Int J Clin Exp Pathol 2018;11(5):2323-2336
www.ijcep.com /ISSN:1936-2625/IJCEP0074100

*Original Article*

# Free amino acid profiling of gastric juice as a method for discovering potential biomarkers of early gastric cancer

Jian Liu[1,2], Sanren Lin[1], Zhengpeng Li[1,3], Liya Zhou[1], Xiue Yan[1], Yan Xue[1], Lingmei Meng[1], Jingjing Lu[1], Baojun Suo[1], Wenjun Jiang[1]

*[1]Department of Gastroenterology, Peking University Third Hospital, Beijing 100191, China; [2]Department of Gastroenterology, Beijing Jishuitan Hospital, Beijing, China; [3]Department of General Surgery, 306 Hospital of PLA, Beijing, China*

Received February 6, 2018; Accepted March 20, 2018; Epub May 1, 2018; Published May 15, 2018

**Abstract:** *Background:* Prior studies showed that aromatic amino acids (AAAs) could be used as potential gastric juice biomarkers in screening gastric cancer (GC). To identify new biomarkers for early diagnosis of GC, the characteristics of gastric juice free amino acid (GJFAA) profiling was determined. *Method:* First, gastric juice was collected from 130 consecutive patients who underwent gastroscopy. They were divided into GC group (n = 47) and non-neoplastic gastric disease (NGD) group (n = 83) according to the pathological diagnosis. The concentrations of 34 GJFAAs were examined by amino acid analyzer. Multivariate and univariate analyses were used for comparing the alterations of GJFAA profiles between the two groups. Then candidate differential GJFAAs were verified by LC-MS/MS in another set of patients, which included 32 GC patients and 38 NGD patients. The diagnostic performance of GJFAAs was evaluated by ROC curve. *Results:* Significant alterations in GJFAA profiles were observed in GC patients compared to NGD patients in the training set. A total of 14 amino acids were screened as differential GJFAAs. Leucine, threonine and serine were the most frequently altered. Combined AUC of the three non-AAAs [0.869 (95% CI, 0.805-0.934)] was superior to the combined three AAAs [0.841 (95% CI, 0.773-0.908)]. In addition, a combined AUC comprisingthe six ones was further improved to 0.871 (95% CI, 0.809-0.933) in the diagnosis of GC. A similar variation trend and diagnostic value were observed in the validation set. *Conclusion:* This study indicates the potential of GJFAA profiling as a promising approach for the early detection and screening of GC.

**Keywords:** Gastric cancer, amino acid profiling, gastric juice free amino acid (GJFAA), aromatic amino acid (AAA), metabolic biomarker

## Introduction

Gastric cancer (GC) is the fifth most common cancer and third leading cause of cancer-related deaths worldwide [1]. The initial phase of GC remains relatively asymptomatic as most patients have already been in the progressive state at the time of diagnosis, and have lost the chance of receiving radical surgery. Therefore, early diagnosis plays a vital role in the improvement of curative rate of GC. Nowadays, clinical diagnosis still relies on endoscopy and histology, whereas GC can be easily missed during conventional endoscopy, especially early gastric cancer (EGC) due to a minor lesion, endoscopy-related inadequate biopsy sampling or incorrect interpretation [2]. Thus, population-based screening by reliable biomarkers is desperately required. However, the poor sensitivity and specificity of conventional tumor markers, like carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9) make it difficult to diagnose GC, while screened EGC alone [3]. Accordingly, more effective biomarkers are urgently needed.

Recently, utility of metabolism-based technologies opened a new window to discover potential biomarkers for metabolic disorders especially malignant diseases, given that pathological alterations in cell functions are frequently accompanied by metabolic reprogramming [4].

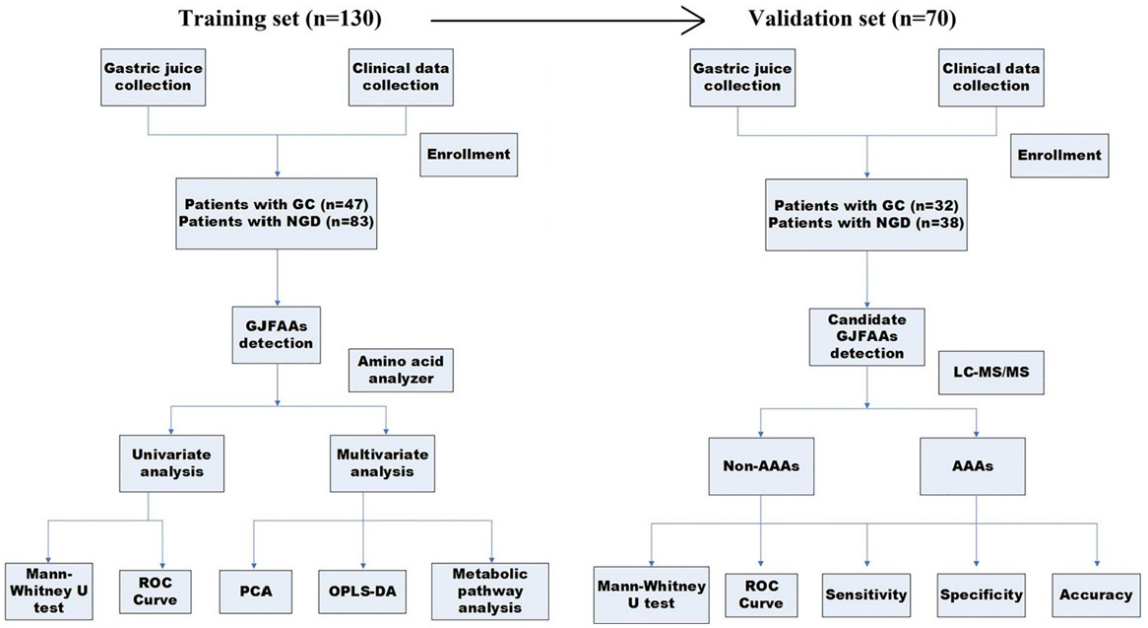


**Figure 1.** Flow diagram of the study. GC, gastric cancer; NGD, non-neoplastic gastric diseases; GJFAAs, gastric juice free amino acids; AAAs, aromatic amino acids; non-AAAs, non-aromatic amino acids; ROC, receiver operating characteristic curve; PCA, principal component analysis; OPLS-DA, orthogonal projections to latent structures discriminant analysis; LC-MS/MS, liquid chromatography-tandem mass spectrometry.

Nevertheless, some practical limitations still remain. These include the necessity to measure a huge number of metabolites, data-redundancy problems, including the false-discovery rate and overfitting, and cost constraints. One approach to overcoming these problems is "focused metabolomics", which limits the objects of the analysis to those that play roles in general metabolism and share physical similarities [5]. Amino acids are among the most suitable candidates as they are either ingested or synthesized endogenously and play essential physiological roles both as basic metabolites and metabolic regulators [6, 7]. To measure amino acids, e.g., plasma free amino acids (PFAAs), which abundantly circulate as a medium linking all organ systems, which are investigated in numerous studies have noted that increased levels of PFAAs in several cancer entities (such as lung, breast, cervix, kidney and digestive organs) compared to normal counterparts [8-15]. However, some investigations have highlighted that a fair number of PFAAs reflecting metabolic disruptions common to many cancers, rather than attacking unique metabolic features in any particular one [16, 17]. Gastric juice has direct contact with the gastric epithelium, and it contains metabolic information of gastric epithelial cells. Thus, gastric juice free amino acids (GJFAAs) may be more specific than PFAAs for the diagnosis of GC. But there are very few studies that cover the application of GJFAAs to GC.

Our previous studies have established several endogenous fluorescence spectra of gastric juice for diagnosing and screening of GC [18, 19]. Three fluorescent substances (aromatic amino acids, AAAs) in gastric juice were isolated and qualitatively identified using liquid chromatography-tandem mass spectrometry (LC-MS/MS) and (1)H-nuclear magnetic resonance ((1)H-NMR) [20]. AAAs in gastric juice can be applied to distinguish advanced gastric cancer and EGC from non-neoplastic gastric disease (NGD) through quantitative measurement using high-performance liquid chromatography (HPLC) [21]. Lately, we have built and tested a prediction model based on the male-to-female ratio, pH values, and AAA concentrations in gastric juice for diagnosing GC [22]. Interestingly, the results revealed that the level of multi-index (gastric juice index, GJI) begins to rise at the early stage of gastric carcinogenesis. It suggested that GJI has great potential for improving the early detection of GC. However, this work has focused on three AAAs that can produce autofluorescence, and the characteristic

**Table 1.** Patients' disease constitution

| Set | Patients | GC | | NGD | | | *P* value |
|---|---|---|---|---|---|---|---|
| | | AGC | EGC | GD | CAG/IM | CSG | |
| Training | 130 | 26 | 21 | 42 | 24 | 17 | 0.197 |
| Validation | 70 | 21 | 11 | 13 | 12 | 13 | |
| Total | 200 | 47 | 32 | 55 | 36 | 30 | |

All data are shown as count. GC, gastric cancer; NGD, non-neoplastic gastric diseases; AGC, advanced gastric cancer; EGC, early gastric cancer; GD, gastric dysplasia; CAG/IM, chronic atrophic gastritis and/or intestinal metastasis; CSG, chronic superficial gastritis. No statistically significant difference of disease constitution was found between the training and validation sets using Pearson's Chi-square test.

signature for the GJFAA profiling has remained poorly understood yet. Hence, this study outlined the metabolic patterns of GJFAA profiles in GC and NGD patients. Then, three most discrepancy GJFAAs were screened out and their diagnostic efficiencies were compared to and combined with three AAAs in gastric juice we focused on before. Finally, the analysis of differential GJFAA metabolic pathways was performed.

## Patients and methods

### *Study design and ethical considerations*

This study protocol was approved by the Peking University Third Hospital Medical Ethics Committee (IRB00006761-2016058). Informed consent was obtained from each patient, and the entire clinical investigation was conducted according to the Declaration of Helsinki. Gastric juice samples were collected from consecutive patients who underwent gastroscopy at our institution from December 2015 to April 2017. Altogether 200 patients were included in our study. Based on the sequence of the patients enrolled in the study, the first 130 cases were taken as the training set to compare the differences of GJFAA profiles between the GC and NGD groups and the latter 70 as the validation set to verify the diagnostic performance of the candidate biomarkers screened from the GJFAA profiles. The demographic and clinicopathological data were taken. The flow diagram of this study was presented in **Figure 1**.

### *Inclusion criteria*

Patients (1) suspected with various benign and malignant gastric diseases as diagnosed and confirmed by mucosal biopsy and/or postoperative pathology; (2) > 18 years of age; (3) who had not received any chemoradiation therapy.

### *Exclusion criteria*

Patients (1) whose gastric juice was contaminated by blood, bile, or gross food residue; diluted in the process of gastroscopy, where < 2 mL was insufficient to conduct the experiment; (2) with recurrent, metastatic, neuroendocrine carcinoma, or lymphoma in the stomach; (3) with gastrointestinal submucosal tumor or gastric polyp; (4) with lesions that predominantly localized in the esophagus or duodenum; (5) with organic diseases or concurrent neoplasms diagnosed by clinical examination; (6) who are pregnant or lactating.

### *Diagnostic criteria*

All patients were histologically confirmed by mucosal biopsy and/or postoperative pathology. Biopsies were obtained from the antrum, corpus and other suspicious sites (≥ 2). Abnormal lesions were sampled for histopathological examination. Diagnosis of each subject depended on the most severe lesion. In case of any discrepancy in pathological diagnosis between biopsy and postoperative specimens, the diagnostic results were mainly based on the latter one. GC was confirmed and classified according to the American Joint Committee on Cancer guidelines in 2010 [23]. The diagnosis of gastric dysplasia and chronic gastritis were according to the Padova International Classification [24] and the updated Sydney system [25], respectively.

### *Sample collection, preservation, and preparation*

Gastric juice (5-10 mL) was collected from the patients who underwent gastroscopy after overnight fasting. The samples were centrifuged at 3000 rpm, 4°C for 10 min. The supernatant was then preserved in 2 mL aliquots at -80°C for subsequent analysis.

### *GJFAA profiling analysis*

The concentrations of 34 amino acids and biogenic amines in gastric juice were targeted measured by Sykam automatic amino acid analyzer S 433-D (Sykam GmbH, Eresing, Germany).

**Table 2.** Patients' baseline characteristics

| Variable | Training set (n = 130) | | *P* value | Validation set (n = 70) | | *P* value |
|---|---|---|---|---|---|---|
| | GC (n = 47) | NGD (n = 83) | | GC (n = 32) | NGD (n = 38) | |
| Age (years) | 59.7 ± 12.9 | 63.2 ± 11.8 | 0.451[a] | 69.6 ± 11.5 | 60.6 ± 15.0 | 0.007[a,**] |
| Gender (M/F) | 32/15 | 35/48 | 0.005** | 24/8 | 16/22 | 0.006** |
| *H. pylori* positive | 14 | 13 | 0.056 | 7 (4 missed) | 5 | 0.218 |

Data on age are shown as mean ± standard deviation, other categorical data are shown as count. GC, gastric cancer; NGD, non-neoplastic gastric diseases; M/F, male/female. **Statistically significant difference ($P < 0.01$) using independent t-test (a) or Pearson's Chi-square test (others).

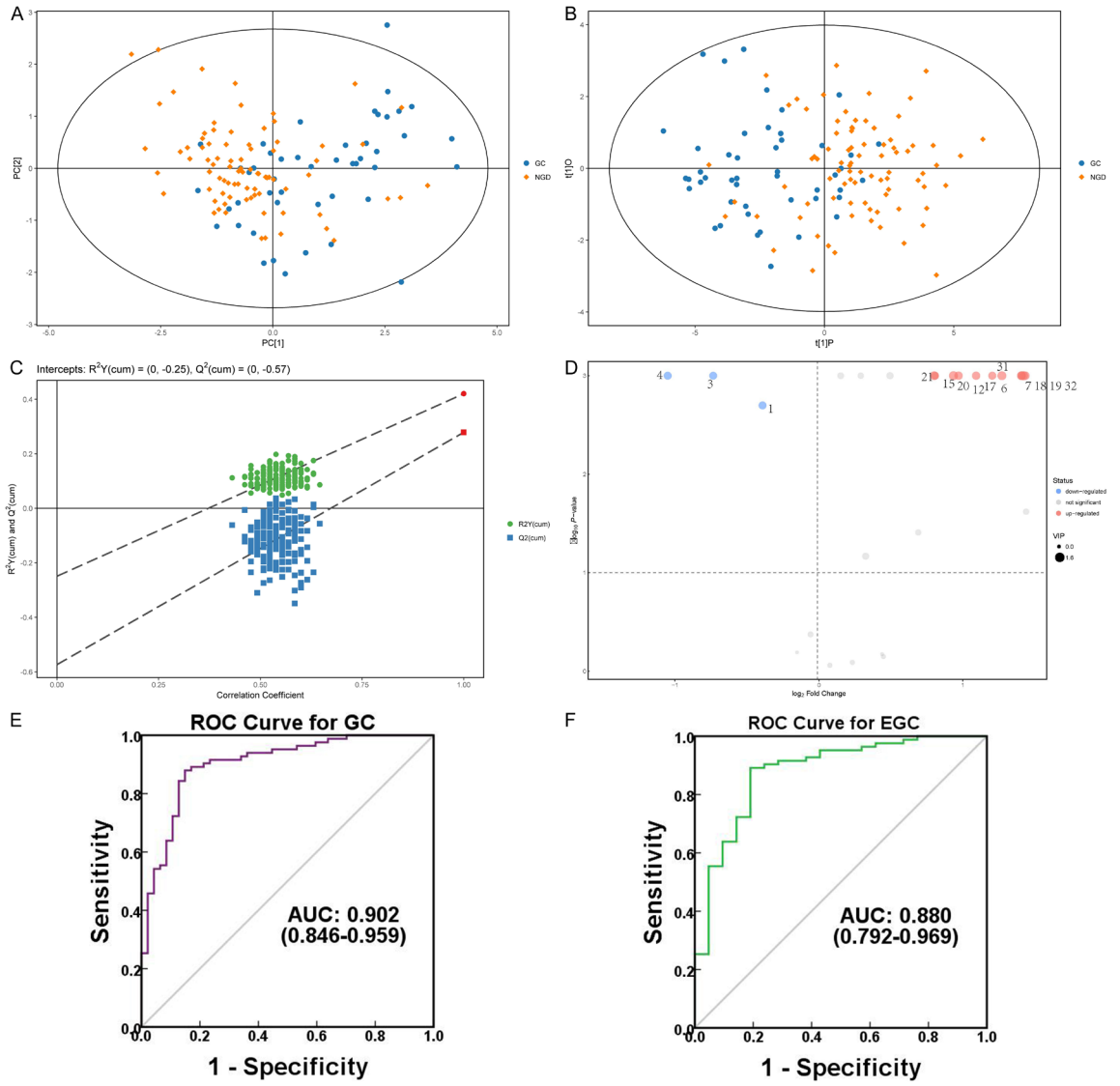


**Figure 2.** Score scatter plot of PCA model for group GC vs. NGD (A); Score scatter plot of OPLS-DA model for group GC vs. NGD (B); Permutation test of OPLS-DA model for group GC vs. NGD (C); Volcano plot for group GC vs. NGD (D); ROC curves for the detection of GC (E) and EGC (F) combined with 14 kinds of differential GJFAAs in the training set. GC, gastric cancer; EGC, early gastric cancer; NGD, non-neoplastic gastric diseases; ROC, receiver operating characteristic curve.

It is equipped with refrigerated reagent organizer S 7130, autosampler S 5200, solvent delivery system S 2100, amino acid reaction module S 4300, and Clarity Amino data analy-

**Table 3.** Differential GJFAAs between the GC and NGD patients in the training set

| GJFAA | Median GC | Median NGD | *P* value | VIP | FC |
|---|---|---|---|---|---|
| PSer | 0.028 | 0.037 | 0.002 | 1.054 | 0.768 |
| PEtN | 0.007 | 0.018 | < 0.001 | 1.028 | 0.606 |
| Urea | 0.178 | 0.604 | < 0.001 | 1.058 | 0.486 |
| Thr | 0.022 | 0.007 | < 0.001 | 1.489 | 2.431 |
| Ser | 0.016 | 0.005 | < 0.001 | 1.420 | 2.671 |
| Ala | 0.033 | 0.016 | < 0.001 | 1.238 | 1.973 |
| Val | 0.025 | 0.013 | < 0.001 | 1.025 | 1.763 |
| Met | 0.017 | 0.007 | < 0.001 | 1.178 | 2.148 |
| Ile | 0.026 | 0.007 | < 0.001 | 1.343 | 2.674 |
| Leu | 0.075 | 0.020 | < 0.001 | 1.626 | 2.697 |
| Tyr | 0.066 | 0.026 | < 0.001 | 1.580 | 1.926 |
| Phe | 0.069 | 0.032 | < 0.001 | 1.415 | 1.754 |
| Lys | 0.044 | 0.015 | < 0.001 | 1.091 | 2.321 |
| Arg | 0.036 | 0.008 | < 0.001 | 1.332 | 2.722 |

*P* value, Statistically significant difference using Mann-Whitney U test; VIP, variable importance in the projection; FC, fold change; GC, gastric cancer; NGD, non-neoplastic gastric diseases; GJFAA, gastric juice free amino acid.

sis software. Additionally, amino acid standard solution, ninhydrine, washing and regenerate solutions were all provided by the manufacturer. The composition of amino acid standard solution was shown in Table S1, and concentrations of them were 1.0 μmol/mL, except for urea which was 10 μmol/mL.

Gastric juice samples were thawed in a 4°C water bath. The cryopreserved gastric juice of 0.5 ml was added and mixed with 8% sulfosalicylic acid of 1.0 ml, and then centrifuged at 4°C at 12000 rpm for 15 min to precipitate proteins. The supernatant of 0.2 ml was transferred to 10 mL glass tube, evaporated to dryness. The sample was dissolved in 2 mL diluent solution, 100 μL was drawn and filtered by 0.22 μm filter unit (Millipore Corp. Carrigtwohill, Co. Cork, Ireland), which was then transferred to the ampere bottle for detection.

Volumes of 50 μL of each sample were injected into the HPLC system for analysis. Each standard was individually run on the gradient noted in the LC parameters program (reaction temperature 130°C, gradient elution analysis). A LCA K07/Li column (4.6 × 150 mm, 7 um) was used for the separation. The flow rate for ninhydrine solution was 0.25 mL/min, and the flow rate for mobile phase was 0.45 mL/min. The run time for one sample was 110 min, and the equilibration time for the column oven was 30 min. Further details like the gradient program steps were shown in Table S2. Integrated dual-channel photometer at 440 nm and 570 nm wavelengths for GJFAA detection was used. The chromatograms thus obtained were compared with the standards to record the areas of the peaks and used for calibration.

*LC-MS/MS analysis*

*Chemicals and reagents:* Leucine, threonine, serine, tyrosine, phenylalanine, tryptophan (purity 99.9%) and their isotope-labeled internal standards (IS) (purity 98.0%) were supplied by Beijing MS Medical Research Co.,Ltd. HPLC-grade formic acid was obtained from Dikma (Lake Forest, USA), methanol and acetonitrile (HPLC grade) were obtained from Fisher Scientific (Fair Lawn, NJ, USA). Water purified by Milli-Q system (Millipore, Bedford, MA, USA) was used for the analysis.

The chromatography system consisted of a Dionex Ultimate -3000 Reagent-Free™ Ion Chromatograph (Dionex Corporation, Sunnyvale, CA, USA) with an AS auto sampler. The chromatographic system was coupled with an API 3200 Q TRAP instrument (AB SCIEX, USA) operated in a multiple reaction monitoring mode (MRM). Data were acquired and processed using Analyst v1.5.2 software (Applied Biosystems).

*Sample treatment*

The gastric juice specimens were thawed at room temperature. Before the chromatographic analysis, 50 μL samples were deproteinized by adding 50 μL methanol (contains norvaline), instantaneous centrifugation at 2000 rpm for 10 s was done, followed by centrifugation at 13200 rpm, 4°C for 4 min. 10 μL of the supernatant was further diluted with 50 μL of buffer solution (contains isotope-labeled IS), and then was instantaneously centrifuged at 2000 rpm for 10 s. Added 20 μL derivatization liquid, instantaneous centrifugation at 2000 rpm for 10 s, followed by derivatization at 55°C for 15 min was done. The sample should be cooled in the refrigerator and then 5 μL was injected into the LC-MS/MS system.

*Chromatographic conditions*

A Phenomenex synergi MSLab45AA-C18 column (5 μm, 4.6 mm' 150 mm) was used for

**Table 4.** Discriminating performance of each GJFAA for the GC and EGC patients in the training set

| GJFAA | GC | | EGC | |
|---|---|---|---|---|
| | AUC (SE) | 95% CI | AUC (SE) | 95% CI |
| PSer | 0.666 (0.054) | 0.561-0.771 | 0.649 (0.076) | 0.500-0.798 |
| PEtN | 0.718 (0.052) | 0.615-0.820 | 0.731 (0.076) | 0.582-0.879 |
| Urea | 0.804 (0.039) | 0.729-0.880 | 0.788 (0.051) | 0.688-0.889 |
| Thr | 0.835 (0.037) | 0.764-0.907 | 0.833 (0.052) | 0.731-0.935 |
| Ser | 0.831 (0.037) | 0.759-0.903 | 0.821 (0.051) | 0.722-0.920 |
| Ala | 0.783 (0.042) | 0.702-0.865 | 0.766 (0.057) | 0.654-0.877 |
| Val | 0.717 (0.049) | 0.621-0.814 | 0.695 (0.071) | 0.556-0.833 |
| Met | 0.797 (0.041) | 0.718-0.877 | 0.745 (0.067) | 0.613-0.877 |
| Ile | 0.812 (0.039) | 0.736-0.887 | 0.827 (0.052) | 0.724-0.929 |
| Leu | 0.868 (0.033) | 0.803-0.933 | 0.857 (0.044) | 0.771-0.942 |
| Tyr | 0.833 (0.035) | 0.765-0.902 | 0.853 (0.043) | 0.769-0.936 |
| Phe | 0.802 (0.041) | 0.720-0.883 | 0.793 (0.059) | 0.677-0.910 |
| Lys | 0.804 (0.040) | 0.725-0.883 | 0.804 (0.055) | 0.697-0.912 |
| Arg | 0.772 (0.044) | 0.686-0.859 | 0.787 (0.066) | 0.657-0.917 |

AUC, area under the ROC curve; SE, the standard error of AUC; 95% CI, 95% confidence interval; GC, gastric cancer; EGC, early gastric cancer; GJFAA, gastric juice free amino acid.

separation of the samples. The column oven was maintained at 50°C. LC separation was performed using a mobile phase consisting of 0.1% formic acid in water (mobile phase A) and 0.1% formic acid in acetonitrile (mobile phase B). The gradient employed was as follows: 0-1 min maintained at 90% A, 1-12 min linear decrease from 90-30% A, 12-12.1 min decrease from 30%-0 A, and hold at 0% for 2.9 min, 15-15.1 min increase from 0-90% A, and hold at 90% A for 4.9 min. The flow rate was 1.0 mL/min, and the run time for one sample was 20.0 min.

The LC-MS/MS data acquisition for the above six amino acids was conducted in a positive ionization mode. The ion source parameters included were collision gas (CAD medium), curtain gas (CUR 20), ion source gas 1 (GS1 55), ion source gas 2 (GS2 60), and temperature (TEM 500°C). The product ions were monitored in a single reaction monitoring mode. Detection of other parameters were summarized in Table S3.

*GJFAA quantification*

Standard solutions that contained equivalent concentrations of the six amino acids were prepared by diluting the stock solution using deionized water at 2.5 μmol/L, 12.5 μmol/L, 25.0 μmol/L, 50.0 μmol/L, 100.0 μmol/L, and 200.0 μmol/L for building the calibration curves and were preserved at -20°C. Six-point calibration curves were used and plotted with the peak area ratio using a weighted ($1/x^2$) quadratic fit.

*Calibration curves, linearity, and retention time of the quantification system*

The calibration curves exhibited excellent linearity for the six amino acids standards, with low interference and high sensitivity. The retention time, linear range of concentration and linear fitting coefficients of them were concluded in Table S4. If some of them fell beyond the measurement range, the samples were re-diluted before injection. In extremely low concentration conditions that was undetectable, the value was replaced by zero.

*Helicobacter pylori infection detection*

*H. pylori* infection was determined by Warthin-Starry (WS) staining in all specimens using the *H. pylori* detection kit (Beijing ShiJi HeLi Biotechnology Co., Ltd, Beijing, China) according to the manufacturer's instructions. The background tissue was stained to yellow and the nucleus was dyed to brown. If there is *H.pylori* infection, the bacteria would be stained to black.

*Statistical analysis*

Firstly, multivariate analysis of pattern recognition was performed by using SIMCA 14.1 software package (MKS Data Analytics Solutions, Umea, Sweden), including principal component analysis (PCA) and orthogonal partial least squares discriminant analysis (OPLS-DA). Afterwards, a seven-fold cross-validation was performed to further validate the method, and the goodness of fit parameter ($R^2$) and the goodness of prediction parameter ($Q^2$) values were used to estimate the robustness and the predictive ability of the models. Secondly, univariate analysis was performed by using SPSS 20.0 software (SPSS Inc., Chicago, IL, USA). The concentrations of GJFAAs were express-

ed as median (interquartile range, IQR), and Mann-Whitney U test was used to assess the differences between the GC and NGD groups. *P*-values < 0.05 were considered as statistically significant. Furthermore, the criteria for selecting differential GJFAAs were a variable importance in the projection (VIP) value > 1, and a *P*-value of Mann-Whitney U test < 0.05 between the two groups [26]. Subsequently, the differential GJFAAs were cross-referenced to the pathways by further searching with a commercial database, KE-GG (http://www.kegg.jp/kegg/pathway.html) and a free and web-based tool, MetaboAnalyst (http://www.metaboanalyst.ca/MetaboAnalyst/) [27, 28].

## Results

No significant difference was found in the disease constitution between the training and validation sets (**Table 1**). The baseline characteristics of the study population are described in **Table 2**. No significant difference was observed in *H. pylori* infection rate, but a significant difference in male/female ratio existed between the GC and NGD groups both in the training set ($P = 0.005$) and validation set ($P = 0.006$). However, no significant difference was observed in age between the GC and NGD groups in the training set, but a significant difference existed in the validation set ($P = 0.007$).

### *Different GJFAA profiles between the GC and NGD groups in the training set*

First, significant alterations in GJFAA profiles were observed in GC patients when compared to NGD patients in the training set (Figure S1). Next, 26 out of 34 GJFAAs were extracted for subsequent analysis. The other 8 GJFAAs (alpha amino acid, citrulline, alpha aminobutyric acid, cystine, 3-methyl histidine, 1-methyl histidine and carnosine and hydroxyproline) were excluded as their median concentrations were 0. The data scale conversion mode used in both GC and NGD samples were log transformation processing and Ctr-formatted (Mean-Centered Scaling) processing. In addition, the missing values of raw data were filled up by half of the minimum values. Thereafter, the normalized data were imported into the SIMCA software for multivariate analysis.

### *Identification of GJFAA characteristic metabolic perturbation of GC patients in the training set*

Initially, PCA was performed using two principal components: $R^2X = 0.517$ and $Q^2 = 0.339$. The score plot (**Figure 2A**) demonstrated that each sample can be clearly divided, and no abnormal sample should be rejected. Certain samples that could not be clearly distinguished were further examined through subsequent discriminating analyses. Subsequently, a loading plot (**Figure 2B**) was constructed based on OPLS-DA model using one predictive and one orthogonal component ($R^2X = 0.455$, $R^2Y = 0.421$ and $Q^2Y = 0.279$). Using this model, the clearest separation point was produced between the GC and NGD groups. These values illustrated that huge variation in the statistical data was attributable to the separation between the two groups. Then a seven-fold cross-validation permutation test was applied (**Figure 2C**), where the $R^2$ and $Q^2$ intercept values were -0.25 and -0.57, respectively. These results excluded the random effects in the constructed model.

### *Screening out potential gastric juice biomarkers in the training set*

In total, 14 amino acids were screened as differential GJFAAs (**Table 3**). Among them, concentrations of 11 GJFAAs' were increased significantly, which included threonine, serine, alanine, valine, methionine, isoleucine, leucine, tyrosine, phenylalanine, lysine and arginine, respectively. The other three GJFAAs' concentrations were decreased significantly, which included phosphoserine, ethanolamine phosphate and urea, respectively. For quick visual identification of the varied GJFAAs, we displayed the statistical results in the form of a volcano plot as depicted in **Figure 2D**.

### *Diagnostic value analysis of 14 differential GJFAAs in the training set*

The area under the curve (AUC) of these differential GJFAAs ranged from 0.666 to 0.868 for the diagnosis of GC individually (**Table 4**). We created logistic regression model by binary logistic regression analysis in the training set. Zero (NGD group) or one (GC group) serves as dichotomous variable, and fourteen differential metabolites as the covariates. The predicted equation was as follows: $P = 28.869 * X_1 -$

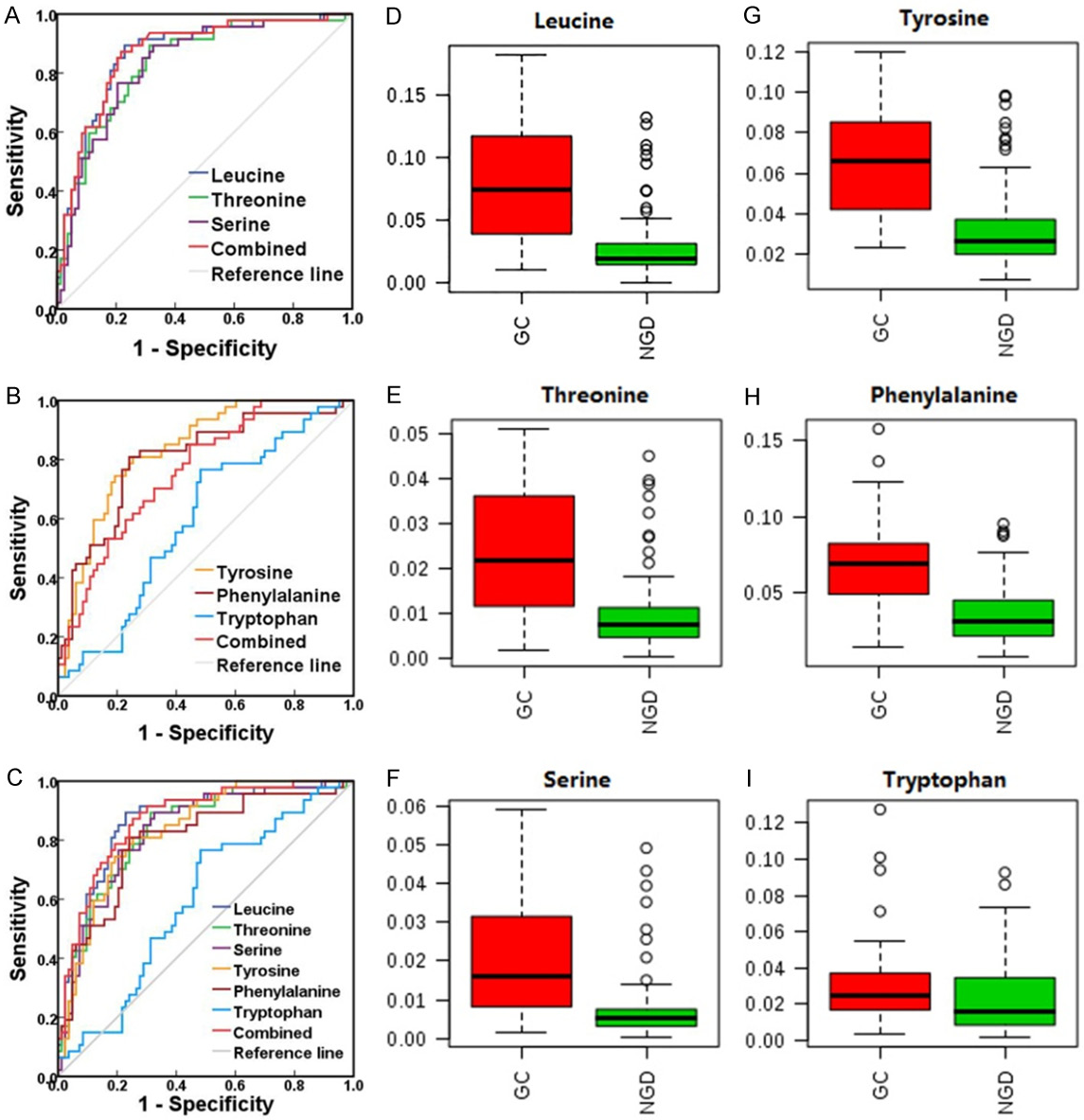


**Figure 3.** ROC curves for the detection of GC (A-C) and box plots of the levels of the six GJFAAs from the GC and NGD groups (D-I) in the training set. GC, gastric cancer; NGD, non-neoplastic gastric diseases; ROC, receiver operating characteristic curve.

25.640 * $X_2$ + 12.840 * $X_3$ + 135.106 * $X_4$ - 44.802 * $X_5$ + 83.488 * $X_6$ - 124.793 * $X_7$ - 1.921 * $X_8$ + 37.750 * $X_9$ + 70.130 * $X_{10}$ + 18.240 * $X_{11}$ + 46.077 * $X_{12}$ + 66.500 * $X_{13}$ - 54.718 * $X_{14}$ + 15.324 (Table S5). *P* referred to the value of predicted probability of each sample based on the levels of 14 candidate metabolites. We used the values of predicted probability as new variables and produced a combined ROC curve with a cutoff value of 30.011. The corresponding sensitivity was 0.851, specificity was 0.892 and accuracy was 0.877. The combined AUC was 0.902 (95% CI, 0.846-0.959) for the diagnosis of GC (**Figure 2E**). Importantly, their AUCs ranged from 0.649 to 0.857 (**Table 4**), and their combined AUCs reached up to 0.880 (95% CI, 0.792-0.969) for the diagnosis of EGC (**Figure 2F**). Particularly, leucine, threonine, and serine were the GJFAAs with the most variance between the two groups, whose fold change was more than 2 and AUC value was greater than 0.8. Moreover, combined AUC of the three non-AAAs [0.869 (95% CI, 0.805-0.934)] for diagnosing GC was

**Table 5.** Discriminating performance of individual and combined detection of the six GJFAAs for GC and EGC in the validation set

| GJFAA | GC | | EGC | |
|---|---|---|---|---|
| | AUC (SE) | 95% CI | AUC (SE) | 95% CI |
| Leucine | 0.872 (0.041) | 0.792-0.951 | 0.895 (0.060) | 0.778-1.000 |
| Threonine | 0.901 (0.036) | 0.831-0.972 | 0.897 (0.063) | 0.774-1.000 |
| Serine | 0.871 (0.043) | 0.786-0.955 | 0.866 (0.082) | 0.706-1.000 |
| Tyrosine | 0.838 (0.047) | 0.745-0.931 | 0.866 (0.072) | 0.726-1.000 |
| Phenylalanine | 0.830 (0.049) | 0.733-0.926 | 0.842 (0.077) | 0.691-0.993 |
| Tryptophan | 0.852 (0.045) | 0.764-0.941 | 0.867 (0.073) | 0.724-1.000 |
| Combined 3 non-AAAs | 0.900 (0.035) | 0.831-0.969 | 0.900 (0.056) | 0.789-1.000 |
| Combined 3 AAAs | 0.851 (0.046) | 0.761-0.941 | 0.866 (0.076) | 0.718-1.000 |
| Combined all | 0.914 (0.032) | 0.850-0.977 | 0.909 (0.050) | 0.812-1.000 |

AUC, area under the ROC curve; SE, the standard error of AUC; AAAs, aromatic amino acids; non-AAAs, non-aromatic amino acids; GJFAA, gastric juice free amino acid; GC, gastric cancer; EGC, early gastric cancer.

slightly higher than the combined three AAAs [0.841 (95% CI, 0.773-0.908)]. Combined detection of the above six GJFAAs could further improve its AUC value to 0.871 (95% CI, 0.809-0.933) (**Figure 3**).

*Validation of the six selected GJFAAs in the validation set*

In the validation set, the combined AUC of the three non-AAAs [0.900 (95% CI, 0.831-0.969)] was also superior to the combined three AAAs [0.851 (95% CI, 0.761-0.941)] in the diagnosis of GC. Combined AUC comprising of the six GJFAAs was 0.914 (95% CI, 0.850-0.977) for GC (**Table 5**; **Figure 4**). The diagnostic panel achieved 71.9% sensitivity, 97.4% specificity, and 85.7% accuracy for the prediction of GC (Table S6). Similarly, combined AUC of the three non-AAAs [0.900 (95% CI, 0.789-1.000)] was still superior over combined three AAAs [0.866 (95% CI, 0.718-1.000)] for the diagnosis of EGC. Combined AUC comprising of the six GJFAAs was 0.909 (95% CI, 0.812-1.000) for EGC (**Table 5**; **Figure 5**). The diagnostic panel achieved 72.7% sensitivity, 97.4% specificity, and 91.8% accuracy for the prediction of EGC (Table S6).

*Different metabolites related pathways between the GC and NGD groups in the training set*

The pathways that matched based on the KEGG database included aminoacyl-tRNA biosynthesis, biosynthesis of amino acids, and others. More details were described in Tables S7 and S8. Among these pathways, both aminoacyl-tRNA biosynthesis and biosynthesis of amino acids pathways included largest number of significantly varied GJFAAs, including all differential GJFAAs except for phosphor-ethanolamine and urea. To identify the pathway that is most relevant to GC development, metabolomics data were comprehensively analyzed using MetaboAnalyst 3.0, which combined pathway enrichment and topology analyses to process our screening data (Table S9). The enrichment analysis showed that the metabolic pathways with raw *P* values < 0.001 included aminoacyl-tRNA biosynthesis; valine, leucine and isoleucine biosynthesis; and cysteine and methionine metabolism. The raw *P* values < 0.001 indicated that the altered pathway significantly contributed to the difference between the two groups. The topology analysis showed that the impact value of the metabolic pathways of aminoacyl-tRNA biosynthesis; glycine serine and threonine metabolism; and phenylalanine metabolism are 0.113, 0.262 and 0.119, respectively. An impact value ≥ 0.1 indicated that the altered pathway affected GC carcinogenesis. Further, a Google map-style interactive visualization system was used to generate **Figure 6** to present the pathway analysis results in an intuitive manner.

## Discussion

*Utility of amino acid analytics in diagnosis of GC*

Employing amino acid metabolomics for diagnosing GC had been discussed by many researchers via various types of biofluids and tissue extracts [29-34]. For example, Chen [32] and colleagues used gas chromatography combined with mass spectrometer for fingerprinting urinary metabolites on 293 urine samples to find practical and cost-effective biomarkers. As a result, 17 metabolites are significantly different between patients and healthy controls in

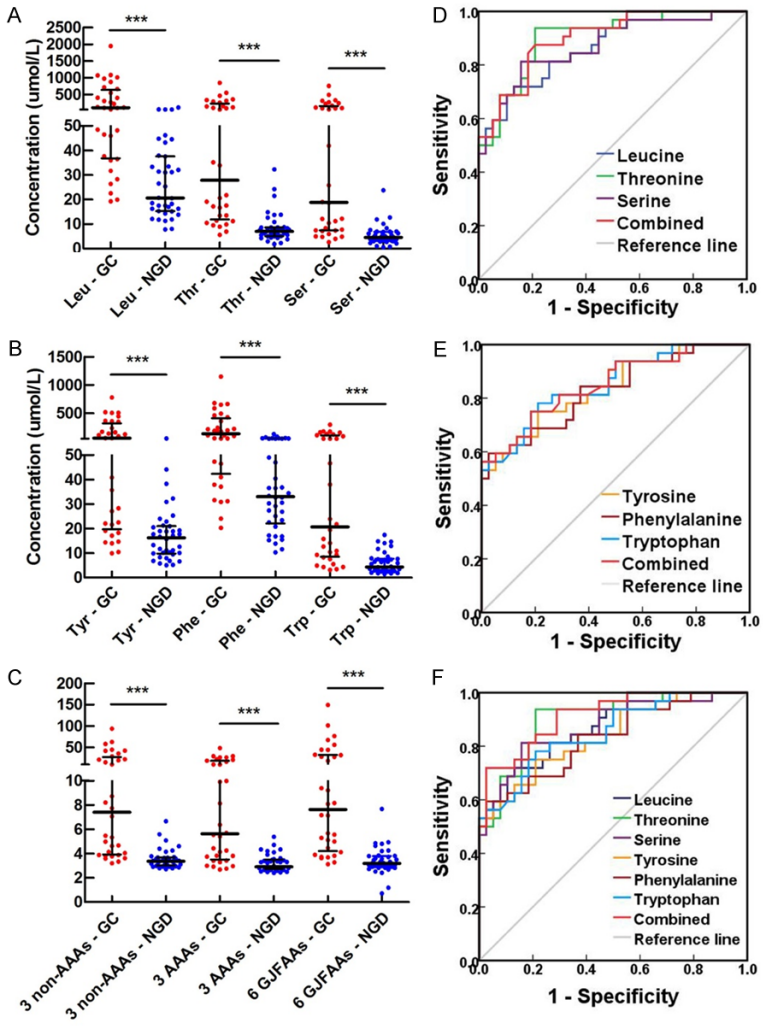


**Figure 4.** Scatter plots of the levels of the six GJFAAs from the GC and NGD patients (A-C) and ROC curves for the detection of GC (D-F) in the validation set. GC, gastric cancer; NGD, non-neoplastic gastric diseases; ROC, receiver operating characteristic curve.

the training set. Among them, 14 metabolites emerged diagnostic value better than traditional serum biomarkers by quantitative assay of the validation set. Of these, 10 are amino acids and four are organic metabolites. Their AUC values ranged between 0.669-0.823, and the combination AUC can reach up to 0.893. The diagnostic efficacy displayed in the study was consistent with ours. In our study, the AUC of 14 differential GJFAAs was between 0.666-0.868, and their combined AUC was 0.902, which was slightly higher than that reported by Chen's study. This may be related to the tissue specificity of gastric juice which was higher than urine.

Wang [33] and colleagues pointed out the metabolic profiling of tissue samples on a large cohort of human GC subjects (n = 125) and normal controls (n = 54) using (1) H-NMR along with multivariate analyses. 48 endogenous distinguishing metabolites (VIP > 1 and $P < 0.05$) were identified altogether. These modified metabolites revealed disturbances of amino acids, glutaminolysis, glycolysis, tricarboxylic acid cycle and choline metabolism, which were correlated with the growth and progression of GC. The AUC value of OPLS-DA model between the neoplastic and normal tissues was 0.945. Although its predictive power was a little higher than ours, the number of distinguishing metabolites selected by OPLS-DA analysis of our study was much less than that of the study by Wang et al. This is probably caused by our study mainly emphasizing amino acid metabolism, instead of global metabolomic profiling. However, the detection method of our study was relatively simple, the expense was comparatively low, and the sample was substantially rich in content and reproducible.

Additionally, PFAA levels from 56 GC patients and 137 age-matched healthy controls were monitored by Gu [17] and colleagues using amino acid analyzer. In contrast to the healthy control group, the concentrations of threonine, arginine and cysteine were increased significantly in the GC group, and the concentrations of 9 amino acids, such as aspartic acid, glutamic acid, serine, etc., were decreased significantly. Instead, the concentrations of 34 GJFAAs were measured by amino acid analyzer from 47 GC patients and 83 NGD patients in the training set of this study. Among the 14 differential GJFAAs, the levels of 11 amino acids, such as threonine, serine, leucine, etc. were increased significantly in the GC group, and the levels of phosphoserine, ethanolamine phosphate and urea were decreased significantly. This elucidates that a changing concentration trend between the majority of

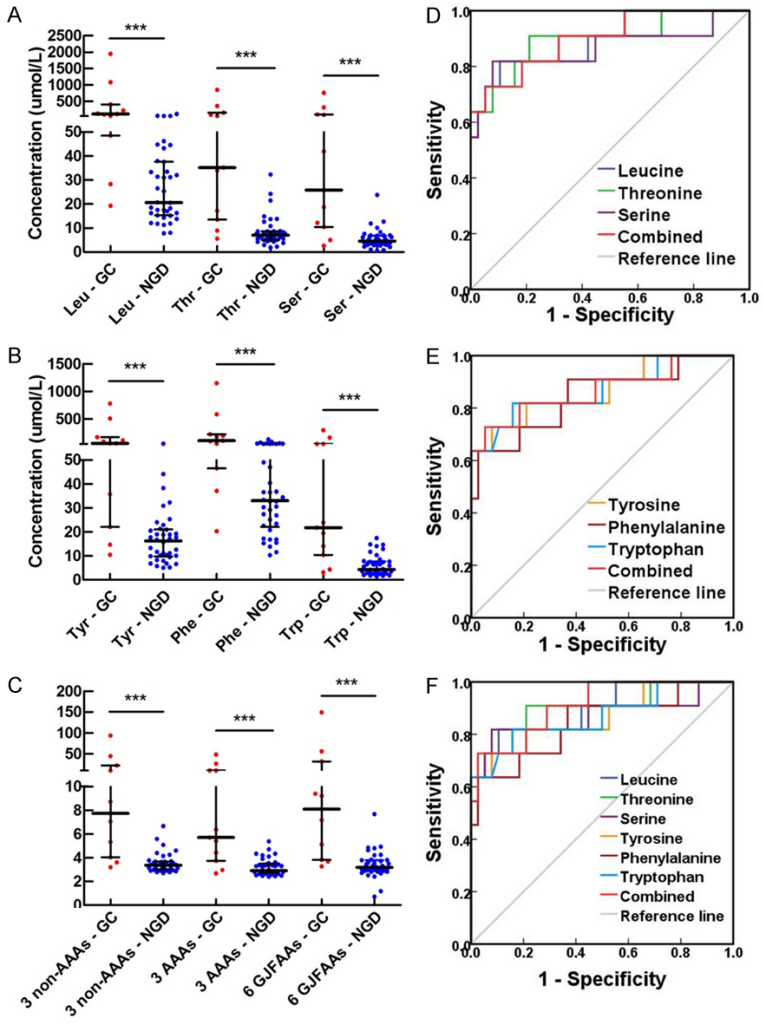


**Figure 5.** Scatter plots of the levels of the six GJFAAs from the EGC and NGD patients (A-C) and ROC curves for the detection of EGC (D-F) in the validation set. EGC, early gastric cancer; NGD, non-neoplastic gastric diseases; ROC, receiver operating characteristic curve.

PFAAs and GJFAAs in GC participants was just the opposite. The abnormal accumulation of "useful" metabolites in tumor microenvironment and their intensive metabolism in GC tissues might explain the inconsistency of the metabolic phenotypes between the plasma and gastric juice in GC sufferers.

*Screening of differential GJFAAs and their biological significance*

To facilitate the clinical application, 14 differential GJFAAs were further screened according to the FC and AUC values. Leucine, threonine, serine, isoleucine, lysine, and urea are the six FAAs that meet both FC > 2.0 and AUC > 0.8 values, and are expected to be the candidate biomarkers for diagnosing GC. We chose the first three for validation, compared and combined their diagnostic capabilities to three AAAs we had concentrated on. More concretely, the combined AUC of the three non-AAAs was superior to the combined three AAAs for diagnosing GC or even EGC in the validation set, too. And the combination of AUC comprised of the above six GJFAAs can be further improved. These findings suggest that this integrated diagnostic panel has great potential for improving early detection of GC.

The biological significance for the tremendous increase of the three non-AAAs' levels in gastric juice of GC patients were as follows. Leucine belongs to the branched chain amino acids, which can promote protein synthesis and inhibit the decomposition, thereby elevated leucine in gastric juice reflects the metabolic pathway of protein biosynthesis was remarkably active in GC tissues. Threonine is a kind of carbohydrate amino acid. Owing to a large amount of carbohydrate is required to meet the energy metabolism of GC tissues, elevated threonine in gastric juice implies the unusual activation of gluconeogenesis in GC tissues. Serine catabolism produces one carbon unit, which can be used as raw material for the synthesis of purines and pyrimidines, so elevated serine in gastric juice represents the tremendously active metabolism of nucleic acids in GC tissues.

*Metabolic pathway analysis of differential GJFAAs*

The metabolic pathway of aminoacyl-tRNA biosynthesis was excessively activated, which was significantly responsible for the metabolic modification in GC. It unveiled that the differences of GJFAA profiles between the GC and NGD groups might be closely related to the dra-

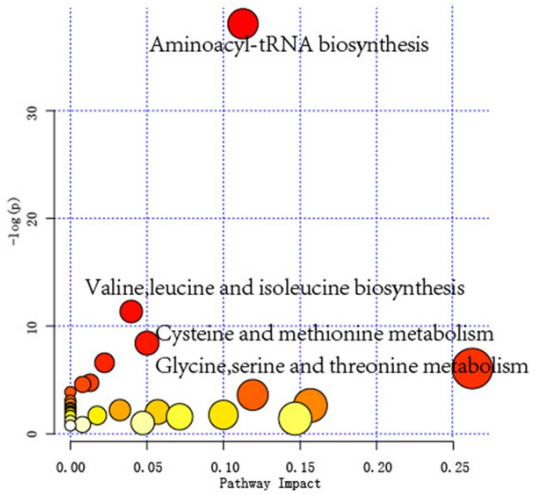


**Figure 6.** Pathway analysis bubble plot for group GC vs. NGD in the training set. GC, gastric cancer; NGD, non-neoplastic gastric diseases.

matic activity in cellular protein biosynthesis during the rapid proliferation of the malignant tumor cells. Another three metabolic pathways are concurrent between the eight main metabolic pathways in this study that are selected from 14 differential GJFAAs and six main metabolic pathways in Chen's study [32] that are selected from 14 differential metabolites in urine, including glycine, serine and threonine metabolism, cystine and methionine metabolism and valine, leucine and isoleucine biosynthesis. This implied that the metabolic patterns of some small molecular substances (especially amino acids) in microenvironment-gastric juice and macroenvironment-urine are more or less the same in the process of gastric carcinogenesis. The metabolic studies based on gastric juice and urine samples manifested a great value for the diagnosis of GC. In addition, combined detection of these might complement each other, as well as enrich our understanding concerning the pathogenesis of GC.

In conclusion, our results suggest that GJFAA profiling might be an ideal tool for improving the prognosis of GC via its early diagnosis and treatment. It also brings deeper insights into the oncogenesis of GC, which will undoubtedly contribute to proposing novel therapeutic targets. Meanwhile, joint examination of the three non-AAAs (leucine, threonine and serine) and three AAAs in gastric juice could further upgrade the diagnostic level of GC especially EGC.

## Acknowledgements

The authors thank Wei-Xiang Zhao (Feed Research Institute, Chinese Academy of Agricultural Sciences, Beijing, China) for assisting in the quantitative determination of the gastric juice specimens, Huan-Huan Qin (Biotree Biotechnology Co., Ltd., Shanghai, China) and Yan Li (Fanxing Bio-technology Co., Ltd., Beijing, China) for providing helps in data analysis. This work was supported by the National Natural Science Foundation of China (No. 81672410) and Key Laboratory for Helicobacter pylori Infection and Upper Gastrointestinal Diseases in Beijing (No. BZ0371).

## Disclosure of conflict of interest

None.

**Address correspondence to:** Dr. Liya Zhou, Department of Gastroenterology, Peking University Third Hospital, 49 North Garden Road, Haidian District, Beijing 100191, China. Tel: +86-10-82266719; Fax: +86-10-62034716; E-mail: zhoumed@126.com

## References

[1] Colquhoun A, Arnold M, Ferlay J, Goodman KJ, Forman D and Soerjomataram I. Global patterns of cardia and non-cardia gastric cancer incidence in 2012. Gut 2015; 64: 1881-1888.

[2] Chadwick G, Groene O, Riley S, Hardwick R, Crosby T, Hoare J, Hanna GB, Greenaway K and Cromwell DA. Gastric cancers missed during endoscopy in England. Clin Gastroenterol Hepatol 2015; 13: 1264-1270, e1261.

[3] He CZ, Zhang KH, Li Q, Liu XH, Hong Y and Lv NH. Combined use of AFP, CEA, CA125 and CA19-9 improves the sensitivity for the diagnosis of gastric cancer. BMC Gastroenterol 2013; 13: 87.

[4] Xiao S and Zhou L. Gastric cancer: metabolic and metabolomics perspectives (review). Int J Oncol 2017; 51: 5-17.

[5] Lario S, Ramírez-Lázaro MJ, Sanjuan-Herráez D, Brunet-Vega A, Pericay C, Gombau L, Junquera F, Quintás G and Calvet X. Plasma sample based analysis of gastric cancer progression using targeted metabolomics. Sci Rep 2017; 7: 17774.

[6] Manig F, Kuhne K, von Neubeck C, Schwarzenbolz U, Yu Z, Kessler BM, Pietzsch J and Kunz-Schughart LA. The why and how of amino acid analytics in cancer diagnostics and therapy. J Biotechnol 2017; 242: 30-54.

[7] Siminska E and Koba M. Amino acid profiling as a method of discovering biomarkers for ear-

ly diagnosis of cancer. Amino Acids 2016; 48: 1339-1345.
[8] Kim HJ, Jang SH, Ryu JS, Lee JE, Kim YC, Lee MK, Jang TW, Lee SY, Nakamura H, Nishikata N, Mori M, Noguchi Y, Miyano H and Lee KY. The performance of a novel amino acid multivariate index for detecting lung cancer: a case control study in Korea. Lung Cancer 2015; 90: 522-527.
[9] Barnes T, Bell K, DiSebastiano KM, Vance V, Hanning R, Russell C, Dubin JA, Bahl M, Califaretti N, Campbell C and Mourtzakis M. Plasma amino acid profiles of breast cancer patients early in the trajectory of the disease differ from healthy comparison groups. Appl Physiol Nutr Metab 2014; 39: 740-744.
[10] Hasim A, Aili A, Maimaiti A, Mamtimin B, Abudula A and Upur H. Plasma-free amino acid profiling of cervical cancer and cervical intraepithelial neoplasia patients and its application for early detection. Mol Biol Rep 2013; 40: 5853-5859.
[11] Zhang F, Ma X, Li H, Guo G, Li P, Li H, Gu L, Li X, Chen L and Zhang X. The predictive and prognostic values of serum amino acid levels for clear cell renal cell carcinoma. Urol Oncol 2017; 35: 392-400.
[12] Fukutake N, Ueno M, Hiraoka N, Shimada K, Shiraishi K, Saruki N, Ito T, Yamakado M, Ono N, Imaizumi A, Kikuchi S, Yamamoto H and Katayama K. A novel multivariate index for pancreatic cancer detection based on the plasma free amino acid profile. PLoS One 2015; 10: e0132223.
[13] Ma H, Hasim A, Mamtimin B, Kong B, Zhang HP and Sheyhidin I. Plasma free amino acid profiling of esophageal cancer using high-performance liquid chromatography spectroscopy. World J Gastroenterol 2014; 20: 8653-8659.
[14] Ikeda A, Nishiumi S, Shinohara M, Yoshie T, Hatano N, Okuno T, Bamba T, Fukusaki E, Takenawa T, Azuma T and Yoshida M. Serum metabolomics as a novel diagnostic approach for gastrointestinal cancer. Biomed Chromatogr 2012; 26: 548-558.
[15] Leichtle AB, Nuoffer JM, Ceglarek U, Kase J, Conrad T, Witzigmann H, Thiery J and Fiedler GM. Serum amino acid profiles and their alterations in colorectal cancer. Metabolomics 2012; 8: 643-653.
[16] Miyagi Y, Higashiyama M, Gochi A, Akaike M, Ishikawa T, Miura T, Saruki N, Bando E, Kimura H, Imamura F, Moriyama M, Ikeda I, Chiba A, Oshita F, Imaizumi A, Yamamoto H, Miyano H, Horimoto K, Tochikubo O, Mitsushima T, Yamakado M and Okamoto N. Plasma free amino acid profiling of five types of cancer patients and its application for early detection. PLoS One 2011; 6: e24143.
[17] Gu Y, Chen T, Fu S, Sun X, Wang L, Wang J, Lu Y, Ding S, Ruan G, Teng L and Wang M. Perioperative dynamics and significance of amino acid profiles in patients with cancer. J Transl Med 2015; 13: 35.
[18] Zhou LY, Lin SR, Li Y, Geng QM, Ding SG, Meng LM, Sun XN, Wang JT, Xu ZJ, Guo HL, Li CG and Cui RL. The intrinsic fluorescence spectrum of dilute gastric juice as a novel diagnostic tool for gastric cancer. J Dig Dis 2011; 12: 279-285.
[19] Deng K, Zhou LY, Lin SR, Li Y, Chen M, Geng QM and Li YW. A novel approach for the detection of early gastric cancer: fluorescence spectroscopy of gastric juice. J Dig Dis 2013; 14: 299-304.
[20] Deng K, Lin S, Zhou L, Geng Q, Li Y, Xu M and Na R. Three aromatic amino acids in gastric juice as potential biomarkers for gastric malignancies. Anal Chim Acta 2011; 694: 100-107.
[21] Deng K, Lin S, Zhou L, Li Y, Chen M, Wang Y and Li Y. High levels of aromatic amino acids in gastric juice during the early stages of gastric cancer progression. PLoS One 2012; 7: e49434.
[22] Liu J, Lin S, Li Z, Zhou L, Xue Y, Yan X, Meng L, Lu J and Suo B. A novel gastric juice index model for detecting early gastric cancer. Int J Clin Exp Med 2017; 10: 14425-14435.
[23] Sun Z, Wang ZN, Zhu Z, Xu YY, Xu Y, Huang BJ, Zhu GL and Xu HM. Evaluation of the seventh edition of American joint committee on cancer TNM staging system for gastric cancer: results from a Chinese monoinstitutional study. Ann Surg Oncol 2012; 19: 1918-1927.
[24] Rugge M, Correa P, Dixon MF, Hattori T, Leandro G, Lewin K, Riddell RH, Sipponen P and Watanabe H. Gastric dysplasia: the Padova international classification. Am J Surg Pathol 2000; 24: 167-176.
[25] Dixon MF, Genta RM, Yardley JH and Correa P. Classification and grading of gastritis. The updated sydney system. International workshop on the histopathology of gastritis, houston 1994. Am J Surg Pathol 1996; 20: 1161-1181.
[26] Saccenti E, Hoefsloot HCJ, Smilde AK, Westerhuis JA and Hendriks MMWB. Reflections on univariate and multivariate analysis of metabolomics data. Metabolomics 2014; 10: 361-374.
[27] Kanehisa M, Sato Y, Kawashima M, Furumichi M and Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 2016; 44: D457-D462.
[28] Xia JG, Sinelnikov IV, Han B and Wishart DS. MetaboAnalyst 3.0-making metabolomics more meaningful. Nucleic Acids Res 2015; 43: W251-W257.
[29] Yu L, Aa J, Xu J, Sun M, Qian S, Cheng L, Yang S and Shi R. Metabolomic phenotype of gastric cancer and precancerous stages based on gas chromatography time-of-flight mass spectrom-

etry. J Gastroenterol Hepatol 2011; 26: 1290-1297.
[30] Fan J, Hong J, Hu JD and Chen JL. Ion chromatography based urine amino acid profiling applied for diagnosis of gastric cancer. Gastroenterol Res Pract 2012; 2012: 474907.
[31] Aa JY, Yu LZ, Sun M, Liu LS, Li MJ, Cao B, Shi J, Xu J, Cheng LP, Zhou J, Zheng T, Wang XW, Zhao CY, Gu RR, Zhang FY, Shi RH and Wang GJ. Metabolic features of the tumor microenvironment of gastric cancer and the link to the systemic macroenvironment. Metabolomics 2012; 8: 164-173.
[32] Chen Y, Zhang J, Guo L, Liu L, Wen J, Xu L, Yan M, Li Z, Zhang X, Nan P, Jiang J, Ji J, Zhang J, Cai W, Zhuang H, Wang Y, Zhu Z and Yu Y. A characteristic biosignature for discrimination of gastric cancer from healthy population by high throughput GC-MS analysis. Oncotarget 2016; 7: 87496-87510.
[33] Wang HJ, Zhang HL, Deng PC, Liu CQ, Li DD, Jie H, Zhang H, Zhou ZG and Zhao YL. Tissue metabolic profiling of human gastric cancer assessed by H-1 NMR. BMC Cancer 2016; 16: 371
[34] Chan AW, Mercier P, Schiller D, Bailey R, Robbins S, Eurich DT, Sawyer MB and Broadhurst D. (1)H-NMR urinary metabolomic profiling for diagnosis of gastric cancer. Br J Cancer 2016; 114: 59-62.

**Table S1.** Amino acid standard solution list

| Amino acid | Compound | Abbreviation |
|---|---|---|
| AA01 | O-Phospho-L-Serine | PSer |
| AA02 | Taurine | Tau |
| AA03 | O-Phospho-Ethanolamine | PEtN |
| AA04 | Urea | Urea |
| AA05 | L-Aspartic acid | Asp |
| AA06 | L-Threonine | Thr |
| AA07 | L-Serine | Ser |
| AA08 | L-Asparagine | Asn |
| AA09 | L-Glutamic acid | Glu |
| AA10 | L-α-Amino-Adipic Acid | Aad |
| AA11 | L-Glycine | Gly |
| AA12 | L-Alanine | Ala |
| AA13 | L-Citrulline | Cit |
| AA14 | L-α-Amino-n-Butyric Acid | Abu |
| AA15 | L-Valine | Val |
| AA16 | L-Cystine | Cys-Cys |
| AA17 | L-Methionine | Met |
| AA18 | L-Isoleucine | Ile |
| AA19 | L-Leucine | Leu |
| AA20 | L-Tyrosine | Tyr |
| AA21 | L-Phenylalanine | Phe |
| AA22 | β-Alanine | bAla |
| AA23 | D, L-β-Amino-Isobutyric Acid | bAib |
| AA24 | γ-Amino-n-Butyric Acid | GABA |
| AA25 | L-Histidine | His |
| AA26 | 3-Methyl-L-Histidine | 3MHis |
| AA27 | 1-Methyl-L-Histidine | 1MHis |
| AA28 | L-Carnosine | Car |
| AA29 | L-Tryptophan | Trp |
| AA30 | L-Ornithine | Orn |
| AA31 | L-Lysine | Lys |
| AA32 | L-Arginine | Arg |
| AA33 | L-Proline | Pro |
| AA34 | L-Hydroxyproline | Hyp |

**Table S2.** Gradient elution program steps of amino acid analyzer

| RT (min) | Buffer A (%) | Buffer B (%) | Buffer C (%) | Diluent (%) |
|---|---|---|---|---|
| 0 | 100 | 0 | 0 | 0 |
| 10 | 100 | 0 | 0 | 0 |
| 11 | 79 | 21 | 0 | 0 |
| 30 | 79 | 21 | 0 | 0 |
| 41 | 62 | 38 | 0 | 0 |
| 63 | 0 | 0 | 100 | 0 |
| 68 | 0 | 0 | 100 | 0 |
| 78 | 0 | 0 | 100 | 0 |
| 81 | 0 | 0 | 86 | 14 |
| 83 | 0 | 0 | 78 | 22 |
| 95 | 0 | 0 | 76 | 24 |
| 102 | 0 | 0 | 0 | 100 |
| 102.1 | 0 | 0 | 0 | 100 |
| 106.4 | 0 | 0 | 0 | 100 |
| 106.5 | 100 | 0 | 0 | 0 |
| 129.8 | 100 | 0 | 0 | 0 |

RT, retention time; Buffer A, B and C was lithium citrate buffer solution at pH 2.90, 4.20 and 8.00, respectively. Additionally, diluent was lithium citrate buffer solution with pH 2.20.

**Table S3.** Multiple reaction monitoring (MRM) conditions for LC-MS/MS detection of the six amino acids

| Compound | Precursor ion (m/z) | Product ion (m/z) | Declustering potential (DP, V) | Entrance potential (EP, V) | Collision energy (CE, V) | Collision cell potential (CXP, V) |
|---|---|---|---|---|---|---|
| Serine | 276.1 | 171.0 | 30 | 10 | 20 | 2 |
| Threonine | 290.1 | 171.0 | 30 | 10 | 20 | 2 |
| Leucine | 302.2 | 171.0 | 35 | 10 | 25 | 2 |
| Phenylalanine | 336.2 | 171.0 | 40 | 10 | 30 | 2 |
| Tyrosine | 352.2 | 171.0 | 35 | 10 | 30 | 2 |
| Tryptophan | 375.2 | 171.0 | 40 | 10 | 30 | 2 |

LC-MS/MS, liquid chromatography-tandem mass spectrometry.

**Table S4.** Retention times, linear range of concentration, and linear fitting coefficient of the six amino acid standards

| Compound | Retention time (min) | Linear range (μmol/L) | $r^2$ |
|---|---|---|---|
| Serine | 3.73 | 2.5-200 | 0.9992 |
| Threonine | 5.63 | 2.5-200 | 0.9997 |
| Leucine | 12.35 | 2.5-200 | 0.9987 |
| Phenylalanine | 12.75 | 1.25-100 | 0.9998 |
| Tyrosine | 10.61 | 1.25-100 | 1.0000 |
| Tryptophan | 12.98 | 1.25-100 | 0.9997 |

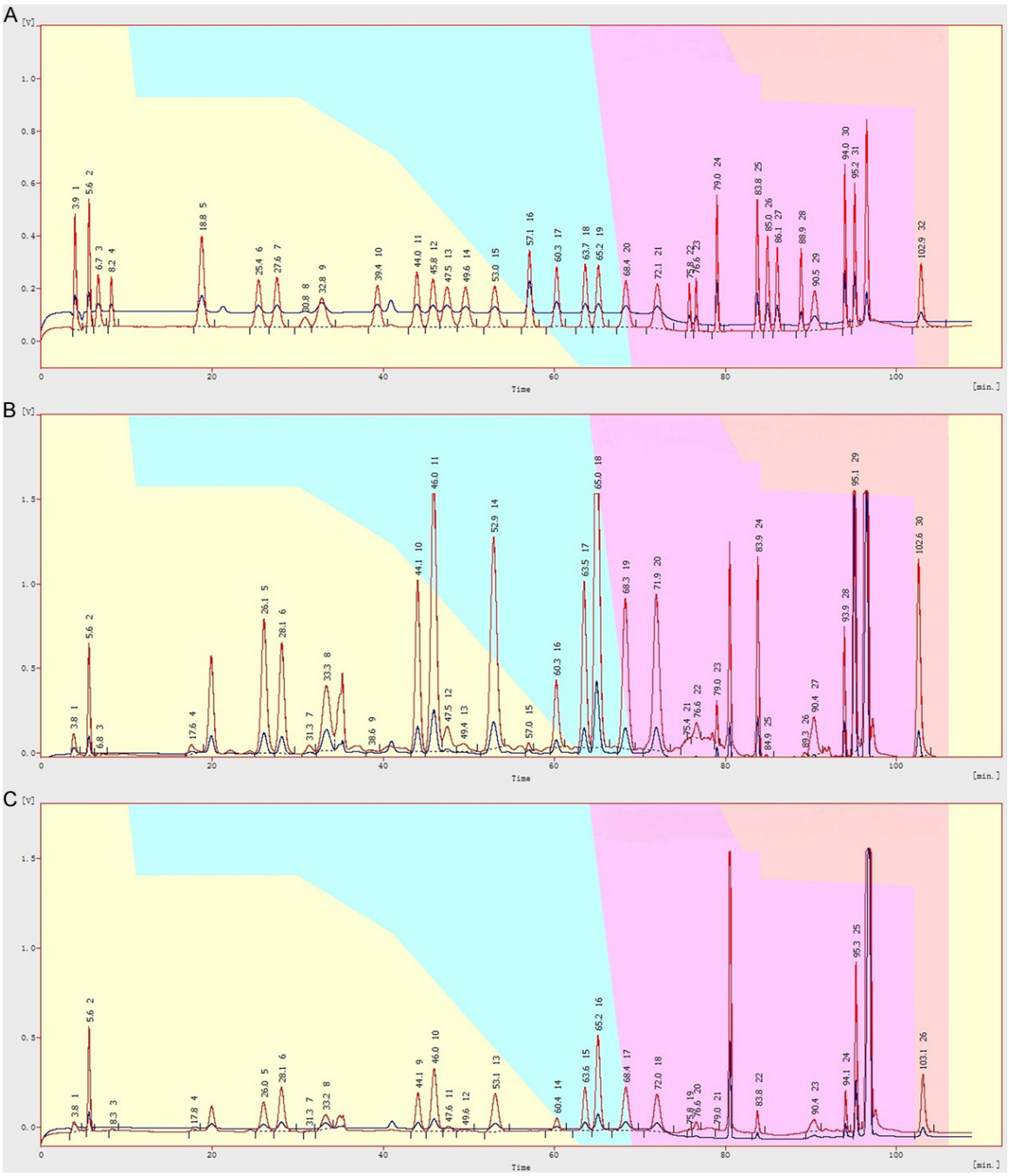


**Figure S1.** Typical chromatograms for 34 amino acids and biogenic amines from amino acid standard solution (A) and gastric juice samples from a GC patient (B) and a NGD patient (C). GC, gastric cancer; NGD, non-neoplastic gastric disease.

**Table S5.** Logistic regression model for the diagnosis of GC in the training set

| Parameter | B | S.E. | Sig. |
|---|---|---|---|
| $X_1$ PSer | -28.869 | 33.959 | 0.395 |
| $X_2$ PEtN | 25.640 | 35.173 | 0.466 |
| $X_3$ Urea | -12.840 | 6.885 | 0.062 |
| $X_4$ Thr | -135.106 | 99.257 | 0.173 |
| $X_5$ Ser | 44.802 | 68.337 | 0.512 |
| $X_6$ Ala | -83.488 | 53.484 | 0.119 |
| $X_7$ Val | 124.793 | 52.525 | 0.018 |
| $X_8$ Met | 1.921 | 71.182 | 0.978 |
| $X_9$ Ile | -37.750 | 44.964 | 0.401 |
| $X_{10}$ Leu | -70.130 | 47.080 | 0.136 |
| $X_{11}$ Tyr | -18.240 | 43.335 | 0.674 |
| $X_{12}$ Phe | -46.077 | 37.167 | 0.215 |
| $X_{13}$ Lys | -66.500 | 45.028 | 0.140 |
| $X_{14}$ Arg | 54.718 | 31.530 | 0.083 |
| Constant | 15.324 | 7.112 | 0.031 |

S.E., standard error; Sig., probability; GC, gastric cancer.

**Table S6.** Diagnostic performance evaluation of individual and combined detection of the six GJFAAs for GC and EGC in the validation set

| GJFAA | Cutoff value | | Sensitivity (%) | | Specificity (%) | | Accuracy (%) | |
|---|---|---|---|---|---|---|---|---|
| | GC | EGC | GC | EGC | GC | EGC | GC | EGC |
| Leucine | 45.40 | 47.35 | 71.9 | 81.8 | 86.8 | 89.5 | 80.0 | 87.8 |
| Threonine | 8.81 | 8.81 | 93.8 | 90.9 | 78.9 | 78.9 | 85.7 | 81.6 |
| Serine | 7.07 | 10.25 | 81.3 | 81.8 | 84.2 | 92.1 | 82.9 | 89.8 |
| Tyrosine | 21.45 | 34.00 | 75.0 | 72.7 | 78.9 | 92.1 | 77.1 | 87.8 |
| Phenylalanine | 93.65 | 93.65 | 59.4 | 63.6 | 97.4 | 97.4 | 80.0 | 89.8 |
| Tryptophan | 8.05 | 10.35 | 78.1 | 81.8 | 78.9 | 84.2 | 78.6 | 83.7 |
| Combined 3 non-AAAs | 3.75 | 5.21 | 87.5 | 72.7 | 78.9 | 94.7 | 82.9 | 89.8 |
| Combined 3 AAAs | 3.68 | 4.39 | 75.0 | 72.7 | 81.6 | 94.7 | 78.6 | 89.8 |
| Combined all | 4.97 | 5.03 | 71.9 | 72.7 | 97.4 | 97.4 | 85.7 | 91.8 |

Cutoff value, optimized cutoff points were the values yielding maximum sums of sensitivity and specificity from the ROC curves; AAAs, aromatic amino acids; non-AAAs, non-aromatic amino acids; GJFAA, gastric juice free amino acid; GC, gastric cancer; EGC, early gastric cancer.

**Table S7.** Mapping information of the altered GJFAAs between the GC and NGD groups in the training set

| Query | Match | HMDB | PubChem | KEGG | Comment |
|---|---|---|---|---|---|
| AA01 | DL-O-Phosphoserine | HMDB01721 | 106 | C01005 | 1 |
| AA03 | O-Phosphoethanolamine | HMDB00224 | 1015 | C00346 | 1 |
| AA04 | Urea | HMDB00294 | 1176 | C00086 | 1 |
| AA06 | L-Threonine | HMDB00167 | 6288 | C00188 | 1 |
| AA07 | L-Serine | HMDB00187 | 5951 | C00065 | 1 |
| AA12 | L-Alanine | HMDB00161 | 5950 | C00041 | 1 |
| AA15 | L-Valine | HMDB00883 | 6287 | C00183 | 1 |
| AA17 | L-Methionine | HMDB00696 | 6137 | C00073 | 1 |
| AA18 | L-Isoleucine | HMDB00172 | 6306 | C00407 | 1 |
| AA19 | L-Leucine | HMDB00687 | 6106 | C00123 | 1 |
| AA20 | L-Tyrosine | HMDB00158 | 6057 | C00082 | 1 |
| AA21 | L-Phenylalanine | HMDB00159 | 6140 | C00079 | 1 |
| AA31 | L-Lysine | HMDB00182 | 5962 | C00047 | 1 |
| AA32 | L-Arginine | HMDB00517 | 6322 | C00062 | 1 |

Query: substance for metabolic mapping; Match: metabolite name; HMDB: compound ID in HMDB database; PubChem: compound ID in PubChem database; KEGG: compound ID in KEGG database; Comment: matching degree. 0 represents no match, 1 exact match, and 2 fuzzy match. HMDB, human metabolome database.

**Table S8.** Representative influenced pathways searched in KEGG database

| KEGG Pathway | Compounds |
|---|---|
| hsa01100 Metabolic pathways - Homo sapiens (human) (14) | cpd:C00041 L-Alanine cpd:C00047 L-Lysine cpd:C00062 L-Arginine cpd:C00065 L-Serine cpd:C00073 L-Methionine cpd:C00079 L-Phenylalanine cpd:C00082 L-Tyrosine cpd:C00086 Urea cpd:C00123 L-Leucine cpd:C00183 L-Valine cpd:C00188 L-Threonine cpd:C00346 Ethanolamine phosphate cpd:C00407 L-Isoleucine cpd:C01005 O-Phospho-L-serine |
| hsa00970 Aminoacyl-tRNA biosynthesis - Homo sapiens (human) (12) | cpd:C00041 L-Alanine cpd:C00047 L-Lysine cpd:C00062 L-Arginine cpd:C00065 L-Serine cpd:C00073 L-Methionine cpd:C00079 L-Phenylalanine cpd:C00082 L-Tyrosine cpd:C00123 L-Leucine cpd:C00183 L-Valine cpd:C00188 L-Threonine cpd:C00407 L-Isoleucine cpd:C01005 O-Phospho-L-serine |
| hsa01230 Biosynthesis of amino acids - Homo sapiens (human) (12) | cpd:C00041 L-Alanine cpd:C00047 L-Lysine cpd:C00062 L-Arginine cpd:C00065 L-Serine cpd:C00073 L-Methionine cpd:C00079 L-Phenylalanine cpd:C00082 L-Tyrosine cpd:C00123 L-Leucine cpd:C00183 L-Valine cpd:C00188 L-Threonine cpd:C00407 L-Isoleucine cpd:C01005 O-Phospho-L-serine |
| hsa04974 Protein digestion and absorption - Homo sapiens (human) (11) | cpd:C00041 L-Alanine cpd:C00047 L-Lysine cpd:C00062 L-Arginine cpd:C00065 L-Serine cpd:C00073 L-Methionine cpd:C00079 L-Phenylalanine cpd:C00082 L-Tyrosine cpd:C00123 L-Leucine cpd:C00183 L-Valine cpd:C00188 L-Threonine cpd:C00407 L-Isoleucine |
| hsa02010 ABC transporters - Homo sapiens (human) (10) | cpd:C00041 L-Alanine cpd:C00047 L-Lysine cpd:C00062 L-Arginine cpd:C00065 L-Serine cpd:C00079 L-Phenylalanine cpd:C00086 Urea cpd:C00123 L-Leucine cpd:C00183 L-Valine cpd:C00188 L-Threonine cpd:C00407 L-Isoleucine |

**Table S9.** Comprehensive analysis of differential amino acid pathways

| Pathway | Total | Hits | Raw p | -log(p) | Holm adjust | FDR | Impact |
|---|---|---|---|---|---|---|---|
| Aminoacyl-tRNA biosynthesis | 75 | 12 | 2.95E-17 | 38.063 | 2.36E-15 | 2.36E-15 | 0.113 |
| Valine, leucine and isoleucine biosynthesis | 27 | 4 | 1.17E-05 | 11.359 | 0.001 | < 0.001 | 0.040 |
| Cysteine and methionine metabolism | 56 | 4 | < 0.001 | 8.416 | 0.017 | 0.006 | 0.050 |
| Valine, leucine and isoleucine degradation | 40 | 3 | 0.001 | 6.597 | 0.105 | 0.027 | 0.022 |
| Glycine, serine and threonine metabolism | 48 | 3 | 0.002 | 6.065 | 0.177 | 0.037 | 0.262 |
| Sphingolipid metabolism | 25 | 2 | 0.009 | 4.741 | 0.655 | 0.116 | 0.013 |
| Phenylalanine, tyrosine and tryptophan biosynthesis | 27 | 2 | 0.010 | 4.590 | 0.751 | 0.116 | 0.008 |
| Phenylalanine metabolism | 45 | 2 | 0.027 | 3.613 | 1.000 | 0.240 | 0.119 |

FDR, false discovery rate.