Tumor Biology

Original Article

A comparative proteomic analysis of bile for biomarkers of cholangiocarcinoma

Marut Laohaviroj[1,2], Jeremy Potriquet[3], Xinying Jia[4],
Sutas Suttiprapa[1,2], Yaovalux Chamgramol[1], Chawalit Pairojkul[1],
Paiboon Sithithaworn[5], Jason Mulvenna[3] and Banchob Sripa[1,2]

Abstract

Cholangiocarcinoma is a primary malignant tumor of the bile duct epithelium. Cholangiocarcinoma is usually detected at an advanced stage when successful treatment is no longer possible. As the tumor originates from the bile duct epithelium, bile is an ideal source of tumor biomarkers for cholangiocarcinoma. In this study, we used a quantitative proteomics approach to identify potential tumor-associated proteins in the bile fluid of six cholangiocarcinoma patients. Three different gross-appearance tumor types were used in the analysis: mass-forming type ($n = 2$), periductal infiltrating type ($n = 2$), and intraductal growth type ($n = 2$). Two bile samples from non-cancerous patients were used as controls. Isobaric labeling, coupled with Tandem mass spectrometry, was used to quantify protein levels in the bile of cholangiocarcinoma and control patients. In all, 63 proteins were significantly increased in cholangiocarcinoma bile compared to normal bile. Alpha-1-antitrypsin was one of the overexpressed proteins that increased in cholangiocarcinoma bile samples. Immunohistochemical analysis revealed that alpha-1-antitrypsin was detected in 177 (50%) of 354 cholangiocarcinoma tissues from our Tissue Bank. Immunoblotting of 54 cholangiocarcinoma bile samples showed that alpha-1-antitrypsin was positive in 38 (70%) samples. Fecal enzyme-linked immunosorbent assay showed that alpha-1-antitrypsin level was able to distinguish cholangiocarcinoma patients from normal individuals. In conclusion, alpha-1-antitrypsin is a potential marker for early diagnosis of cholangiocarcinoma.

Introduction

Cholangiocarcinoma (CCA) is a primary malignant tumor of the biliary tract epithelium. The highest incidence of CCA in the world is in northeast Thailand where CCA is strongly associated with infection by the carcinogenic liver fluke, *Opisthorchis viverrini*. CCA is a major public health problem in this area and nearby countries of the Greater Mekong Subregion. While CCA is asymptomatic in the early stage, it grows rapidly and metastasizes aggressively in the late stage. CCA is usually detected in the advanced stage when treatment options are limited. Surgical resection is the most common treatment for CCA, but 5-year survival rate is less than 5%,[1] and this survival rate has not improved over the last 30 years.[2,3] Since early detection significantly improves the survival of CCA

[1]Department of Pathology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand
[2]WHO Collaborating Centre for Research and Control of Opisthorchiasis (Southeast Asian Liver Fluke Disease), Tropical Disease Research Center (TDRC), Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand
[3]Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia
[4]Centre for Advanced Imaging, The University of Queensland St Lucia, Brisbane, QLD, Australia
[5]Department of Parasitology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand

Corresponding author:
Banchob Sripa, Department of Pathology, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand.
Email: banchob@kku.ac.th

cases,[4] there is an urgent need for an effective diagnostic tool to detect early-stage CCA in high-risk groups. Currently, there are no non-invasive and effective tumor markers specific for *O. viverrini*–associated or non–*O. viverrini*–associated CCA. Although both CA19-9 and carcinoembryonic antigen (CEA) are elevated in the serum of CCA patients, these markers are also increased in diseases such as alcoholic liver disease, clinical viral hepatitis, primary sclerosing cholangitis (PSC), cholestasis, liver injury, and other cancers.[1] Thus, diagnostic values of these markers for CCA is limited.[5] The discovery of novel and specific markers for CCA is essential for improving the survival rate of CCA patients.

Given that CCA originates from the biliary epithelium, tumor cells and/or cancer-derived proteins can be released into bile fluid and be detected in bile fluid and/or feces. Proteomic analyses of bile, including biomarker discovery studies, have been performed for a number of different diseases including pancreas and hepatobiliary diseases.[6–10] In these studies, candidate proteins were verified in bile and blood of patients, but none were verified in fecal samples. In this study, we aimed to find accessible diagnostic markers for CCA by identifying potential protein markers in the bile of CCA patients and evaluating their suitability as a diagnostic marker in feces. In particular, comparative proteomic analysis was employed to identify elevated proteins in the bile fluid of CCA patients, candidate proteins were then verified in cancer tissue, considered the tissue of origin, by immunohistochemistry (IHC), and immunoblot was used to confirm their presence in the bile fluid. Finally, diagnostic potential of candidate protein was assessed by fecal enzyme-linked immunosorbent assay (ELISA) technique.

Materials and methods

Ethics statement

All human specimens and the protocols in these studies were approved by the Human Ethics Committee of Khon Kaen University, based on the ethics of human specimen experimentation of the National Research Council of Thailand (HE581073).

Patients and specimens for proteomic experiments

Human bile was obtained with informed consent from eight patients (six CCA patients with different gross types—mass forming (MF, $n=2$), periductal infiltrating (PI, $n=2$), and intraductal growth (IG, $n=2$)—and two non-CCA patients) in *O. viverrini* endemic area who underwent hepatectomy and had bile collected from their gallbladder at Srinagarind Hospital, Khon Kaen University, Khon Kaen, Thailand. Patients' profiles and clinical information were obtained from medical records. Histopathological classification of each tumor was performed according to the World Health

Organization (WHO), Pathology and Genetics of Tumors of the Digestive System Guidelines.[11] Tumor stage was determined using the International Hepato-Pancreato-Biliary Association (IHPBA) classification. Six CCA tissue samples had the same histological papillary type.

Protein purification and processing for mass spectrometry

A bile sample of 200 μL was taken from each patient and mixed with 1 mL of lysis buffer (7 M urea, 2 M thiourea, and 4% w/v 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS)). Cleanascite™ (Biotech Support Group, USA), lipid removal reagent, was added to the bile and the sample was vertically shaken for 1 h at 4°C before centrifugation at $10,000 \times g$ for 1 min. The supernatant was then placed in a 3-kDa Amicon Ultra-15 Centrifugal Filter Unit (Merck Millipore, USA) and washed three times with 3 mL of 0.1 M triethylammonium bicarbonate buffer (TEAB; Sigma-Aldrich, Australia). Proteins in 250 μL of buffer were then precipitated using 1 mL acetone with incubation at −20°C overnight. The sample was then centrifuged at $10,000 \times g$ for 10 min and the supernatant removed. The pellet was washed three times with acetone, air-dried at room temperature, and dissolved in 200 μL of 0.5 M TEAB. After measuring protein concentration with the Bio-Rad Protein Assay (Bio-Rad Laboratories Pty., Ltd., USA), each sample was divided into 200 μg aliquots of bile protein. Each sample was mixed with sodium dodecyl sulfate (SDS) and dithiothreitol (DTT) to concentrations of 10% SDS and 10 mM DTT. The mixture was then incubated at 65°C for 1 h. Freshly prepared iodoacetic acid (IAA) was then added to a concentration of 50 mM and the solution incubated at 37°C for 45 min in the dark. Proteins were then acetone-precipitated as described above and then solubilized in 100 μL 0.5 M TEAB. Proteins were digested by the addition of 2 μg of trypsin with overnight incubation at 37°C. The trypsin digestion was terminated by the addition of formic acid to a final concentration of 1%.

Isobaric labeling

Tryptic peptides from 100 μg of total protein were labeled with 8-plex iTRAQ reagents according to the manufacturer's protocol (Applied Biosystems, USA). Labeled peptides from all eight samples were combined and salt and unbound iTRAQ reagents were removed using in sequence: (1) a HiTrap SP HP column (GE Healthcare, Germany) and (2) a SepPak C18 cartridge (Waters Corporation, USA).

Peptide OFFGEL fractionation

Purified peptides were fractionated according to their isoelectric point using a 3100 OFFGEL Fractionator (Agilent Technologies, Germany) on a 24-well immobilized pH

gradient (IPG) strip. Peptide samples were diluted to a final volume of 3.6 mL using OFFGEL peptide sample solution, and 150 μL was added to wells that had been rehydrated for 15 min with the peptide IPG strip rehydration solution. Electrophoresis was performed at room temperature with 50 μA until the 50 kV h level was reached. After electrophoresis, each of the 24 fractions was removed and the wells rinsed with a further 200 μL of a solution of water:methanol:formic acid (49:50:1) for 15 min. Rinse solutions from each of the 24 wells were combined with their corresponding fractions and were lyophilized and stored at −80°C prior to analysis.

Tandem mass spectrometry

Lyophilized OFFGEL fractions were solubilized in 20 μL of 0.1% trifluoroacetic acid (TFA). Labeled peptides were analyzed by liquid chromatography–tandem mass spectrometry (LC-MS/MS) on a Shimadzu Prominence Nano HPLC (Japan) coupled to a Triple TOF 5600 Mass Spectrometer (AB SCIEX, Canada) equipped with a nano electrospray ion source; 2 μL of each OFFGEL fraction was injected onto a 50 mm × 300 μm C18 trap column (Agilent Technologies) at 20 μL/min. The samples were de-salted on the trap column for 5 min using 0.1% formic acid (aq) at 20 μL/min. The trap column was then placed in-line with an analytical nano-HPLC column (150 mm × 75 μm; C18, 5 μm; Vydac, USA) for mass spectrometry analysis. A linear gradient of 1%–80% solvent B (90/10 acetonitrile/0.1% formic acid (aq)) over 120 min at 800 nL/min flow rate, followed by a steeper gradient from 40% to 80% solvent B in 5 min, was used for peptide elution. The ion spray voltage was set to 2000 V, declustering potential 100 V, curtain gas flow at 25, nebulizer gas 1 (GS1) at 10, and interface heater at 150°C; 500 ms full-scan time-of-flight mass spectrometry (TOF-MS) data were acquired followed by 20 × 50 ms full-scan product ion data in an Information-Dependant Acquisition (IDA) mode. Full-scan TOF-MS data were acquired over the mass range 350–1800 and for product ions 100–1800. Ions observed in the TOF-MS scan exceeding a threshold of 100 counts and a charge state of +2 to +5 were set to trigger the acquisition of product ion spectra for a maximum of 20 of the most intense ions. The data were acquired and processed using Analyst TF 1.5.1 software (AB SCIEX).

Peptide and protein identification and database searches

Searches were performed using ProteinPilot v4 (AB SCIEX) with the following parameters: allowing for methionine oxidation as a variable modification; carbamidomethylation (or modification with methyl methanethiosulfonate (MMTS) where appropriate) as a fixed modification; two missed cleavages; charge states +2, +3, and +4; and trypsin as the

enzyme. Searches were conducted against the UniProt *Homo sapiens* proteome dataset (proteome ID UP000005640) as of Oct 2014 (23,878 entries). Proteins were grouped using ProteinPilot's ProGroup algorithm, single-peptide identifications were not considered, and only proteins containing at least one unique significant peptide identification were reported. Searches were also conducted with X! TANDEM Jackhammer TPP (GPM, 2013.06.15.1)[12] using the same database appended with reversed sequences with the following parameters: enzyme = trypsin; precursor ion mass tolerance = ±0.1 Da; fragment ion tolerance = ±0.1 Da; fixed modifications = carbamidomethylation (or modification with MMTS where appropriate) and iTRAQ modification of Lys and N-term free-amines (using modification masses 304.205360 and 304.199040); variable modifications = methionine oxidation and variable labeling of Tyr residues with iTRAQ reagents; number of missed cleavages allowed = 2; allowed charge states = +2 to +4; and "k-score" as the scoring algorithm. The Trans Proteomic Pipeline (TPP)[12] was used to validate peptide and protein identifications using PeptideProphet[11] and ProteinProphet,[13] and Mayu[14] was used for false discovery rate (FDR) estimation. Using the TPP and the same parameters, with the exception of Lys and N-terminal modification by iTRAQ reagent specified as a variable modification, further searches were conducted to estimate the iTRAQ labeling efficiency. iTRAQ reporter ion intensities from peptides identified by ProteinPilot as suitable for quantitation and which possessed a probability greater than 0.95 were used in the R package iQuantitator[15] to generate credible intervals for protein expression differences in specific samples.

IHC of CCA tissue microarrays

A CCA tissue microarray (TMA, $n = 354$ and control, $n = 2$) was constructed by the Department of Pathology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. IHC reactions were performed on 4-μm-thick paraffin sections of CCA TMA on silane-coated slides (Sigma Chemical, USA) using the immunoperoxidase method.[16] The paraffin sections were de-waxed in xylene and rehydrated in serial dilution of ethanol. Antigen retrieval was performed in citrate buffer pH 6.0 under high pressure. Endogenous peroxidase activity was blocked in methanol with 3% hydrogen peroxide (HO). Non-specific binding was blocked by incubation in 5% normal horse serum, followed by overnight incubation at 4°C with 1:300 diluted rabbit anti-alpha 1 antitrypsin antibody (polyclonal, ab9373; Abcam, USA). Phosphate-buffered saline (PBS) was used instead of primary antibody in the control sections. Sections were washed with PBS and incubated with the specific secondary antibodies for 30 min at room temperature (1:300 diluted anti-rabbit horseradish peroxidase (HRP) conjugated, ab97051; Abcam). Slides were subsequently counterstained with Mayer's™ hematoxylin,
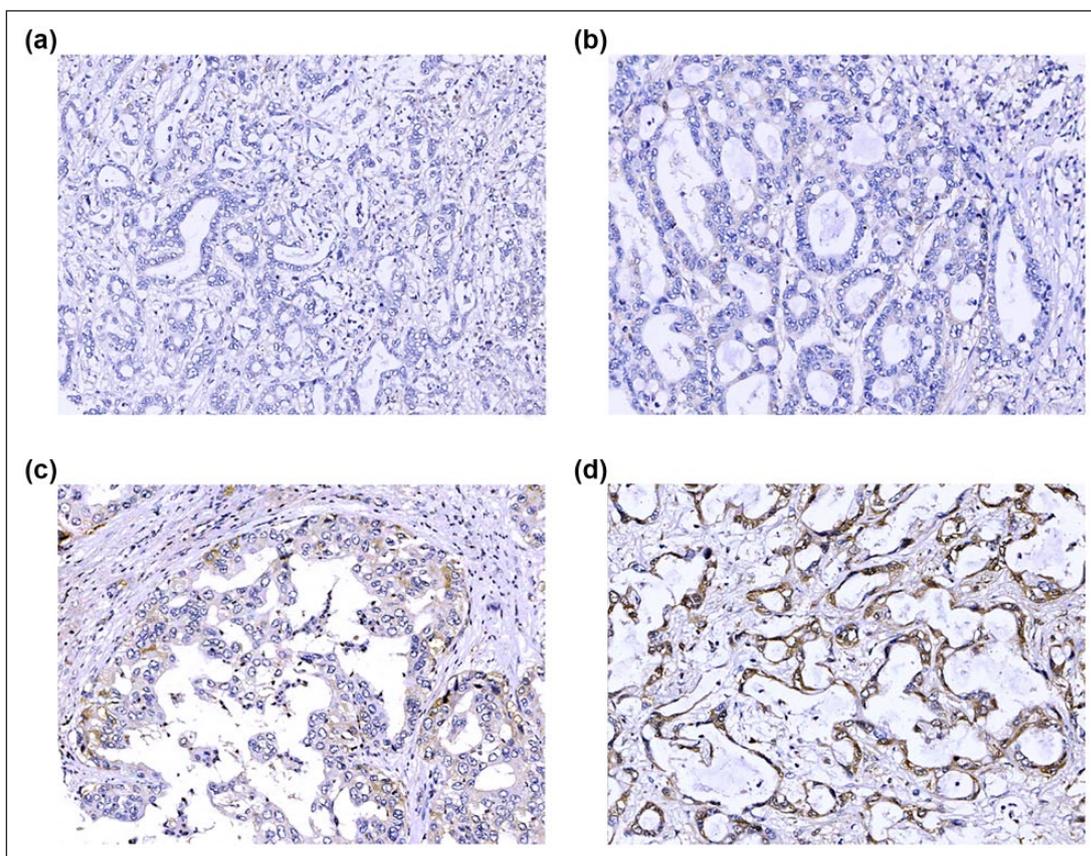
Figure 1. Representative cell grading of immunohistochemical staining of AAT: (a) 0 or absent; (b) 1 or absent; (c) 2 or present; and (d) 3 or present (original magnification of 100×).

dehydrated, cleaned, and mounted. All tissue slides were histopathologically evaluated by M.L., B.S., and the senior gastrointestinal pathologist, C.P.

Scoring and statistical analysis

All sections were evaluated by two investigators blinded to grading of human sections. The staining density was quantified as the percentage of cells stained positively in tissue as follows: 0, <10%; 1, positive staining in 11%–33%; 2, 34%–66%; and 3, 67%–100%. Intensity scores were multiplied by the density score to yield an overall score of 0–9 for each sample. As described previously,[16] to decrease human bias, scores of 0 or 1 were defined as negative/absent or low expression. A score of 2–3 was defined as positive/present or high expression (Figure 1). The SPSS V.19.0 (IBM) statistical package and Prism 6.0 (GraphPad) software were used to analyze data and create graphical images. For cross-sectional analyses, the chi-squared test and chi-squared test for trend were utilized to analyze the relationship between a candidate protein expression and categorical parameters according to clinicopathological parameters (e.g. age, gross type, and tumor invasion). Values of $p < 0.05$ were considered statistically significant.

Western blotting

Human gallbladder bile was obtained from 54 CCA patients in *O. viverrini* endemic areas who had undergone hepatectomy. Bile samples were collected from Srinagarind Hospital, Khon Kaen University, Khon Kaen, Thailand. Patient profiles, clinical information, and tumor classification were also obtained. Five CCA tissues and 5 non-CCA adjacent normal tissues were included as positive and negative controls. Western blot analysis was performed as previously described.[17] Briefly, 5 μL of bile sample was separated by 15% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) with 20 mA for 2 h. Separated proteins were then transferred to nitrocellulose membrane. Non-specific binding was blocked by incubating membrane in 5% skimmed milk in PBS with 0.05% Tween 20 (PBST) at 25°C for 1 h. The membrane was then incubated for 1 h with 1:1000 diluted rabbit anti-alpha 1 antitrypsin antibody (polyclonal, ab9373; Abcam) in 2% skimmed milk in PBST at 25°C. After washing with PBST, the membrane was incubated in diluted 1:1000 anti-rabbit HRP-conjugated ab97051 (Abcam) in 2% skimmed milk in PBST at 25°C for 1 h. The membrane was then washed with PBST and PBS, and reactive band was visualized with 3,3′-diaminobenzidine tetrahydrochloride solutions (DAB; Sigma Chemical, USA).
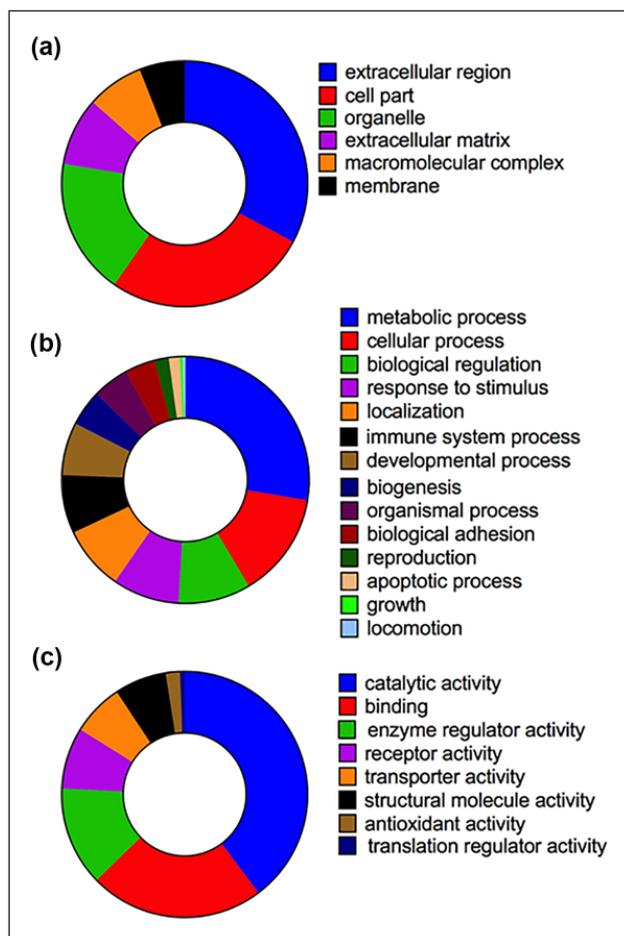
Figure 2. 147 significantly dysregulated proteins were classified by PANTHER classification system version 10.0. Classification based on (a) cellular component, (b) biological process, and (c) molecular function.

ELISA detection of alpha-1 antichymotrypsin in fecal samples

Human feces were obtained with informed consent from 12 CCA patients in Srinagarind Hospital, Khon Kaen University, Khon Kaen, Thailand. They were admitted to receive hepatectomy, and fecal samples were collected before surgical operation. These 12 CCA cases were identified as moderate-to-advanced-stage CCA. In all, 11 suspected CCA cases diagnosed by ultrasonography with subclinical symptoms were included and identified as early-stage CCA. Other stages of *O. viverrini* infection including 30 *O. viverrini* positive cases with hepatobiliary abnormality (HBA+) cases, 21 *O. viverrini* positive cases without HBA (HBA−) cases, and 10 uninfected individuals were also recruited from *O. viverrini* endemic area in Khon Kaen, Thailand. Hospital case profiles and clinical information were obtained from Srinagarind Hospital, while the profiles of patients (e.g. sex, age, fecal examination, and hepatobiliary ultrasonography results) were

obtained from the Tropical Diseases Research Center (TDRC), Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. The patients with obesity, diabetes mellitus, chronic viral hepatitis, inflammatory bowel diseases, and other cancers are excluded from this study. All feces were frozen at −20°C immediately after collection. An amount of 50 mg of each fecal sample was mixed with 2.5 mL wash buffer. The sandwich ELISA for detection of alpha-1 antichymotrypsin (AAT) was performed according to the manufacturer's procedure of Alpha-1-Antitrypsin in Stool Human ELISA Kit (Biovender, USA). The SPSS V.19.0 statistical package and Prism 6.0 were used to analyze data and create graphical images. For cross-sectional analyses, one-way analysis of variance (ANOVA) and receiver operating characteristic (ROC) curves were utilized to analyze the candidate protein expression level among the group. Values of $p < 0.05$ were considered statistically significant.

Results

Protein identification using tandem mass spectrometry

In total, 1243 proteins were identified. Among them, 147 proteins were differentially expressed in CCA and non-CCA bile samples (Supplementary Table 1). These proteins could be classified into six categories for cellular components using the PANTHER classification system (Figure 2). A majority of proteins identified (59.70%) were classified as extracellular (32.80%) or cytoplasmic (26.90%). Other identified proteins were classified as originating in organelles (17.90%), the extracellular matrix (9%), the macromolecular complex (7.50%), and membranes (6%). In biological aspects, proteins involved in metabolic process were the most abundant, while the second and third most abundant biological process categories were cellular processes and biological regulation, respectively. Proteins with catalytic, binding, and enzyme regulatory activities were the most abundant in molecular function categories.

Selection of potential protein marker

Alpha-1-antitrypsin was one of the proteins that was highly increased in bile of CCA cases in this study, and it has never been tested in bile and fecal samples from CCA patients. Moreover, AAT has been suggested by other groups to be involved in carcinogenesis of CCA, and serum-oxidized AAT has been suggested as a potential biomarker to screen and predict *O. viverrini*-related CCA by ELISA.[18] Together, with the availability of commercial ELISA kit for detection of AAT in fecal samples, AAT was chosen for further evaluation for diagnostic potential.

Table 1. Clinicopathological parameters and AAT expression in CCA tissue.

Clinicopathological parameters	AAT		
	Absent	Present	*p*
Gross type			*0.01*
Mass forming	123	95	
Periductal infiltrating	25	39	
Intraductal growth	14	5	
Tumor size (cm)			NS
≤5	48	47	
>5	89	81	
Vascular invasion			*0.05*
Absent	54	52	
Present	51	82	
Lymphatic invasion			NS
Absent	38	45	
Present	84	112	
Perineural invasion			NS
Absent	66	73	
Present	36	61	
Metastasis			NS
Absent	60	77	
Present	65	76	
Stage			*0.003*
1	7	9	
2	9	28	
3	37	25	

AAT: alpha-1 antichymotrypsin; CCA: cholangiocarcinoma.

Immunohistochemical staining of AAT in CCA tissues

IHC analyses were carried out on the CCA TMAs (*n*=354) with normal case sections to determine whether AAT was expressed in CCA cells. AAT expression was found in 177 out of 354 patients (50%). AAT staining was preferably located in the cytoplasm (Figure 1). Increased expression of AAT did not show a preference for any histological subtypes of tumor. However, AAT expression levels in CCA cells were associated with gross type (*p*=0.01), vascular invasion (*p*=0.05), and stage (*p*=0.003; Table 1).
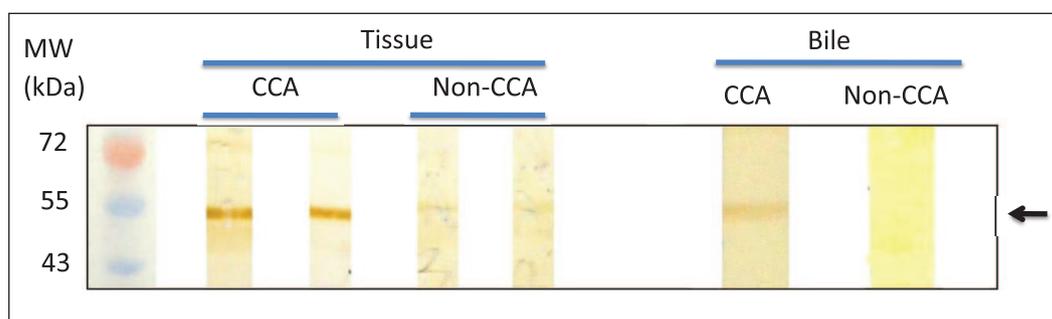
Immunoblot verification of AAT in bile samples

A total of 54 bile samples were collected from hepatectomy cases that were confirmed as CCA positive by postoperative histopathology. The target band size of AAT was 52 kDa (Figure 3), and a positive band at the expected size was seen in all CCA tissues, while there were no corresponding bands in normal control tissues. AAT was 70% positive (38/54) in CCA bile samples.

### Sandwich ELISA for AAT in the feces of CCA patients and patients with hepatobiliary abnormalities

The level of AAT in feces was determined using sandwich ELISA in three groups: CCA patients, patients with hepatobiliary abnormalities (HBA 3+), and the control group. The CCA group displayed significantly higher levels of AAT than both the HBA3+ and control groups (Figure 4(a); Table 2). ROC curve analysis demonstrated that the AAT level could be used to differentiate between CCA and controls with an area under the curve (AUC) of 0.83, at cut-off <173.4 µg/g, and sensitivity and specificity of 80% and 75%, respectively (Figure 4). Similarly, AUC was 0.73 with cut-off <97.75 µg/g (sensitivity=70% and specificity=90%) when compared between suspected CCA cases and control.

Discussion

Bile is an important potential source for identifying early diagnostic markers for CCA as CCA originates and develops in the bile duct epithelium, and hence, cancer cells and/or cancer-derived proteins can be released into the bile ducts. Several groups have employed different proteomic approaches to identify tumor-specific makers for CCA, and in these studies, protein candidates were verified in bile and blood by immunoblot and/or ELISA.[6–10] In this study, we employed a quantitative proteomic approach to investigate the differential protein expression in bile samples from different gross types of CCA and non-CCA patients. A protein candidate was then verified in CCA tissues, the

Figure 3. Representation of immunoblot of AAT in CCA tissues and bile samples (MW: molecular weight of protein markers (kDa)). Expected size of AAT is 52 kDa (arrow).
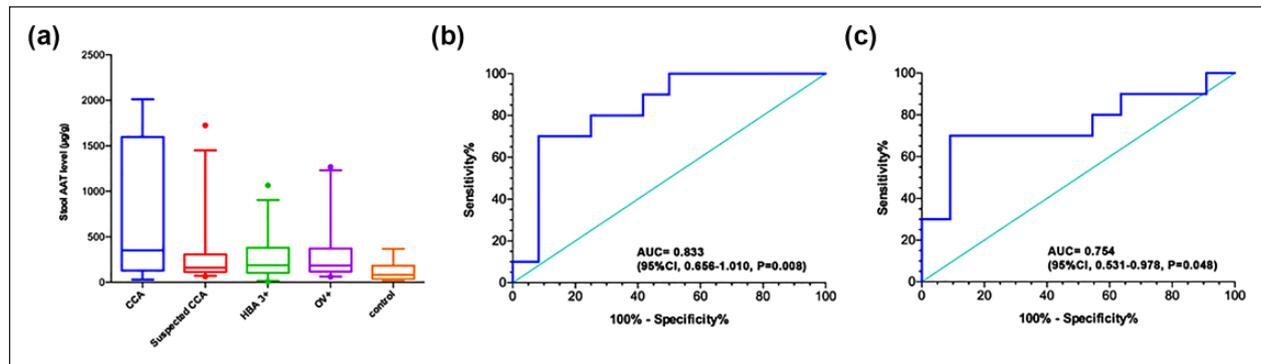
Figure 4. The fecal AAT level in five groups of samples, and the performance of AAT as a detection marker. (a) Level of AAT in feces (μg/g) in each group. (b) The ROC curve of control versus CCA. (c) The ROC curve of control versus suspected CCA.

Table 2. Performance of diagnosis of fecal AAT by sandwich ELISA.

Diagnostic comparison	Cut-off (μg/g)	Sensitivity (95% CI)	Specificity (95% CI)	Likelihood ratio	AUC (95% CI)	p^a
Control versus CCA	<173.40	80 (44.39–97.48)	75 (42.81–94.51)	3.20	0.833±0.090 (0.656–1.010)	0.008
Control versus suspected CCA	<96.75	70 (34.75–93.33)	90.91 (58.72–99.77)	7.70	0.754±0.114 (0.531–0.978)	0.048

AAT: alpha-1 antichymotrypsin; CCA: cholangiocarcinoma; ELISA: enzyme-linked immunosorbent assay; CI: confidence interval; AUC: area under the curve.

$^a p$ value was significantly higher than 0.5 cut-off.

upstream protein sources of the tumor protein in bile. The candidate marker was then verified in bile fluid to confirm its presence in the mid-stream flow of tumor-derived proteins. Finally, the candidate CCA marker protein was verified in stool samples, representing the down-stream destination of the tumor marker and the most accessible matrix for assaying at-risk individuals for CCA.

Given the complexity of bile fluid, sample preparation steps prior to mass spectral analysis and the technology available for performing mass spectrometry have tremendous effects on resulting protein profile at both quality and quantity. The first large-scale proteomic study of bile was performed in 2004,[19] revealing 87 proteins in the bile of CCA patients. Later, using more advanced technology and sample preparation techniques, up to 2552 bile proteins have been discovered.[20] The comparative study of protein profiles in bile from non-malignant and malignant patients has also been performed with different proteomic approaches including conventional one-dimensional (1D) and two-dimensional (2D) SDS-PAGE protein separation followed by mass spectrometry and labeled or non-labeled quantitative proteomics.[8–10] Both single-protein biomarker and peptide panels have been validated for diagnostic potential for CCA diagnosis.[6–10]

In this study, using iTRAQ labeling, peptide fractionation by OFFGEL electrophoresis and LC-MS/MS identified 1243 unique proteins. In all, 63 proteins were overexpressed in CCA bile samples compared to non-CCA bile. A similar workflow was performed by Farina et al.,[21] in which 416 proteins were identified with 66 proteins upregulated in the bile from cancer patients and 30% of the

proteins identified as overexpressed were also identified in our study. These include neutrophil gelatinase–associated lipocalin (NGAL) and galectin-3-binding protein (LG3BP), which have been proposed as markers to differentiate malignant from non-malignant stenosis. However, CEA-related cell adhesion molecule 6 (CEAM6), which could discriminate between malignant and non-malignant causes of biliary stenosis, was not increased in our study.

Different protein profiles in different studies might be due to the different gross/histological types of cancer or the heterogeneity of cancer tissue itself. However, there were no differences in the detection of AAT in bile by western blot analysis among different gross/histological types in this study (Supplementary Tables 2 and 3). Nonetheless, this observation needs to be confirmed by higher number of CCA cases. Since CCA in the northeast Thailand is associated with chronic *O. viverrini* infection, the difference might also related to *O. viverrini*–associated and non-associated carcinogenesis, as there are significant difference in the frequency of mutated genes between *O. viverrini*–associated and non- associated CCA.[22]

AAT and oxidized AAT (ox-AAT) were detected in the tumor of CCA patients and overexpression is associated with poor-prognosis. Serum levels of ox-AAT were also significantly higher in groups of patients with heavy *O. viverrini* infection, advanced periductal fibrosis (APF), and CCA when compared with healthy controls.[18] In this study, we found that AAT expression was found in 50% of the CCA tissues examined. Moreover, AAT expression in CCA cells was associated with gross type, vascular invasion, and cancer stage, which was not found in the

previous study.[18] This could be because we examined more (354) CCA tissues as compared to 42 tissues in the previous study.

Detection of a potential cancer biomarker in stool samples has been attempted in gastrointestinal and pancreatobiliary cancers.[23,24] We also evaluated the detection of AAT in bile and feces, and AAT was detected in the majority of the CCA bile samples examined (70%). The result from fecal ELISA indicated that AAT could be an early diagnostic marker for CCA and could differentiate between normal and early-stage CCA.

The higher rate of detection of AAT in bile fluid and fecal samples when compared with detection in the tissue by the IHC was observed. Since cancers are heterogeneous, there could be a subset of cells within the tumor that express AAT. One of the limitations of TMAs is the sampled cancer tissue which may or may not include areas containing cells expressing the potential marker. Thus, results from TMAs may not accurately represent expression of the marker over the whole tumor. However, TMAs do provide a robust analysis capacity for staining and analyzing a large number of samples, reducing the inherent variability in staining procedures. By attempting to detect secreted AAT in the bile fluid and feces, which represents the overall secretion of AAT from the tumor tissue, we sought to increase our ability to successfully detect AAT when compared with the subset of CCA cells that are analyzed in TMAs. Another possibility is that tissue fixation and processing could reduce antigen integrity and eventually effect the staining of AAT in the CCA tissue during TMAs. In contrast, bile and fecal samples are fresh or frozen and are not subjected to fixation that could result in increased sensitivity to the antigen in western blots and ELISA experiments.

As stated above, tumors contain a heterogeneous mixture of cells and each individual tumor will differ from others. Accordingly, each CCA case will result in different measured levels of expressed or secreted AAT. In this case, 70% of the cases examined displayed overexpression of AAT and we would therefore anticipate using a panel of markers, which included AAT, to provide better sensitivity than any single marker.

There were limitations in our ability to acquire the stool samples from CCA patients as hepatobiliary surgery disturbs gastrointestinal function. Therefore, we could not collect stool samples in most cases during hospitalized until discharged. Moreover, identification of suspected early cases of CCA was practically performed by ultrasonography of the liver in high-risk groups who infected with the liver fluke. The rate of the suspected CCA is <0.2% in our series of over 6000 (Sripa, unpublished data). Therefore, these all limit our CCA or suspected CCA case recruitment. Future studies, with a higher number of CCA cases, will further verify the diagnostic potential of AAT.

In summary, we have employed a quantitative proteomic approach to identify tumor biomarker for CCA. AAT was found to be one of the proteins that was increased in all gross types of CCA bile. For the first time, we showed that fecal ELISA detection of AAT could be an alternative method for early detection of CCA. Further investigation of other overexpressed proteins found in this study could lead us to a protein or protein panel for developing early diagnostic marker for CCA, which will increase the treatment options and survival rate of the patients.

References

1. Carriaga MT and Henson DE. Liver, gallbladder, extrahepatic bile ducts, and pancreas. *Cancer* 1995; 75(1 Suppl): 171–190.
2. Uttaravichien T, Bhudhisawasdi V, Pairojkul C, et al. Intrahepatic cholangiocarcinoma in Thailand. *J Hepatobiliary Pancreat Surg* 1999; 6(2): 128–135.
3. Shaib Y and El-Serag HB. The epidemiology of cholangiocarcinoma. *Semin Liver Dis* 2004; 24(2): 115–125.
4. Gores GJ. Early detection and treatment of cholangiocarcinoma. *Liver Transpl* 2000; 6(6 Suppl 2): S30–S34.
5. Razumilava N and Gores GJ. Cholangiocarcinoma. *Lancet* 2014; 383(9935): 2168–2179.
6. Lankisch TO, Metzger J, Negm AA, et al. Bile proteomic profiles differentiate cholangiocarcinoma from primary

sclerosing cholangitis and choledocholithiasis. *Hepatology* 2011; 53(3): 875–884.

7. Lukic N, Visentin R, Delhaye M, et al. An integrated approach for comparative proteomic analysis of human bile reveals overexpressed cancer-associated proteins in malignant biliary stenosis. *Biochim Biophys Acta* 2014; 1844(5): 1026–1033.

8. Shen J, Wang W, Wu J, et al. Comparative proteomic profiling of human bile reveals SSP411 as a novel biomarker of cholangiocarcinoma. *PLoS ONE* 2012; 7(10): e47476.

9. Sriwanitchrak P, Viyanant V, Chaijaroenkul W, et al. Proteomics analysis and evaluation of biomarkers for detection of cholangiocarcinoma. *Asian Pac J Cancer Prev* 2011; 12(6): 1503–1510.

10. Zabron AA, Horneffer-van der Sluis VM, Wadsworth CA, et al. Elevated levels of neutrophil gelatinase-associated lipocalin in bile from patients with malignant pancreatobiliary disease. *Am J Gastroenterol* 2011; 106(9): 1711–1717.

11. Keller A, Nesvizhskii AI, Kolker E, et al. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* 2002; 74(20): 5383–5392.

12. Deutsch EW, Mendoza L, Shteynberg D, et al. A guided tour of the trans-proteomic pipeline. *Proteomics* 2010; 10(6): 1150–1159.

13. Nesvizhskii AI, Keller A, Kolker E, et al. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 2003; 75(17): 4646–4658.

14. Reiter L, Claassen M, Schrimpf SP, et al. Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Mol Cell Proteomics* 2009; 8(11): 2405–2417.

15. Schwacke JH, Hill EG, Krug EL, et al. iQuantitator: a tool for protein expression inference using iTRAQ. *BMC Bioinformatics* 2009; 10: 342.

16. Yonglitthipagon P, Pairojkul C, Chamgramol Y, et al. Up-regulation of annexin A2 in cholangiocarcinoma caused by Opisthorchis viverrini and its implication as a prognostic marker. *Int J Parasitol* 2010; 40(10): 1203–1212.

17. Suttiprapa S, Loukas A, Laha T, et al. Characterization of the antioxidant enzyme, thioredoxin peroxidase, from the carcinogenic human liver fluke, Opisthorchis viverrini. *Mol Biochem Parasitol* 2008; 160(2): 116–122.

18. Jamnongkan W, Techasen A, Thanan R, et al. Oxidized alpha-1 antitrypsin as a predictive risk marker of Opisthorchiasis-associated cholangiocarcinoma. *Tumour Biol* 2013; 34(2): 695–704.

19. Kristiansen TZ, Bunkenborg J, Gronborg M, et al. A proteomic analysis of human bile. *Mol Cell Proteomics* 2004; 3(7): 715–728.

20. Barbhuiya MA, Sahasrabuddhe NA, Pinto SM, et al. Comprehensive proteomic analysis of human bile. *Proteomics* 2011; 11(23): 4443–4453.

21. Farina A, Dumonceau JM, Antinori P, et al. Bile carcinoembryonic cell adhesion molecule 6 (CEAM6) as a biomarker of malignant biliary stenoses. *Biochim Biophys Acta* 2014; 1844(5): 1018–1025.

22. Chan-On W, Nairismagi ML, Ong CK, et al. Exome sequencing identifies distinct mutational patterns in liver fluke-related and non-infection-related bile duct cancers. *Nat Genet* 2013; 45(12): 1474–1478.

23. Yang JY, Sun YW, Liu DJ, et al. MicroRNAs in stool samples as potential screening biomarkers for pancreatic ductal adenocarcinoma cancer. *Am J Cancer Res* 2014; 4(6): 663–673.

24. Bailey JR, Aggarwal A and Imperiale TF. Colorectal cancer screening: stool DNA and other noninvasive modalities. *Gut Liver* 2016; 10(2): 204–211.