biochim. Pol. 2014;61:615-631. https://doi.org/10.18388/abp.2014_1823.
23. Karna E, Szoka L, Huynh T, Palka J. Proline-dependent regulation of collagen metabolism. Cell Mol. Life Sci. 2020;77:1911-1918. https://doi.org/10.1007/s00018-019-03363-3.
24. Surazynski A, Donald S, Cooper S, Whiteside M, Salnikow K, Liu Y, et al. Extracellular matrix and HIF-1 signaling: the role of prolidase. Int. J. Cancer. 2008;122:1435-1440. https://doi.org/10.1002/ijc.23263.
25. Fultang L, Vardon A, Santo C, Mussai F. Molecular basis and current strategies of therapeutic arginine depletion for cancer. Int. J. Cancer. 2016;139:501-509. https://doi.org/10.1002/ijc.30051.
26. Zou S, Wang X, Liu P, Ke C, Xu S. Arginine metabolism and deprivation in cancer therapy. Biomed. Pharmacother. 2019;118:109210. https://doi.org/10.1016/j.biopha.2019.109210.
27. Lohavanichbutr P, Zhang Y, Wang P, Gu H, Nagana Gowda G, Djukovic D, et al. Salivary metabolite profiling distinguishes patients with oral cavity squamous cell carcinoma from normal controls. PLoS One. 2018;13:e0204249. https://doi.org/10.1371/journal.pone.0204249.
28. Davis V, Schiller D, Eurich D, Sawyer M. Urinary metabolomic signature of esophageal cancer and Barrett's esophagus. World J. Surg. Oncol. 2012;10: 271-271. https://doi.org/10.1186/1477-7819-10-271.
29. Hasim A, Ma H, Mamtimin B, Abudula A, Niyaz M, Zhang L, et al. Revealing the metabonomic variation of EC using 1H-NMR spectroscopy and its association with the clinicopathological characteristics. Mol. Biol. Rep. 2012;39:8955-8964. https://doi.org/10.1007/s11033-012-1764-z.
30. Liang J, Lin Y, Ouyang T, Tang W, Huang Y, Zhao J, et al. Nuclear magnetic resonance-based metabolomics and Ye metabolic pathway networks from patient-matched esophageal carcinoma, adjacent noncancerous tissues and urine. World J. Gastroenterol. 2019;25:3218-3230. https://doi.org/10.3748/wjg.v25.i25.3218.

689 31. Nagana Gowda G, Raftery D. Overview of NMR spectroscopy-based metabolomics:
690 opportunities and challenges. Methods Mol. Biol. 2019;2037:3-14.
691 https://doi.org/10.1007/978-1-4939-9690-2_1.
692 32. Wang H, Xu J, Chen Y, Zhang R, He J, Wang Z, et al. Optimization and evaluation
693 strategy of esophageal tissue preparation protocols for metabolomics by LC-MS. Anal.
694 Chem. 2016;88:3459-3464. https://doi.org/10.1021/acs.analchem.5b04709.
695 33. Stefanescu H, Suciu A, Romanciuc F, Crisan D, Procopet B, Radu C, et al. Lyso-
696 phosphatidylcholine: A potential metabolomic biomarker for alcoholic liver disease?
697 Hepatology. 2016;64:678-679. https://doi.org/10.1002/hep.28630.
698 34. Huang X, Zeng J, Zhou L, Hu C, Yin P, Lin X. A new strategy for analyzing time-series
699 data using dynamic networks: Identifying prospective biomarkers of hepatocellular
700 carcinoma. Sci. Rep. 2016;6:32448. https://doi.org/10.1038/srep32448.
701

Figures

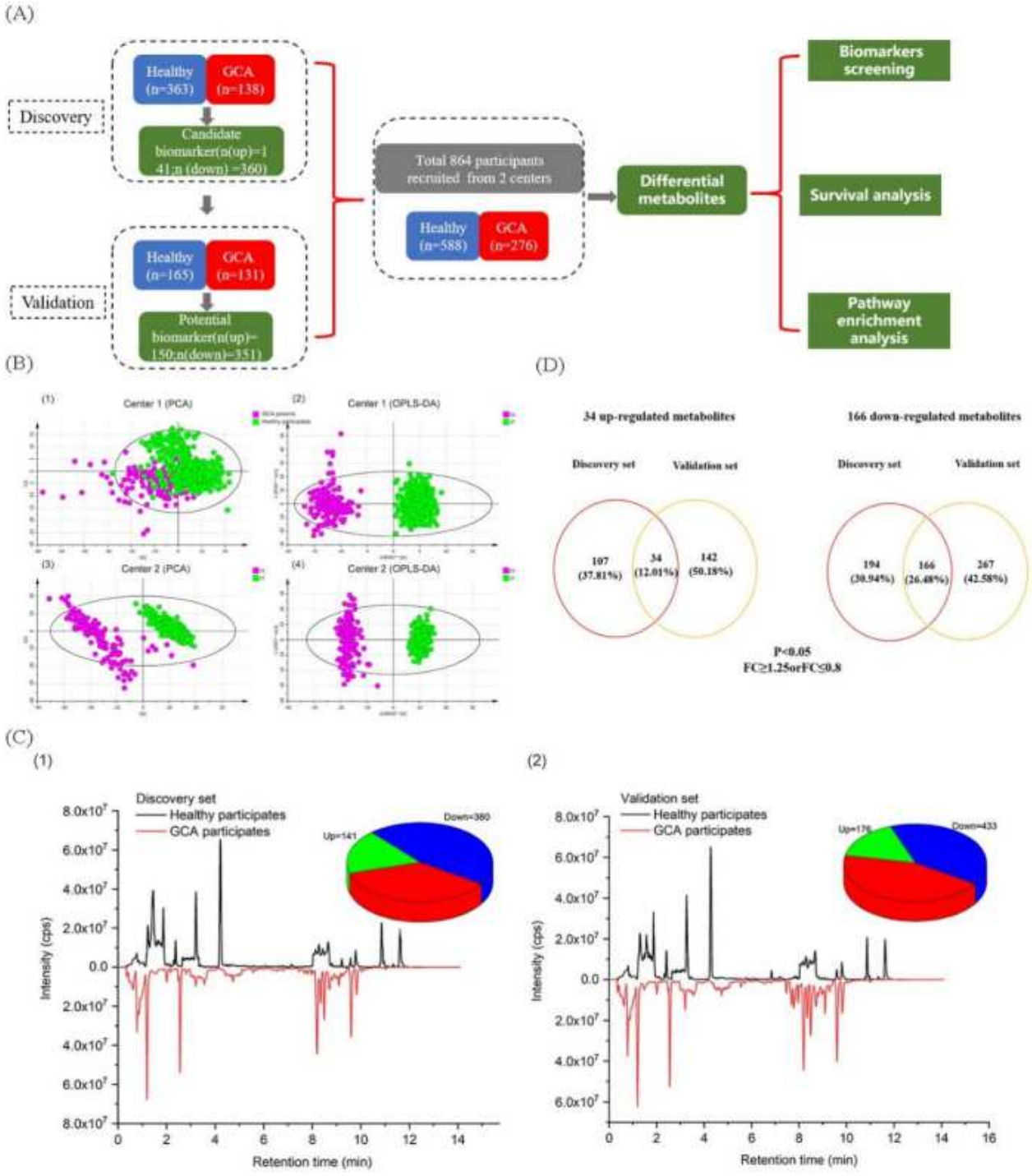

Figure 1

(A) Study design. There are a total of 864 participants, including 588 healthy volunteers, 276 GCA patients in the discovery and validation sets. (B). Score plots of principal component analysis based on the combined data of ESI+ and ESI− modes from the discovery set. (C) Total ion chromatograms of the discovery set and validation set for the healthy vs GCA groups. (The x-axis represents retention time, and

the y-axis represents the charge-to-mass ratio of the features. Each circle in the cloud plot represents 1 differential feature, and the circle size represents the relative concentration of the feature. Before differential metabolite analysis, first perform principal component analysis on the grouped samples for difference comparison, and observe the degree of variability between the difference groups and the samples within the group.) (D). Venn diagram displaying the 200 differential metabolitesthat were altered as biomarker candidates from the 2 cohorts in the discovery set and validation set.

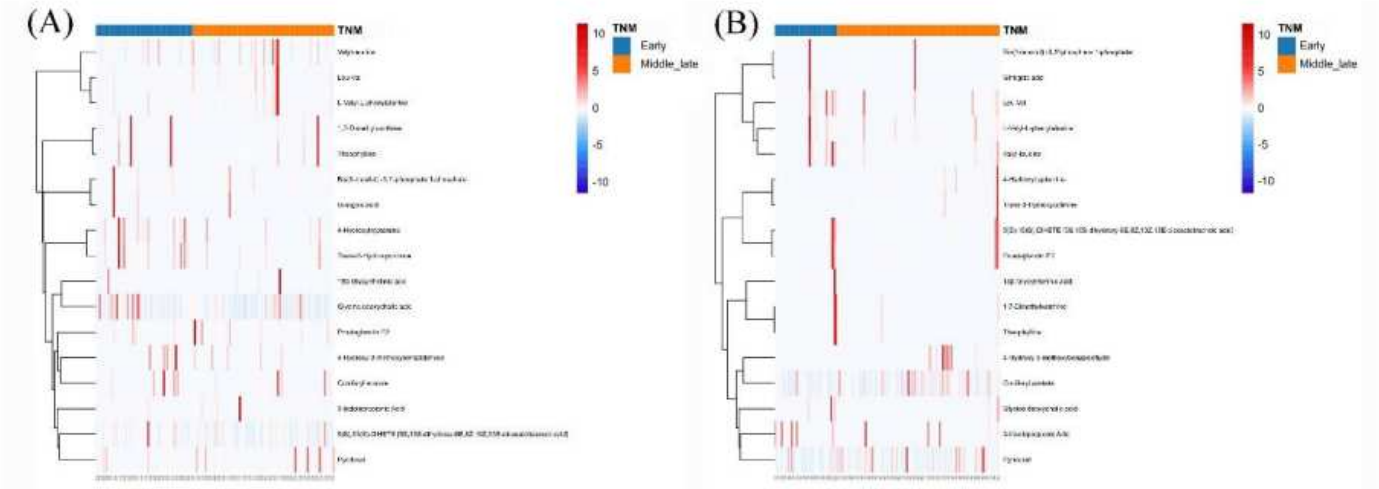

Figure 2

Metabolomic profile of matched early and middle-late stage of GCA for discovery set (A) and validation set (B).

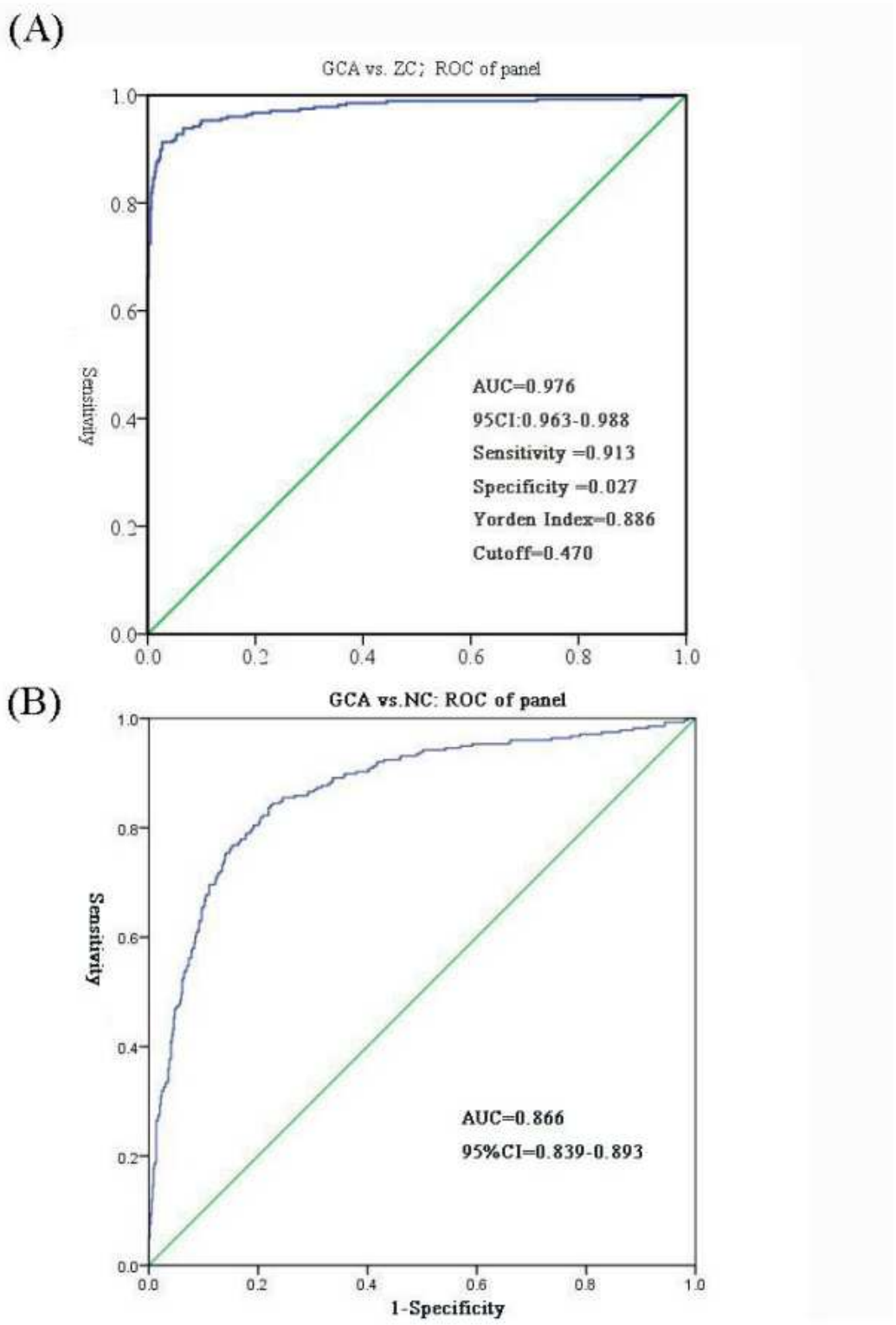

Figure 3

(A). GCA vs. NC. Roc of panel (9 different metabolites of 12-ketolithocholic acid, 2-Hydroxybutanoic Acid, Aldosterone, All-Trans-13,14-Dihydroretinol, Hododeoxycholic acid, L-Histidine, Malonic acid, Prostaglandin E2, and Sphingosine).(B). GCA vs. NC. Roc of panel (3 different metabolites of 2-Hydroxybutanoic Acid, All-Trans-13,14-Dihydroretinol, and Malonic acid)

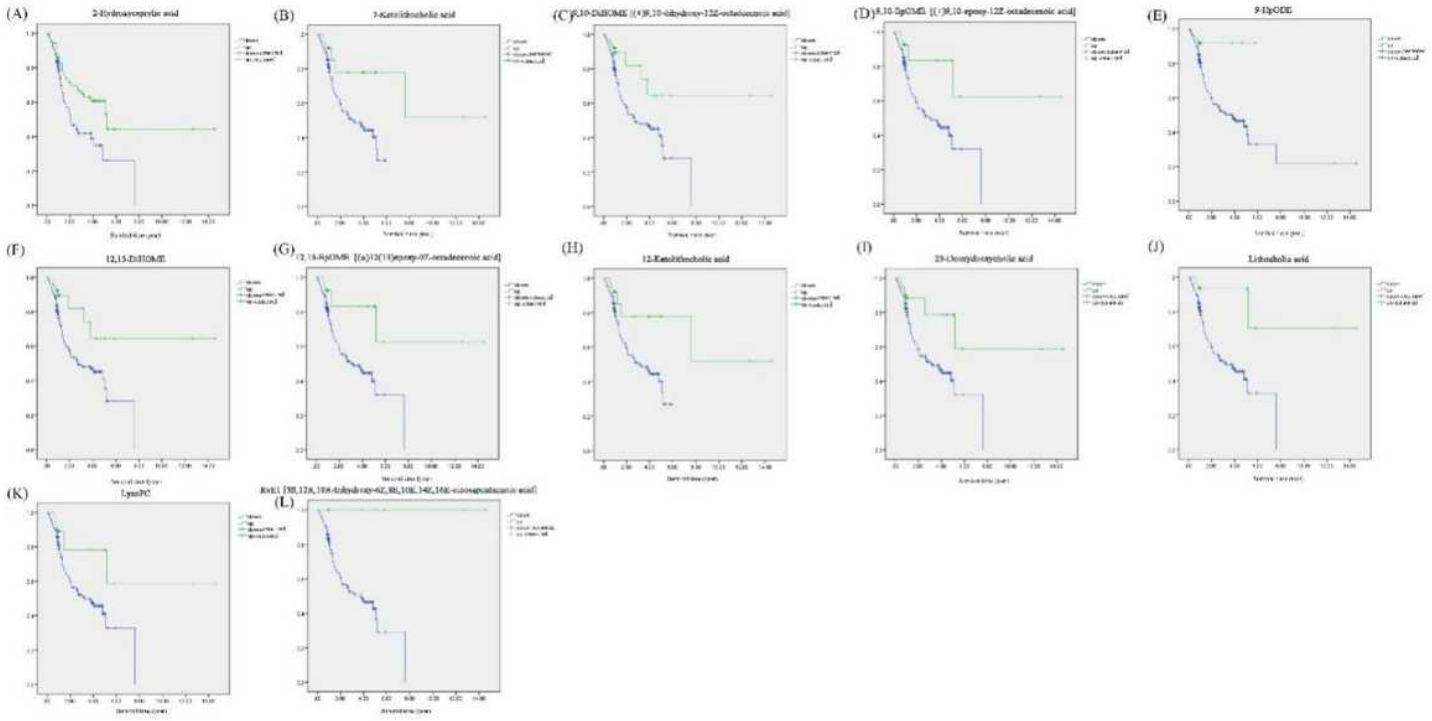

Figure 4

Survival analysis. We used the median of the metabolite in GCA as the cut-off value to set the up and down group.

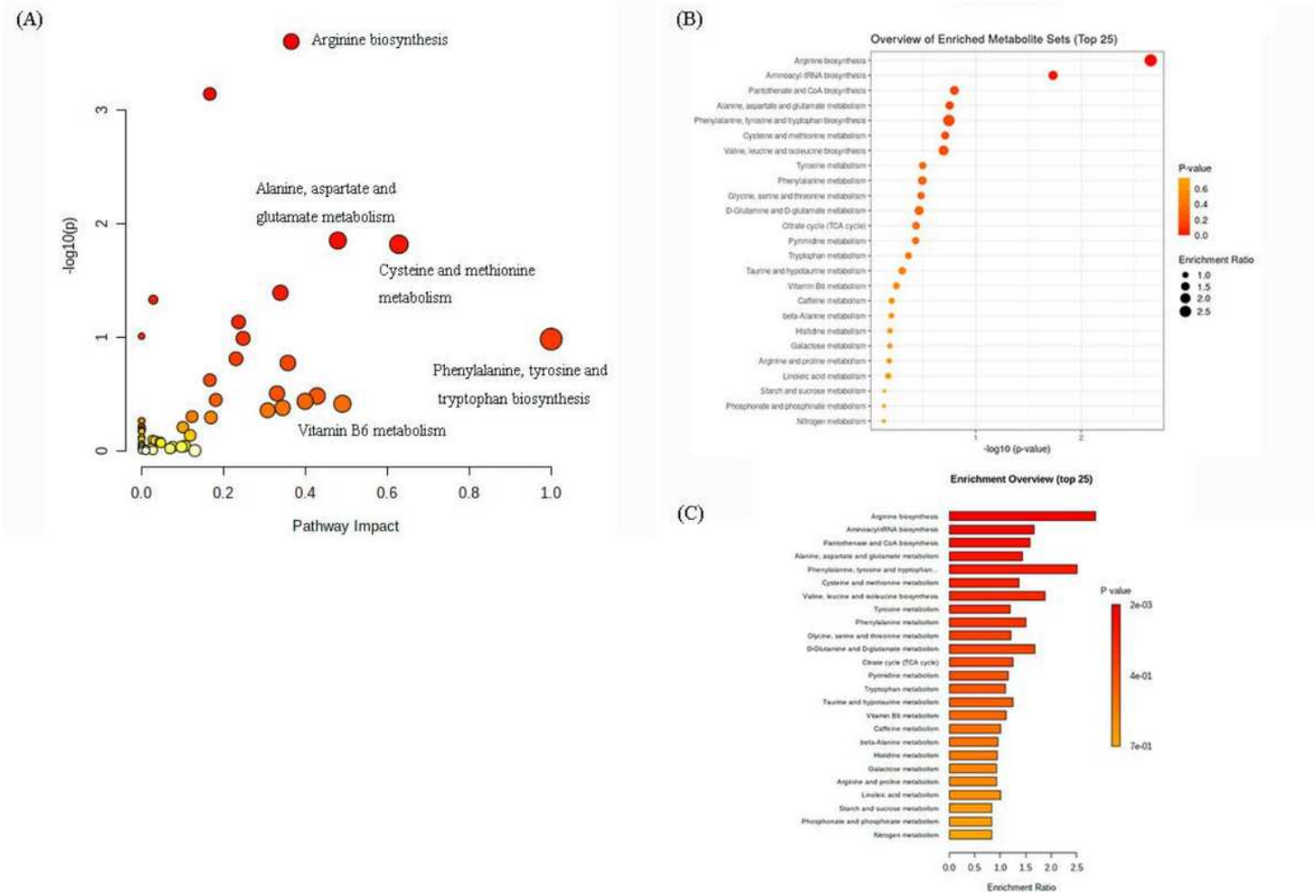

Figure 5

(A). Disturbed metabolic pathways identified from the candidate biomarkers in the 2 sets; (B). Overview of enriched metabolite sets (Top 25); (C). Metabolite set enrichment analysis ofmetabolite variations.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigS1.png
- Table12andTableS1S2.pdf