Cysteinylglycine for potential diagnosis of extrahepatic cholangiocarcinoma using a novel metabolomic approach

Longjiu Cui[1*], Xiaoqing Jiang[1*], Liang Zhao[1*], Yuanzhuo Chen[2], Zhangjun Cheng[3], Jiehua Li[2], Weifeng Tan[1,2], Alex F. Chen[2]

[1]Department of Biliary Surgery, Eastern Hepatobiliary Surgery Hospital, The Second Military Medical University, Shanghai 200438, China; [2]Department of Surgery, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; [3]Department of General Surgery, Zhongda Hospital, School of Medicine, Southeast University, Nanjing 210009, China

Contributions: (I) Conception and design: W Tan, AF Chen; (II) Administrative support: AF Chen; (III) Provision of study materials or patients: W Tan; (IV) Collection and assembly of data: J Li, L Cui, X Jiang; (V) Data analysis and interpretation: Z Cheng, Y Chen, L Zhao; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

*These authors contributed equally to this work.

Correspondence to: Weifeng Tan, MD, PhD. Department of Biliary Surgery, Eastern Hepatobiliary Surgery Hospital, The Second Military Medical University, Shanghai 200438, China. Email: twf1231@163.com; Alex F. Chen, MD, PhD. Department of Surgery, University of Pittsburgh School of Medicine, Pittsburgh, PA 15237, USA. Email: Chena5@upmc.edu.

Background: Currently, the diagnosis of extrahepatic cholangiocarcinoma (EHCC) is clinically difficult and remains unsatisfactory. In the present study we aimed to differentiate EHCC from benign biliary tract diseases in an effective and time-saving manner, by applying a novel metabolomic approach.

Methods: Bile was collected from patients with EHCC (n=32), primary bile duct stones (n=17) and choledochocysts (n=19), and processed for quantitative metabolomic analysis based on gas chromatography-mass spectrometry (GC-MS). The data obtained from the bile samples were analyzed by using principal component analysis (PCA), partial least squares discriminant analysis (PLS-DA), and orthogonal partial least square discriminant analysis. The content of cysteinylglycine in bile collected from different patients was verified using ultra-performance liquid chromatography/triple quadrupole tandem mass spectrometry. In patients with EHCC, the specificity and sensitity of well-known tumor markers cysteinylglycine, carbohydrate antigen 19-9 (CA19-9), and carcinoembryonic antigen (CEA)—alone or in combination—was calculated and compared.

Results: The content of cysteinylglycine in bile of patients with EHCC was significantly lower than that in bile of patients with primary bile duct stones and choledochal cysts (P<0.01). The sensitivity and specificity of cysteinylglycine content as a diagnostic value for EHCC was 91.7% and 84.4%, respectively. Moreover, the sensitivity and specificity rates for CA19-9 were 75% and 86.1%, CEA 12.5% and 100%, and the combination of cysteinylglycine and CA19-9, 84.4% and 88.9%, respectively.

Conclusions: Cysteinylglycine can be used as a novel potential biomarker to differentiate EHCC from primary bile duct stones and choledochal cysts.

Keywords: Extrahepatic cholangiocarcinoma (EHCC); bile metabolomics; gas chromatography-mass spectrometry (GC-MS); ultra-pressure liquid chromatography/triple quadrupole tandem mass spectrometry (UPLC-MS/MS)

Introduction

Cholangiocarcinoma (CCA), first described by Durand-Fardel in 1840 (1), is a malignant neoplasm arising from the biliary epithelium. The incidence and mortality rates from CCA have markedly increased worldwide over the past 30 years.

According to the international classification of American Joint Committee on cancer (AJCC), CCAs are currently classified into two groups: intrahepatic cholangiocarcinoma (IHCC) and extrahepatic cholangiocarcinoma (EHCC) (2). In Patients with a 2-year overall survival rate (23%) have a particularly poor prognosis, largely due to the advanced disease stage at the moment of diagnosis. Hence, the early detection of EHCC and resection of the tumor tissue would undoubtedly improve the patients' overall survival.

EHCC has been associated with long-standing cholangitis. Previous studies have shown that several factors such as primary sclerosing cholangitis, biliary tract inflammation, and hepatolithiasis are related with EHCC. CCA is usually clinically silent and notoriously difficult to diagnose before the tumor obstructs the bile ducts (3).

Current methods for the diagnosis of CCA employ multiple criteria including imaging, biliary cytology, and serum tumor markers. Additionally, ultrasonography, computed tomography (CT), magnetic resonance imaging (MRI), percutaneous transhepatic cholangiography (PTC), and other imaging technologies have also been used.

Contrast-enhanced MRI is reported to be effective in the differentiation of benign bile duct structures from malignant disorders (4). However, imaging of benign biliary tract diseases is often similar to the structures observed in patients with EHCC, thus affecting the diagnosis accuracy. Hence, tumor markers such as carbohydrate antigen 19-9 (CA19-9) and carcinoembryonic antigen (CEA) have been recently introduced. Although serum levels of CA19-9 were often used as a marker for the diagnosis of pancreatic cancer, CCA, and other malignancies, the clinical utility of this marker is rather limited since cholangitis patients also have elevated CA19-9 serum levels (5). Moreover, high levels of CEA, an acid glycoprotein produced by tumor tissue, are detected in serum of patients with colon cancer, pancreatic cancer, gastric carcinoma, or biliary tract carcinoma. However, its sensitivity is rather low, thereby restricting its use (6). Hence, a novel diagnostic approach is needed to differentiate EHCC from benign biliary tract diseases in an effective and time-saving manner.

Metabolomics is a novel methodology arising from the post-genomics era. This comprehensive approach allows the ideal measurement of all endogenous metabolites in a cell or body fluid. The use of metabolomics has markedly increased in pharmaceutical research and biomarker development for disease diagnosis, prognosis, and risk prediction (7). This technique combines data-rich analytical chemical methods such as nuclear magnetic resonance spectroscopy and mass spectrometry with chemometric for profiling metabolism and interpreting metabolic fingerprints in complex biological systems (8). Metabolomics has been applied in many areas, including drug safety assessment, characterization of genetically modified animal models of disease, diagnosis of human disease, understanding physiological variations, and drug therapy monitoring (9).

The principal component analysis (PCA) is the most commonly used "unsupervised" technique for the analysis of metabolomics data (10). Partial least squares discriminant analysis (PLS-DA), a "supervised" learning technique, is often used to maximize the covariance between the independent and dependent variables. Orthogonal partial least-squares discriminant analysis (OPLS-DA) is often used to maximize covariance between the measured data and the response variable (11).

In this study, we used a gas chromatography-mass spectrometry (GC-MS)-based bile metabolomic approach to screen potential biomarkers for differentiating EHCC from benign biliary tract diseases. The specificities and sensitivities of the screened biomarkers were then compared with those of traditional tumor biomarkers such as CA19-9 and CEA.

Methods

Patient samples

Reports from the information database of our Radiology Department from February 2013 to March 2013 were reviewed to identify patients with EHCC, stones, or cysts at Shanghai Eastern Hepatobiliary Surgery Hospital affiliated to the Second Military Medical University (Shanghai, China). Patients with hepatitis C virus (HCV), primary hepatocellular carcinoma, gout, hypothyroidism, or hematopathy due to hormonal use or presenting extensive burns were excluded from the analysis. Pregnant or breastfeeding women were also excluded. Finally, 32 patients with EHCC {22 men and 10 women; mean age: 58 [41–74] years}, 17 patients with stones {7 men and 10 women; mean age: 51.9 [39–78] years}, and 19 patients with cysts {4 men and 15 women; mean age: 45.5 [35–72] years} were confirmed by histopathologic

Table 1 Clinicopathological characteristics of the patients' cohorts

Index/patients	EHCC (n=32)	Stone (n=17)	Cyst (n=19)
Age (year)	58.0±8.0	51.9±14.0	45.5±12.7
Gender (male/female)	22/10	7/10	4/15
AFP (0–20 µg/L)	3.0±1.3	3.2±3.0	3.0±3.1
CEA (0–10 µg/L)	2.7 (0.7–8.9)	1.7 (0.4–3.6)	1.3 (0.2–4.5)
CA19-9 (0–39 U/mL)	78.2 (0.7–1,000)	13.5 (4.8–1,000)	8.4 (0.6–110.9)
STB (5.1–18.8 µmol/L)	101.55 (3–547.1)	11.9 (5.9–127.9)	11.9 (5.9–205.7)
CB (1.7–6.8 µmol/L)	77.3 (3.4–424.6)	4.8 (1.7–105)	4.4 (2.1–169.4)
BA (0–12 µmol/L)	40.4 (0.6–153.2)	4.8 (1.5–135.6)	5.2 (0.6–126.1)
STP (66–87 g/L)	65.5±5.8	71.6±7.1	69.7±6.0
ALB (34–48 g/L)	38.8±4.6	42.2±4.0	42.8±3.4
ALT (0–41 U/L)	91 [7–359]	23 [9–708]	13 [5–113]
AST (0–37 U/L)	61.5 [14–322]	24 [9–793]	17 [11–57]
GGT (5–61 U/L)	463 [10–2,110]	104 [20–1,155]	26 [11–473]
ALP (40–129 U/L)	298 [56–1,639]	85 [43–663]	57 [35–234]

EHCC, extrahepatic cholangiocarcinoma; AFP, alpha fetoprotein; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9; STB, serum total bilirubin; CB, conjugated bilirubin; BA, bile acid; STP, serum total protein; ALB, albumin; ALT, alanine aminotransferase; GGT, γ-glutamine transferase; ALP, alkaline phosphatase; EHCC, extrahepatic cholangiocarcinoma.

Table 2 Pathological diagnosis of TNM stage

Diagnostic basis	CCA (n=32)		Stone (n=17)	Cyst (n=19)
	Hilar CCA (n=23)	Distal CCA (n=9)		
Radical operation*	I: 2	Ia: 1; Ib: 1	17	19
	II: 10	IIa: 6; IIb: 1		
	IIIa: 5; IIIb: 5	–		
	IVb: 1	–		

Tumor staging was performed according to the 7th edition of the AJCC. *, EHCC was excision based on the postoperative pathological report. The radical operation for distal CCA included pylorus-preserving pancreaticoduodenectomy. TNM, tumor node metastasis; EHCC, extrahepatic cholangiocarcinoma; CCA, cholangiocarcinoma; AJCC, American Joint Committee on Cancer.

findings.

Demographic information and laboratory reports of the patients' cohorts are shown in *Table 1*. Pathological tumor node metastasis (TNM) stages of malignancy after surgery are shown in *Table 2*. No statistical significance was found in age or sex between patients with EHCC, nor patients with stones and patients with cysts.

This study was approved by the Ethics Committee of the Shanghai Eastern Hepatobiliary Surgery Hospital affiliated to the Second Military Medical University and consent

forms were authorized by patients or their family.

Bile sample collection

Bile was extracted from patients with EHCC, stones, and cysts using ultrasonic extraction (KQ-100DV; Shumei, Kunshan, China). Bile was then centrifuged at 3,000 ×g for 15 min (TGL-16C; Anting, Shanghai, China), mixed uniformly using a XW-80A Vortex mixer (Jingke, Shanghai, China), and kept in a 4 ℃ freezer for 40 min. The

supernatant (500 µL) was transferred into several 1.5 mL tubes and numbered. All samples were stored in a −80 ℃ refrigerator (Thermo Electron, MA, USA).

Bile sample preparation

Samples (100 µL) were transferred into 1.5 mL tubes after thawing at room temperature, and methanol and an internal standard solution including 0.1 mg/mL L-2-chlorobenzene alanine was added. The mixture was then centrifuged at 4 ℃, 13,000 ×g for 15 min, and the supernatant (400 µL) was transferred into a 2-mL glass centrifuge tube to dry in a vacuum drying chamber (DZ-1BC, Kexiao, Shanghai, China). Next, 20 mg/mL of methoxylamine pyridines (80 µL) were added into the centrifuge tube and shaken at room temperature for 30 s. The oximation reaction was first performed at 37 ℃ for 90 min to close carbonyl. The BSTFA derivatization reagent containing 1% TMCS (Sigma-Supelco, Bellefonte, Pennsylvania, USA) was added quickly and incubated at 70 ℃ for 60 min to increase the volatility of metabolites. After mixing and centrifugation at 3,000 ×g for 15 min, the supernatant was kept for GC-MS analysis.

GC-MS analysis

The GC-MS analysis was performed on an Agilent 7890A gas chromatograph (Agilent Corporation, Santa Clara, CA, USA) and LECO Chroma TOF Pegasus 4D mass spectrometer (LECO Corporation, St. Joseph, MI, USA). Separation was carried out by using a DB-5MS capillary column (250 cm × 4.6 mm *id*, 0.25 µm film thickness, J & W Scientific, Folsom, CA, USA). Helium (He) was used as the carrier gas. The flow rate was 1.0 mL/min and 1 µL of sample was injected. The column temperature started with a temperature of 50 ℃ for 1.0 min, increased to 330 ℃ at 10 ℃/min, and maintained for 5 min. The temperatures of the front inlet, transmission line, and ion source were 260, 280, and 220 ℃, respectively. The ionization voltage was −70 eV and solvent delay was 366 s. The scanning mode was full scan with m/z 85–600 and scanning frequency was 20 spectra per second.

Data processing and analysis

Primary GC-MS data were first processed using Chroma TOF4.3X software to identify the original peaks, filter noise, and original baseline; to conduct peak alignment and spectra deconvolution analyses; and to analyze the peaks qualitatively and quantitatively. A normalization method was used to dispose the exported data, and peak areas were imported into SIMCA-P v13.0 software (Umetrics AB, Umeå, Sweden) for multivariate statistical analysis. Variables were pretreated with pattern recognition technologies, and PCA was used to observe the separation trends of samples from EHCC, stone, and cyst patients. Sample models were constructed with PLS-DA. Subsequently, cross-validation was performed and the obtained R2 and Q2 values were used for evaluating the validity of sample models. R2(X) was the proportion of the total variance of the dependent variables explained by the model; R2(Y) was the proportion of the total variance of the response variable explained by the model, and; Q2(Y) was similar to R2(Y) except that it was computed by cross-validation. In order to identify differences between the groups exclusively, orthogonal signal correction (OSC) technologies were used and OPLS-DA models were constructed.

Identification of metabolites

To identify potential biomarkers for EHCC patients, the nuclear/cytoplasmic ratio, retention time, and mass spectrum of the metabolites were analyzed. First, we sought possible materials in the human metabolomics database (HMDB; www.hmdb.ca) according to the obtained nuclear/cytoplasmic ratio, and compared them with the standards. "Similarity" was used to match the obtained material and the standards. The total amount of points was 1,000 and compounds were reliably identified with a "similarity" corresponding to >700 points. Moreover, it is important to know how much an individual metabolite contributes to the principal components of the PLS-DA model. These quantities can be expressed as variable importance in the projection (VIP) values, and variables with VIP >1.0 are considered relevant to group discrimination. Herein, VIP statistics and similarity were applied to obtain the significant variables for subsequent metabolic pathway analysis.

Apart from the multivariate approaches, one univariate method (the Student's *t*-test) was selected to analyze the significance of each metabolite in EHCC patients as well as in patients with stones or cysts. Finally, differential metabolites were identified based on a similarity >700, VIP >1 and P<0.05. If both the retention time and mass spectrogram were the same, then there were no differences between the obtained material and the standards.

72

Cui et al. Cysteinylglycine, a novel biomarker for diagnosing EHCC

Ultra-pressure liquid chromatography/triple quadrupole tandem mass spectrometry (UPLC-MS/MS) analysis

To verify the different contents of the screened biomarker cysteinylglycine, bile samples were collected from 10 patients with EHCCs, 10 patients with stones, and 10 patients with cysts. The content of cysteinylglycine in each sample was determined by UPLC-MS/MS analysis. Bile samples (100 µL) were put into 1.5 mL centrifuge tubes, and a 0.90-mL solution including methanol and acetonitrile (5:2, v/v) was added. The mixture was centrifuged at 12,000 ×g for 10 min and 0.8 mL of supernatant was taken for UPLC-MS/MS analysis. For the cysteinylglycine standard solution, 1 g of cysteinylglycine was dissolved in double that amount of distilled water, diluted to 1.0 L, and stored at −20 ℃. Eight successive dilutions were prepared with methanol into the following concentrations: 0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.5, and 1.0 µg/mL. These dilutions were required for the determination of the calibration curve, and the concentration of cysteinylglycine in the test samples was calculated from the standard curve.

Determination of cysteinylglycine in bile samples was performed using UPLC (Waters Acquity System, Milford, MA, USA) coupled with an API4000 triple quadrupole mass spectrometry system (AB Sciex, Foster City, USA). The analytical column used was a Waters ACQUITY UPLC BEH HILIC (2.1 mm × 1 mm, 1.7 µm) with a pre-column (2.1 mm × 5 mm, 1.7 µm), maintained at 40 ℃. Mobile phases consisted of water/formic acid (99.9:0.1, v/v) (A) and acetonitrile/formic acid (99.9:0.1, v/v) (B), and performed at a flow rate of 0.4 mL/min. The linear gradient was as follows: 95% B at 0 min, 60% B at 4 min, 80% B at 8 min, 70% B at 9 min, and 95% B at 11 min, then a post-run of 2 min was performed for column equilibration. The final volume of the injection was 5 µL. Electrospray ionization in the positive (ESI+) mode was used with nitrogen (600 ℃, flow rate of 600 L/h) as the desolvation gas. The capillary voltage was set to 3 KV, and the ion source was kept at 350 ℃. The analysis was performed using a multiple reaction monitoring (MRM) model, and the precursor-to-product ion pair for cysteinylglycine was m/z 179.3–160.9 with a collision energy (CE) 11 V.

Statistical analysis

The data were calculated with SPSS version 18.0 software and shown as mean ± standard deviation (SD). The correlation studies were performed by linear regression analysis. Data comparison among groups was performed by one-way analysis of variance (ANOVA) using the Bonferroni correction as a post-hoc test, and molecules with a P value less than 0.05 were considered to be significant. Receiver operator characteristic (ROC) curves were constructed for CEA, CA 19-9, the selected biomarker cysteinylglycine, and the combination of CA 19-9 and cysteinylglycine.

Results

Metabolomics profiling of EHCC patients

Bile samples collected from 32 patients with EHCC, 17 patients with stones, and 19 patients with cysts were analyzed using GC-MS, and total ion chromatograms (TIC) were obtained. A representative TIC of a bile sample is shown in *Figure 1*. As shown in the figure, the number of compounds in the different patients' cohorts did not differ. Moreover, the reproducibility of spectral peaks was adequate and the instrument was stable.

PCA of GC-MS data

PCA was performed on the normalized GC-MS dataset using mean-centered data. The PCA 2D and 3D score plots of samples from patients with EHCC, stones, and cysts are shown in *Figure 2A,B*, respectively. In the score plots, the confidence interval was defined by Hotelling's T2 ellipse (95% confidence interval), and observations outside the confidence ellipse were considered outliers.

As shown in *Figure 2*, most observations were inside the confidence ellipse. The samples from patients with stones and cysts had a separation tendency to the top-left corner, while samples from EHCC patients presented a tendency to the lower-right corner. Altogether, these results indicate that the metabolic patterns of samples from EHCC patients were different from those with stones or cysts.

Variables in the blocks were mean-centered and scaled to unit variance (UV). The cumulative explanation rate of the model was obtained by automatic modeling analysis. The quality of all models was judged by the goodness-of-fit parameter (R2X) and the predictive ability parameter (Q2) which was calculated by a 7-fold cross-validation test (12). R2X and Q2 for patients with EHCC *vs.* patients with stones were 0.354 and 0.179, respectively. R2X and Q2 for patients with EHCC *vs.* patients with cysts were 0.322 and 0.165, respectively. Finally, the values were 0.331 and 0.172 for patients with stones *vs.* patients with cysts, respectively.
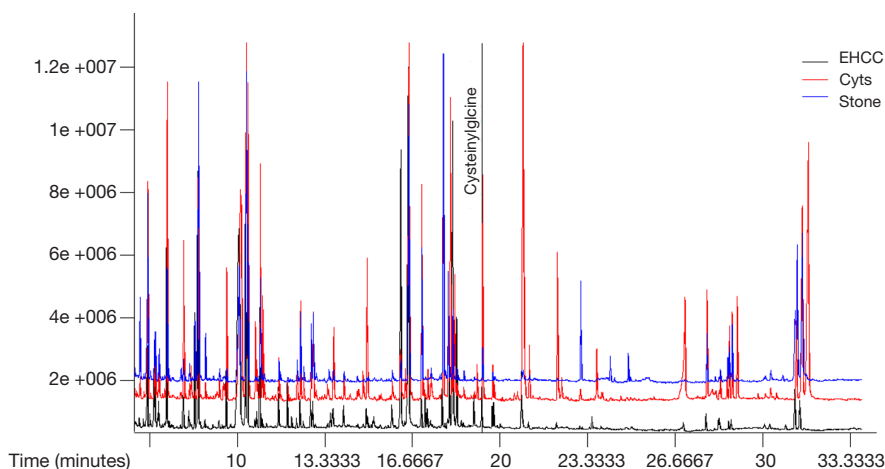
Figure 1 Bile total ion current profiles of one sample in the three different patients' cohorts (EHCC, stones, and cysts) using GC-MS. EHCC, extrahepatic cholangiocarcinoma; GC-MS, gas chromatography-mass spectrometry.

PLS-DA of GC-MS data

In order to distinguish bile metabolites of patients with EHCC from patients with stones and cysts, PLS was used on the X and the Y variables (PLS-DA uses data scale conversion processed by UV to perform modeling research on the first and second principal components). Next, the order of classified variable Y was rearranged randomly several times (n=200) and the corresponding Q2 values were also obtained, which helped to further test the validity of the model. R2X, R2Y, and Q2 for patients with EHCC *vs.* patients with stones were 0.291, 0.980, and 0.953, respectively. For patients with EHCC *vs.* patients with cysts, values were 0.314, 0.969, and 0.944, respectively. Finally, R2X, R2Y, and Q2 for patients with stones *vs.* patients with cysts were 0.323, 0.885, and 0.797, respectively.

The PLS-DA 2D score plot of samples from patients with EHCC and patients with cysts is shown in *Figure 2C*, and the result of the permutation test is shown in *Figure 2D*. The samples from patients with cysts are differentially separated from those of EHCC patients. Indeed, the cysts samples scattered on the left side while the samples from EHCC patients are grouped on the right side, thus indicating that the reference spectra for both patients' cohorts was different. The intercept of R2 and Q2 plots in the Y-vector permutation test was 0.452 and –0.176, respectively, which accurately reflects the authenticity of the model.

The PLS-DA 2D score plot of samples from patients with stones and samples from patients with EHCC as well as the permutation test result are shown in *Figure 2E,F*,

respectively. The samples from patients with stones separated well from those of patients with EHCC in the region constituted by the first and second principal components (t1 and t2), which suggested a different metabolic pattern in the two group of patients. The R2 of 0.46 and Q2 of –0.197 indicated that the model was fairly robust.

OPLS-DA of GC-MS data

To maximize the difference and dispersion degree among groups, OPLS-DA was applied. The correlation score plots of OPLS-DA of the metabolites identified differential metabolites as responsible for the differences found in the samples from EHCC patients and patients with stones or cysts. OPLS-DA used UV-scaled data to construct the model for the first and second principal components. OPLS-DA produced two key parameters: R2Y (which is the cumulative model variation in Y) and Q2.

The cumulative explanation rate of the OPLS-DA model showed that the R2 and Q2 were significantly high.

The OPLS-DA 2D score plot of samples from patients with cysts and patients with EHCC is shown in *Figure 2G*, and the load diagram in *Figure 2H*. The OPLS-DA 2D score plot for samples from patients with stones and patients with EHCC is shown in *Figure 2I* and the load diagram in *Figure 2J*. Samples from patients with stones and cysts scattered on the left side, while the samples from the EHCC patients grouped on the right side, thus indicating a significantly different metabolic pattern in
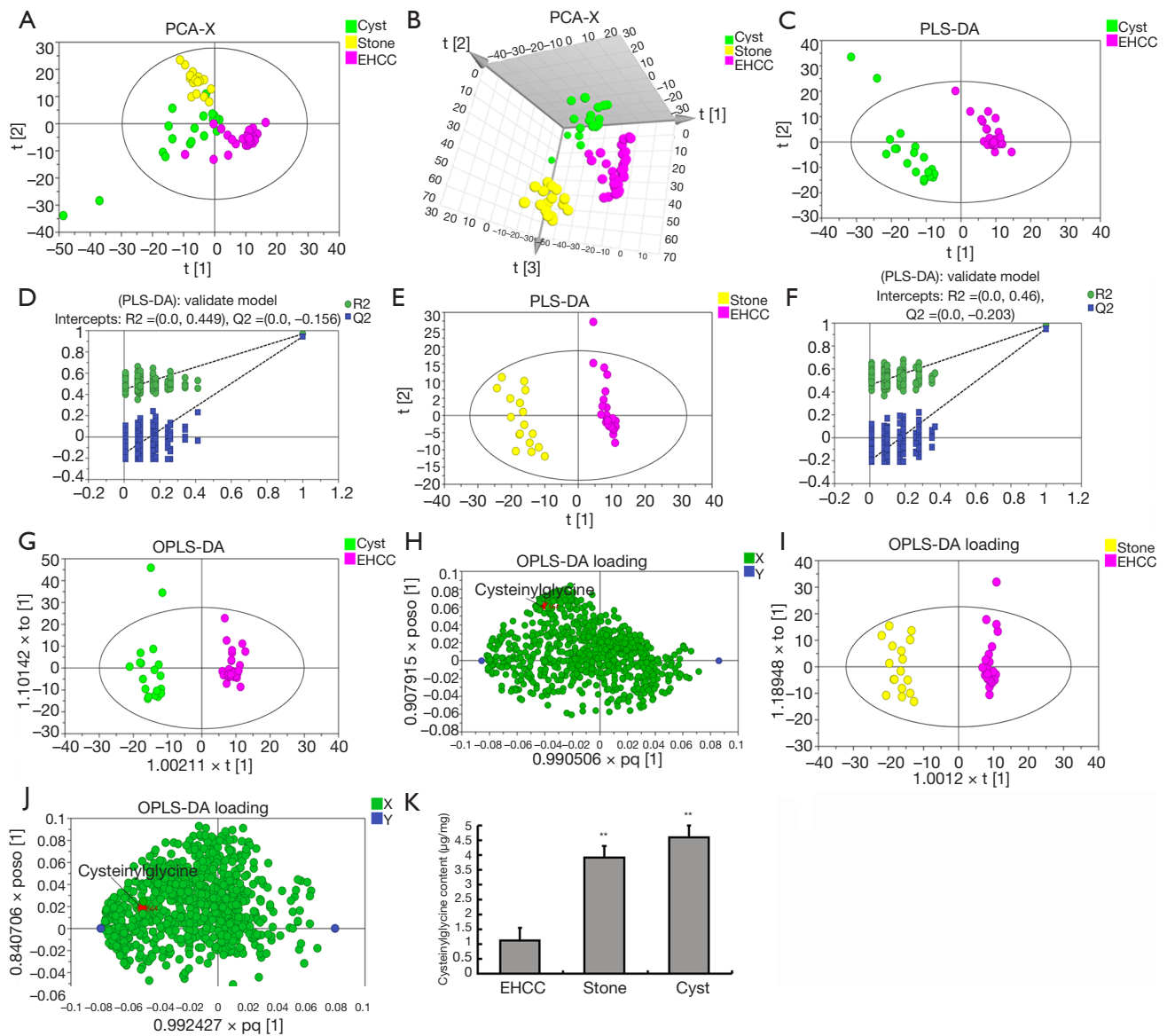
Figure 2 PCA score plot of total samples, score plot of PLS-DA model, OPLS-DA model and permutation test obtained from EHCC and benign diseases. (A) 2D PCA score plot of total samples; (B) 3D PCA score plot of total samples; (C) score plot of PLS-DA model obtained from patients with EHCC and cysts; (D) permutation test obtained from patients with EHCC and cysts; (E) score plot of PLS-DA model obtained from patients with EHCC and stones; (F) permutation test obtained from patients with EHCC and stones; (G) score plot of OPLS-DA model obtained from patients with EHCC and cysts; (H) load diagram obtained from patients with EHCC and cysts; (I) score plot of PLS-DA model obtained from patients with EHCC and stones; (J) load diagram obtained from patients with EHCC and stones; (K) the differences of cysteinylglycine between patient with EHCC and stones, EHCC and cysts both show statistical significance (**P<0.01). There is no statistical significance between patients with stones and cysts. PCA, principal component analysis; PLS-DA, partial least squares discriminant analysis; OPLS-DA, orthogonal partial least-squares discriminant analysis; EHCC, extrahepatic cholangiocarcinoma.
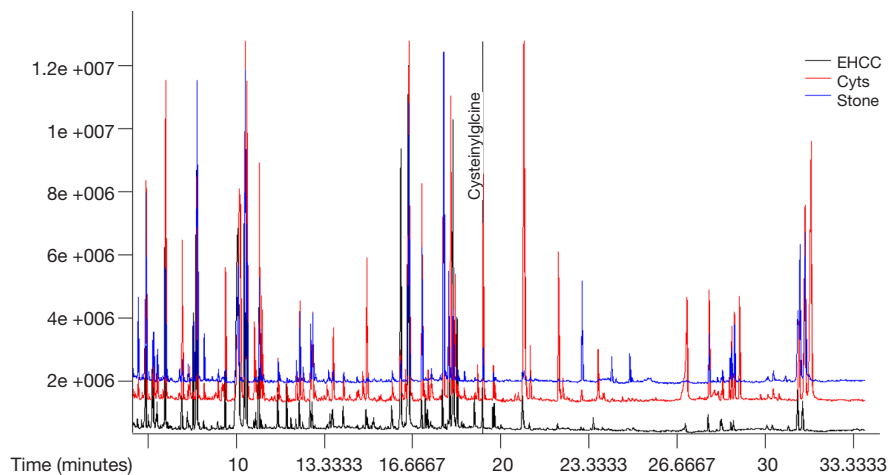
Figure 3 The ROC curve graphs of indexes in patients with EHCC and with benign biliary tract diseases. ROC, receiver operator characteristic; EHCC, extrahepatic cholangiocarcinoma.

Table 3 Comparison of diagnostic efficiency between traditional tumor markers and cysteinylglycine

Biomarker	CA19-9 (>100 U/mL)* (%)	CEA[#] (>10 µg/L) (%)	Cysteinylglycine (%)	CA19-9 + cysteinylglycine (%)
Sensitivity	75	12.5	91.7	84.4
Specificity	86.1	100	84.4	88.9

*, selecting value >100 U/mL derived from reference (13). [#], CEA selected diagnostic efficiency with specificity of 100%; the others selected the highest threshold. CA19-9, carbohydrate antigen 19-9; CEA, carcinoembryonic antigen.

patients with EHCC compared to the patients with stones or cysts. These results were consistent with those of PLS-DA discriminant.

Screening and identification of metabolites

Differential metabolites were identified based on similarity >700, VIP >1 and P<0.05. In the present study, we used ID as the data number imported into the software, peak was the name of every qualitative substance, RT was the retention time, MASS was the nucleocytoplasmic ratio of characteristic ions, MEAN was the mean peak area of every group after normalization, VIP was the weight of differences between the two groups caused by different substances, the P value was the result of a Student's t-test, and the fold change was the multiple relations between the two groups.

Cysteinylglycine was the screened metabolite in EHCC samples and verified by the standard. The content of cysteinylglycine in EHCC patients was significantly lower than in the patients with stones or cysts.

Analysis of the UPLC-MS/MS

The linear regression equation of cysteinylglycine obtained by UPLC-MS/MS was y =853x–723 and the correlation coefficient "r" was 0.9995. Based on this equation, we obtained the contents of cysteinylglycine in the samples of the three groups of patients: stones, cysts, and EHCC which were 3.906±2.105, 4.593±2.212, and 1.122±0.5233 µg/mL, respectively. The statistical analysis of the data of the stone and cyst samples (3.906 *vs.* 4.593) was not statistically different (P>0.05). However, the analysis of the comparisons stone and EHCC samples, and cyst and EHCC samples were both statistically significant (P<0.01) (*Figure 2K*).

Comparison of cysteinylglycine content with CA19-9 and CEA levels

The ROC curves of cysteinylglycine, CA19-9, CEA, and the combination of cysteinylglycine and CA19-9 for patients with EHCC and patients with benign biliary tract diseases were constructed with SPSS version 18.0 software.

The results are shown in *Figure 3*. Sensitivity and specificity were also calculated in *Table 3*.

Discussion

EHCC is a devastating malignancy and is notoriously difficult to diagnose and treat due to the similarity of its clinical manifestations to many benign biliary tract diseases (14). Here we used a bile metabonomics approach to differentiate EHCC from benign biliary tract diseases.

In the present study, bile samples were collected from 32 patients with EHCC, 17 patients with stones, and 19 patients with cysts, and a metabonomics analysis was performed. The cysteinylglycine content in the bile of patients with EHCC, stones, and cysts was then determined, and cysteinylglycine was screened out as a differential metabolite.

In order to evaluate the diagnostic use of cysteinylglycine in EHCC, the sensitivity and specificity of cysteinylglycine for EHCC was compared with traditional tumor markers such as CA19-9, CEA, and the combination of cysteinylglycine with CA19-9. Increased serum levels of CA19-9 are often used as a tumor marker for pancreatic cancer and CCA (5). However, the content of CA19-9 is also high in other disorders including bacterial cholangitis, extrahepatic bile cysts, and inflammation of the female reproductive tract, thus limiting its efficiency (15). Indeed, we demonstrate that the sensitivity of cysteinylglycine for EHCC is far higher than CA19-9, and present a similar specificity. However, the sensitivity of CA19-9 combined with cysteinylglycine for the diagnosis of EHCC was greater again, thus indicating that cysteinylglycine is more sensitive for the diagnosis of EHCC than CA19-9.

CEA has been shown to be a suitable diagnostic tool not only for CCA, but also for pancreatic and colon cancers (16). High levels of CEA in combination with CA19-9 have been often used for the diagnosis of CCA (17). However, the serum levels of CEA are also increased in benign disorders related to aging or smoking (18). Our results show that the specificity of CEA for the diagnosis of EHCC is 100%, but the sensitivity was extremely low, negatively influencing the diagnosis. Moreover, sensitivity is more important than specificity in the diagnosis of serious diseases such as cancer. Therefore, cysteinylglycine may become a novel potential biomarker for the differential diagnosis of EHCC.

Cysteinylglycine is an aminothiol produced by intracellular glutathione (γ-glutamyl-cysteinyl-glycine; GSH), a reaction catalyzed by γ-glutamyl transpeptidase (19). Cysteinylglycine provides reducing power to the cell and plays a key role in body metabolism (20). In many reactions, GSH is first oxidized to GSSG (oxidized glutathione) and then reduced to GSH by the NADPH-dependent glutathione reductase (19).

Several publications have shown that GSH plays an important role in human physiology, and cysteinylglycine is a key factor in the conversion of GSSG to GSH. Regarding its metabolic function, cysteinylglycine not only participates in the tricarboxylic acid cycle and glucose metabolism in the body, but also activates a large number of enzymes, thus promoting carbohydrate, fat, and protein metabolism (21,22). During redox reactions, cysteinylglycine reacts with hydrogen peroxide and free radicals, inhibiting the deleterious oxidative effects of sulfhydryl to cytomembranes and organs (23).

GSH is essential in regulation cell function including gene expression, DNA and protein synthesis, cell proliferation, apoptosis, and signal transduction (24,25). The reduction of GSH is an early apoptotic stimulus which promotes the generation of oxidative stress. Therefore, GHS plays a key role in the development of Parkinson's disease, hepatosteatosis, cancer, and heart diseases (26). Therefore, we speculated that the decrease of cysteinylglycine content in the bile samples of patients with EHCC inhibits the synthesis of GSH, thus altering carbohydrate, fat and protein metabolism and producing redox imbalance.

Conclusions

Based on our results, we conclude that cysteinylglycine may be considered as a biomarker to differentiate EHCC from benign biliary tract diseases, and this finding warrants further investigation.

The majority of patients with EHCC are currently diagnosed using a combination of cytology, imaging, and tumor markers in serum. This process is complicated and the results are not reliable enough. In this study, we used a metabolomics approach to analyze and compare the bile from EHCC patients with patients with benign biliary tract diseases. Altogether the screening showed that cysteinylglycine is a differential metabolite for the diagnosis of EHCC, and can be used a reliable, convenient, and minimally invasive diagnostic tool.

Acknowledgments

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/tcr.2018.01.04). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the Ethics Committee of the Shanghai Eastern Hepatobiliary Surgery Hospital affiliated to the Second Military Medical University and consent forms were authorized by patients or their family.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Olnes MJ, Erlich R. A review and update on cholangiocarcinoma. Oncology 2004;66:167-79.
2. Sempoux C, Jibara G, Ward SC, et al. Intrahepatic cholangiocarcinoma: new insights in pathology. Semin Liver Dis 2011;31:49-60.
3. Khan SA, Thomas HC, Davidson BR, et al. Cholangiocarcinoma. Lancet 2005;366:1303-14.
4. Chung YE, Kim MJ, Park YN, et al. Varying appearances of cholangiocarcinoma: radiologic-pathologic correlation. Radiographics 2009;29:683-700.
5. Singh S, Tang SJ, Sreenarasimhaiah J, et al. The clinical utility and limitations of serum carbohydrate antigen (CA19-9) as a diagnostic tool for pancreatic cancer and cholangiocarcinoma. Dig Dis Sci 2011;56:2491-6.
6. Ikeguchi M, Ohro S, Maeda Y, et al. Detection of cancer cells in the peripheral blood of gastric cancer patients. Int J Mol Med 2003;11:217-21.
7. Xia J, Broadhurst DI, Wilson M, et al. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. Metabolomics 2013;9:280-99.
8. Lindon JC, Holmes E, Nicholson JK. Metabonomics: systems biology in pharmaceutical research and development. Curr Opin Mol Ther 2004;6:265-72.
9. Yang J, Xu G, Zheng Y, et al. Diagnosis of liver cancer using HPLC-based metabonomics avoiding false-positive result from hepatitis and hepatocirrhosis diseases. J Chromatogr B Analyt Technol Biomed Life Sci 2004;813:59-65.
10. Wagner S, Scholz K, Donegan M, et al. Metabonomics and biomarker discovery: LC-MS metabolic profiling and constant neutral loss scanning combined with multivariate data analysis for mercapturic acid analysis. Anal Chem 2006;78:1296-305.
11. Mishur RJ, Rea SL. Applications of mass spectrometry to metabolomics and metabonomics: detection of biomarkers of aging and of age-related diseases. Mass Spectrom Rev 2012;31:70-95.
12. Walsh MC, Brennan L, Pujos-Guillot E, et al. Influence of acute phytochemical intake on human urinary metabolomic profiles. Am J Clin Nutr 2007;86:1687-93.
13. Patel AH, Harnois DM, Klee GG, et al. The utility of CA 19-9 in the diagnoses of cholangiocarcinoma in patients without primary sclerosing cholangitis. Am J Gastroenterol 2000;95:204-7.
14. Park MS, Kim TK, Kim KW, et al. Differentiation of extrahepatic bile duct cholangiocarcinoma from benign stricture: findings at MRCP versus ERCP. Radiology 2004;233:234-40.
15. Li C, Wang W, Ding H, et al. Value of contrast-enhanced sonography in the diagnosis of peripheral intrahepatic cholangiocarcinoma. J Clin Ultrasound 2011;39:447-53.
16. Kim YT, Byun JS, Kim J, et al. Factors predicting concurrent cholangiocarcinomas associated with hepatolithiasis. Hepatogastroenterology 2003;50:8-12.
17. Chen J, Yi S, He D, et al. Study of relationship between cholangiocarcinoma with CEA,CA19-9 and CA242 in bile and blood. Chin Med Herald 2008;5:96-7.
18. Zhang D, Yu M, Xu T, et al. Predictive value of serum CEA, CA19-9 and CA125 in diagnosis of colorectal liver metastasis in Chinese population. Hepatogastroenterology

78

Cui et al. Cysteinylglycine, a novel biomarker for diagnosing EHCC

2013;60:1297-301.

19. Wu G, Fang YZ, Yang S, et al. Glutathione metabolism and its implications for health. J Nutr 2004;134:489-92.

20. Katrusiak AE, Paterson PG, Kamencic H, et al. Pre-column derivatization high-performance liquid chromatographic method for determination of cysteine, cysteinyl-glycine, homocysteine and glutathione in plasma and cell extracts. J Chromatogr B Biomed Sci Appl 2001;758:207-12.

21. Gänzle MG, Vermeulen N, Vogel RF. Carbohydrate, peptide and lipid metabolism of lactic acid bacteria in sourdough. Food Microbiol 2007;24:128-38.

22. Ohmori S, Kawase T, Higashiura M, et al. High-performance liquid chromatographic method to analyze picomole levels of glutathione, cysteine and

cysteinylglycine and its application to pre-cancerous rat livers. J Chromatogr B Biomed Sci Appl 2001;762:25-32.

23. Shan XQ, Aw TY, Jones DP. Glutathione-dependent protection against oxidative injury. Pharmacol Ther 1990;47:61-71.

24. Townsend DM, Tew KD, Tapiero H. The importance of glutathione in human disease. Biomed Pharmacother 2003;57:145-55.

25. Jones DP. Redox potential of GSH/GSSG couple: assay and biological significance. Methods Enzymol 2002;348:93-112.

26. Mülle T, Muhlack S. Cysteinyl-glycine reduction as marker for levodopa-induced oxidative stress in Parkinson's disease patients. Mov Disord 2011;26:543-6.