

# Bile Proteomic Profiles Differentiate Cholangiocarcinoma from Primary Sclerosing Cholangitis and Choledocholithiasis

Tim O. Lankisch,<sup>1,2\*</sup> Jochen Metzger,<sup>3\*</sup> Ahmed A. Negm,<sup>1</sup> Katja Voßkuhl,<sup>1</sup> Eric Schiffer,<sup>3</sup> Justyna Siwy,<sup>3</sup> Tobias J. Weismüller,<sup>1,2</sup> Andrea S. Schneider,<sup>1</sup> Kathrin Thedieck,<sup>4</sup> Ralf Baumeister,<sup>4,5</sup> Petra Züribig,<sup>3</sup> Eva M. Weissinger,<sup>6</sup> Michael P. Manns,<sup>1</sup> Harald Mischak,<sup>3,7</sup> and Jochen Wedemeyer<sup>1,2</sup>

Early detection of malignant biliary tract diseases, especially cholangiocarcinoma (CC) in patients with primary sclerosing cholangitis (PSC), is very difficult and often comes too late to give the patient a therapeutic benefit. We hypothesize that bile proteomic analysis distinguishes CC from nonmalignant lesions. We used capillary electrophoresis mass spectrometry (CE-MS) to identify disease-specific peptide patterns in patients with choledocholithiasis (n = 16), PSC (n = 18), and CC (n = 16) in a training set. A model for differentiation of choledocholithiasis from PSC and CC (PSC/CC model) and another model distinguishing CC from PSC (CC model) were subsequently validated in independent cohorts (choledocholithiasis [n = 14], PSC [n = 18] and CC [n = 25]). Peptides were characterized by sequencing. Application of the PSC/CC model in the independent test cohort resulted in correct exclusion of 12/14 bile samples from patients with choledocholithiasis and identification of 40/43 patients with PSC or CC (86% specificity, 93% sensitivity). The corresponding receiver operating characteristic (ROC) analysis revealed an area under the curve (AUC) of 0.93 (95% confidence interval [CI]: 0.82-0.98,  $P = 0.0001$ ). The CC model succeeded in an accurate detection of 14/18 bile samples from patients with PSC and 21/25 samples with CC (78% specificity, 84% sensitivity) in the independent cohort, resulting in an AUC value of 0.87 (95% CI: 0.73-0.95,  $P = 0.0001$ ) in ROC analysis. Eight out of 10 samples of patients with CC complicating PSC were identified. **Conclusion:** Bile proteomic analysis discriminates benign conditions from CC accurately. This method may become a diagnostic tool in future as it offers a new possibility to diagnose malignant bile duct disease and thus enables efficient therapy particularly in patients with PSC. (HEPATOLOGY 2011;53:875-884)

Cholangiocarcinoma (CC) is the second most common hepatobiliary cancer and arises from cholangiocytes of the intra- and extrahepatic biliary tract.<sup>1</sup> Unfortunately, CC is often detected in an unresectable stage and hence it is associated with a poor prognosis and a life expectancy of 6-12 months.<sup>2</sup> The diagnosis of CC is based on a combination of imaging techniques and tissue sampling. Tumor

Abbreviations: AUC, area under the curve; CA 19-9, carbohydrate antigen 19-9; CC, cholangiocarcinoma; CEA, carcinoembryonic antigen; CE-MS, capillary electrophoresis mass spectrometry; ERCP, endoscopic retrograde cholangiopancreatography; LC-MS, liquid chromatography-mass spectrometry; PSC, primary sclerosing cholangitis; PTC, percutaneous transhepatic cholangiographies; ROC, receiver operating characteristic; SDS-PAGE, sodium dodecylsulfate polyacrylamide gel electrophoresis; SVM, support vector machine.

From the <sup>1</sup>Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany; <sup>2</sup>Integrated Research and Treatment Center, Transplantation (IFB-Tx), Hannover Medical School Hannover, Germany; <sup>3</sup>mosaiques diagnostics, Hannover, Germany; <sup>4</sup>Bioinformatics and Molecular Genetics (Faculty of Biology) and Centre for Biological Signalling Studies BIOS, Albert-Ludwigs-Universität Freiburg, Germany; <sup>5</sup>ZBMZ (Faculty of Medicine) and Freiburg Institute for Advanced Studies (FRIAS), Albert-Ludwigs-Universität Freiburg, Germany; <sup>6</sup>Department of Hematology, Hemostasis and Oncology, Hannover Medical School, Hannover, Germany; and <sup>7</sup>BHF Glasgow Cardiovascular Research Centre, University of Glasgow, Faculty of Medicine, Glasgow, UK.

Received September 2, 2010; accepted November 24, 2010.

E.S. and H.M. were supported by grant 202230 from the GENINCA EU-project. This work was supported by the Niedersächsische Krebsgesellschaft and a grant from the German Federal Ministry of Education and Research through the Integrated Research and Treatment Center, Transplantation (reference number: 01EO0802). This work was also supported by the Schlieben-Lange-Programm, the EC 6th FP NoE LifeSpan (LSHG-CT-2007-036894), the DFG Excellence Initiative (EXC 294, FRIAS LifeNet), and FORSYS (#03139219). The contents of this article are the sole responsibility of the authors.

\*These authors contributed equally to this work.

Address reprint requests to: Tim O. Lankisch, M.D., Department of Gastroenterology, Hepatology and Endocrinology, Medical School Hannover, Carl-Neuberg Str. 1, 30625 Hannover, Germany. E-mail: lankisch.tim@mh-hannover.de; fax: +49-511-532 3351.

markers, like serum carbohydrate antigen 19-9 (CA 19-9) and carcinoembryonic antigen (CEA), have a low specificity and sensitivity (<80%), which limit their potential to differentiate between benign and malignant bile duct stenosis.<sup>1,3</sup> Magnetic resonance cholangiopancreatography and endoscopic retrograde cholangiopancreatography (ERCP) provide imaging tools leading to a diagnostic accuracy of only 70%-80%.<sup>4,5</sup> Despite being more invasive, ERCP allows tissue sampling by brush cytology or biopsy of the suspected biliary stenosis. Unfortunately, ERCP-guided biopsies and brush cytology can provide a diagnosis in only 36%-46% of cases, although expert groups achieved a remarkably higher yield in diagnosing dysplasia and CC in patients with primary sclerosing cholangitis (PSC).<sup>6,7</sup>

The main predisposing factor for the development of CC in Western countries is PSC.<sup>3,8</sup> PSC is associated with a >160-fold increased risk to develop hepatobiliary malignancies.<sup>3,9</sup> In patients with PSC, the differentiation between benign and malignant strictures is particularly difficult, because CC as well as chronic or acute inflammation frequently result in similar cholangiographic findings.<sup>10</sup>

An interesting approach for the early detection of CC in PSC is the identification of markers in bile, as the development of carcinoma takes place at the biliary epithelium and tumor-related proteins become detectable in bile rather than serum. The measurement of established serum biomarkers in bile has been performed, but is only of limited clinical value.<sup>11,12</sup>

A novel concept to detect CC is the analysis of protein patterns instead of focusing on single proteins. In a small number of preliminary descriptive studies proteomic analyses with only a few samples were performed to evaluate the role of proteomics in bile without developing disease-specific models.<sup>13-18</sup> Proteomic analysis may be a valuable tool in the discovery and differentiation of various diseases, but has not been clinically evaluated in hepatobiliary diseases.<sup>19,20</sup> We hypothesize that proteomic analysis of bile can be applied to distinguish CC from nonmalignant biliary tract pathologies.

## Patients and Methods

**Patients.** Bile samples were collected at the gastrointestinal endoscopy unit of the Hannover Medical

School, Germany. We performed 142 endoscopic procedures and bile aspiration was successful in 75% of cases. From 94 consecutive patients included in the study, bile samples were successfully collected during 107 interventions (102 ERCs, five percutaneous transhepatic cholangiographies [PTC]). Indications for cholangiographic interventions were: PSC, CC, and choledocholithiasis. Ten patients developed CC in addition to PSC. Clinical cholangitis was present pre-intervention in one patient with choledocholithiasis, in three patients with PSC, and in two patients with CC. We defined cholangitis as the presence of fever, elevated C-reactive protein (CRP), and alkaline phosphatase (AP). Antibiotic treatment before intervention was initiated in 30 out of 107 endoscopic procedures (seven choledocholithiasis, eight PSC, 15 CC). We performed microbiological bile analysis in 93/107 examinations and bile remained sterile in only 15% of cases (1/30 choledocholithiasis, 7/36 PSC, 8/41 CC). Coexistent bacterial infection (bacteriobilia with fever, elevated CRP, AP, and gamma-GT) was present in six patients. Detailed patients characteristics and laboratory data are given in Table 1.

The diagnosis of PSC was based on typical cholangiographic findings such as strictures or irregularity of intrahepatic or extrahepatic bile ducts after exclusion of secondary causes for sclerosing cholangitis. CC was proven histologically in 35 out of 38 patients. In three patients a definite histology could not be obtained, but clinical, laboratory, radiological, and ERCP findings were consistent with a diagnosis of CC. None of the patients with CC received chemotherapy before the cholangiographic intervention. The diagnosis of choledocholithiasis was based on ultrasound and/or endoscopic ultrasound and confirmed by ERC. The trial was approved by the local ethical committee of Hannover Medical School and written informed consent was obtained from all patients.

**Collection of Bile.** Bile collection was performed as previously described.<sup>21</sup> In brief, bile was aspirated by placing a 5F standard ERC catheter (without previous flushing) into the bile duct before contrast dye injection. Approximately 0.5 to 6 mL of bile (mean 2 mL) were collected and transferred into a sterile tube. In five patients (four patients with CC and one patient with choledocholithiasis) bile was aspirated during PTC

**Table 1. Patient Demographic and Laboratory Characteristics**

	Patients/ Samples	Age (Range)	Female/ Male	AP	$\gamma$ -GT	Bilirubin	Leukocytes	CRP	ALT	AST	CA 19-9	LDH
Choledocholithiasis	22/30	62 (25-84)	10/12	211 (46-924)	305 (11-2626)	43 (3-300)	7218 (2300-18100)	30 (1-187)	76 (12-411)	59 (14-176)	77 (6-435)	221 (156-320)
PSC	34/36	48 (22-68)	8/26	279 (48-1475)	215 (18-1160)	37 (6-225)	6628 (3000-11600)	12 (0-103)	85 (14-392)	79 (20-378)	37 (1-227)	196 (119-358)
CC	38/41	59 (25-82)	9/29	445 (96-2144)	480 (38-1468)	161 (5-506)	8413 (3400-20800)	41 (2-184)	104 (14-716)	117 (23-479)	5254 (1-81243)	233 (124-470)

Abbreviations: AP, alkaline phosphatase;  $\gamma$ -GT, gamma glutamyltransferase; AST, aspartate aminotransferase; ALT, alanine aminotransferase; CRP, C-reactive protein; carbohydrate antigen, CA 19-9; LDH, Lactate dehydrogenase.

directly after sonographic-guided percutaneous puncture of the bile ducts and before contrast injection. Bile samples were directly frozen at  $-80^{\circ}\text{C}$  and were thawed only once just before proteomic analysis. Bile samples were diluted in  $\text{H}_2\text{O}$  to a final protein concentration of 1 mg/mL, as verified with the bicinchoninic acid assay (Interchim, Montlucon, France).

**Bile Sample Preparation.** For CE-MS analysis, 0.7 mL diluted bile was added to 0.7 mL *n*-butanol/*iso*-proyl ether 4:6 (v/v) and centrifuged for 10 minutes at 14,000 rpm and  $4^{\circ}\text{C}$ . The lower aqueous phase was extracted and diluted with 0.5 mL of 8 M urea, followed by 1 mL  $\text{H}_2\text{O}$ , and passed over a 10 kDa MWCO Centrisart ultrafilter (Sartorius, Goettingen, Germany) at 3,000 rpm until 1.4 mL filtrate was obtained. The filtrate was desalted on a PD-10 column (GE Healthcare, München, Germany) preequilibrated in 0.01% aqueous  $\text{NH}_4\text{OH}$  (Roth, Karlsruhe, Germany). After elution with ammonium buffer, the sample was lyophilized, stored at  $4^{\circ}\text{C}$ , and resuspended in CE-MS running buffer containing 20% acetonitrile and 1% formic acid before analysis.

**Capillary Electrophoresis Mass Spectrometry Analysis.** CE-MS analysis was performed as described using a P/ACE MDQ capillary electrophoresis system (Beckman Coulter, Fullerton, CA) on-line coupled to a Micro-TOF MS (Bruker Daltonic, Bremen, Germany).<sup>19,22</sup> The ESI sprayer (Agilent Technologies, Palo Alto, CA) was grounded, and the ion spray interface potential was set between  $-4.0$  and  $-4.5$  kV. Data acquisition and MS acquisition methods were automatically controlled by the CE via contact-close-relays. Spectra were accumulated every 3 seconds over a range of  $m/z$  350 to 3,000. Details regarding accuracy, precision, selectivity, sensitivity, reproducibility, and stability of the CE-MS method have been described.<sup>19</sup>

**Data Processing.** Mass spectral ion peaks representing identical molecules at different charge states were deconvoluted into single masses using MosaiquesVisu software.<sup>23</sup> Only signals were included with a charge  $>1$  observed in a minimum of three consecutive spectra and

with signal-to-noise ratios  $>4$ .<sup>24</sup> The software employs probabilistic clustering and uses isotopic distribution and conjugated masses for charge-state determination of peptides/proteins. The resulting peak list characterizes each peptide by its molecular mass, CE-migration time, and ion signal intensity (amplitude). Because these parameters are influenced by the amount of salt and peptides in the sample, comparison of peptide spectra requires normalization. CE migration time and MS-detected mass were normalized by the definition of 339 clusters of peptides covering a range of 19.39 to 37.93 minutes in CE-migration time and 0.830 to 6.456 kDa in molecular mass. Amplitude calibration was based on 38 peptides with  $>60\%$  abundance,  $>100$  counts ion signal intensity above baseline, and  $<130\%$  amplitude deviation.

Detected peptides were deposited, matched, and annotated in a Microsoft SQL database, allowing comparison of multiple samples (patient groups). Peptides were considered identical within different samples when mass deviation was lower than 50 ppm for small peptides and 75 ppm for large peptides and proteins. In the data clustering process, analyte diffusion was compensated by linearly increasing cluster widths over the entire electropherogram (19-45 minutes) from 2%-5%. After calibration, deviation of migration time had to be below 0.35 minutes.

**Statistical Analysis.** Sensitivity, specificity, and 95% confidence intervals (95% CI) were calculated based on receiver operating characteristic (ROC) analysis (MedCalc Software, Belgium).<sup>25</sup> ROC plots were obtained by plotting all sensitivity values (true-positive fraction) on the y axis against their equivalent (1-specificity) values (false-positive fraction) for all available thresholds on the x axis. The area under the ROC curve (AUC) was evaluated, as it provides the single best measure of overall accuracy independent of any threshold.<sup>25</sup> For biomarker discovery, *P*-values were calculated using the natural-logarithm transformed intensities and the Wilcoxon rank sum test.

**Classification.** Disease-type specific peptide marker models were generated using the Support Vector Machine (SVM)-based MosaCluster software.<sup>19</sup> Sample

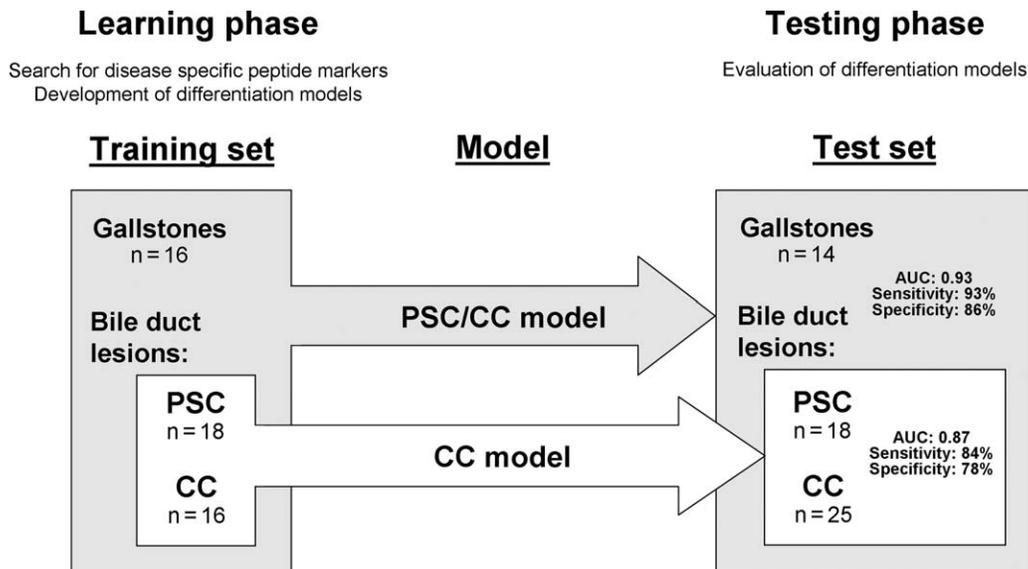


Fig. 1. Study design for the discrimination of controls (gallstones), PSC, and CC by the established proteomic two marker model approach. AUC, sensitivity, and specificity values for the two models are presented for the independent patient sets.

classification was performed by determining the Euclidian distance of a particular dataset to the maximal margin of the SVM hyperplane and assignment of a positive or negative value depending on which side of the hyperplane, case or control, the data point was located.

**Sequencing of Peptides.** Samples were stage tip-purified using Empore Disk C18 as described.<sup>26</sup> The peptides were analyzed by reversed phase chromatography-tandem MS using an LTQ Orbitrap XL (Thermo, Bremen, Germany) coupled to an Agilent 1200 nano-flow-HPLC (high-performance liquid chromatography) (Agilent, Waldbronn, Germany). HPLC-column tips (fused silica) with 75  $\mu$ m inner diameter (New Objective, Woburn, MA) were self-packed with Reprosil-Pur 120 ODS-3 (Dr. Maisch, Ammerbuch, Germany) to a length of 20 cm.<sup>27</sup> Samples were applied directly onto the column without precolumn. The peptides were injected onto the separation column with a linear 140 minutes gradient from 2%-80% B (0.5% acetic acid in 80% acetonitrile [LC-MS grade, Wako, Germany]) in solvent A (0.5% acetic acid [LGC Promochem, Wesel, Germany] in ddH<sub>2</sub>O). The flow rate was 250 nl/min for operation and 500 nl/min for sample application. The mass spectrometer was operated in the data-dependent mode and switched automatically between MS (maximum  $1 \times 10^6$  ions, mass range  $m/z = 350$  to 2,000, resolution 60,000) and MS/MS. Each MS scan was followed by a maximum of five MS/MS scans in the linear ion trap (collision energy 35%, target value 30,000). Singly charged parent ions and unassigned charge states were excluded for fragmentation. MS

parameters were 2.3 kV spray voltage, no sheath, and auxiliary gas flow and 125°C ion-transfer tube temperature.

Individual MS/MS spectra were searched against the IPI human database using the Proteome Discoverer 1.1.0.263 software and an in-house Mascot server (parent mass deviation 10 ppm, fragment ion mass deviation 0.6 Da, decoy database search activated: strict false-discovery rate [FDR] 0.01, relaxed FDR 0.05). An additional search was employed against the NCBI human nonredundant database using the Open Mass Spectrometry Search Algorithm.

## Results

**Biomarker Discovery.** CE-MS measurements revealed high variability in the composition of the low molecular weight proteome in the range of 0.8 to 10 kDa. An average of 1,680 peptides (minimum 469, maximum 3,309) was detected in the 0.8-10 kDa mass range of 1 mg/mL-diluted bile by CE-MS. This high variability in peptide composition necessitates normalization of peptide amplitudes as described in Patients and Methods.

A training set of 50 samples from choledocholithiasis (n = 16), PSC (n = 18), and CC (n = 16) patients was used for the identification of differentially expressed peptides (Fig. 1). We evaluated the data with respect to marker candidates with Wilcoxon *P*-values < 0.05. This resulted in a list of 83 peptides for the differentiation of PSC/CC from choledocholithiasis and 90 for the differentiation of PSC from CC.

**Table 2. Characteristics of Bile Peptides Included in the PSC/CC Marker Model**

CE-MS Characteristics			Sequence Information			Peptide Distribution in the Training Cohort					
						Cholelithiasis (n=16)		PSC/CC (n=34)		P <sup>‡</sup>	AUC
Peptide-ID <sup>†</sup>	Mass [Da]	CE migration time [min]	Sequence	Protein	AA	Frequency	Mean amp.	Frequency	Mean amp.		
164	834.5	22.03	n.i.			0.44	487	0.14	16	0.0157	0.66
1027	922.5	31.99	n.i.			0.56	428	0.23	35	0.0155	0.68
2861	1,053.6	32.52	n.i.			0.25	35	0.58	255	0.0248	0.68
3605	1,104.6	25.14	n.i.			0.56	726	0.14	106	0.0024	0.71
4847	1,213.6	24.59	IKDPDASKPED	Calreticulin	208-218	0.58	5,547	0.25	1,092	0.0100	0.68
5936	1,332.7	34.11	KVVATTQMQAADA	Aminopeptidase N	205-217	0.75	1,496	0.26	366	0.0437	0.66
6289	1,372.7	34.29	LTEPADTITDAVK	Kallikrein-1	126-138	0.67	5,493	0.35	2,185	0.0377	0.64
7101	1,466.8	34.88	TNAVAHVDDMPNAL	Hemoglobin subunit alpha	68-81	0.29	929	0.53	7,231	0.0495	0.62
7143	1,470.8	28.42	DGLAHLNLDKGTFA	Hemoglobin subunit beta	74-87	0.25	37	0.49	2,014	0.0066	0.72
7183	1,474.8	28.06	n.i.			0.56	580	0.23	65	0.0096	0.69
7238	1,479.9	34.98	SLDKFLASVSTVLT	Hemoglobin subunit alpha	125-138	0.29	324	0.58	1,462	0.0118	0.71
7636	1,528.8	27.47	VGAHAGEYGAEALER	Hemoglobin subunit alpha	18-32	0.96	21,948	0.63	10,654	0.0005	0.80
8182	1,596.9	35.53	n.i.			0.19	606	0.43	908	0.0396	0.63
8491	1,632.8	35.37	MLLADQGSWKEEV	Glutathione S-transferase P	20-33	0.50	6,826	0.28	1,674	0.0478	0.63
13670	2,346.3	33.21	n.i.			0.56	282	0.17	34	0.0041	0.70
14835	2,530.4	33.71	n.i.			0.56	1,135	0.14	62	0.0011	0.73
16428	2,809.5	20.90	n.i.			0.13	28	0.46	995	0.0158	0.68
18815	3,307.8	23.73	n.i.			0.06	94	0.43	940	0.0114	0.68

Abbreviations: AA, amino acid; n.i., not identified

<sup>†</sup>Peptide identification numbers; <sup>‡</sup>according to UniProt Knowledge Base numbering; <sup>‡</sup>Wilcoxon *P* values

Shown is the peptide identification number (Peptide-ID), mass (in Da) and CE migration time (in min). For all sequence-identified peptides, the amino acid sequence, the name of the protein precursor and the amino acid positions within the protein's primary sequence (according to the UniProt Knowledge Base) are presented. For accurate sequence annotation, only sequences were accepted with a mass tolerance of less than 10 ppm in MS/MS. In addition, the frequency and the mean amplitude in the choledocholithiasis control and in the PSC/CC case group of the training cohort, as well as the *P* and AUC value of the intergroup comparison of peptide distribution is provided.

On the basis of the two sets of preselected candidate peptide markers, peptide patterns were established to differentiate PSC and CC from choledocholithiasis (PSC/CC model), and in another model to distinguish PSC from CC (CC model). These two models were chosen to construct independent classification schemes for discrimination of sclerosing/malignant lesions from gallstones and of CC from PSC.

**PSC/CC Model to Differentiate Gallstones from PSC or CC.** The PSC/CC model was constructed by selection of 18 out of the 83 PSC/CC peptide marker candidates (Table 2), yielding best classification performance on the training set. This PSC/CC model differentiates PSC and CC from choledocholithiasis with an AUC of 0.90 (95% CI: 0.79-0.97, *P* = 0.0001) in ROC analysis after total cross-validation of training set data (not shown). In Fig. 2A the compiled CE-MS profiles of PSC/CC marker candidate distribution in patients with choledocholithiasis, PSC, and CC are shown representing disease-specific signatures.

Reliability of the PSC/CC peptide model was evaluated by classification of an independent set of 57 patient samples. As presented in Fig. 2B, the PSC/CC model showed an AUC of 0.93 (95% CI: 0.82-0.98,

*P* = 0.0001) to differentiate choledocholithiasis from PSC and CC. At the best cutoff of 0.013 this resulted in correct classification of 12 from 14 choledocholithiasis and 40 from 43 PSC and CC bile samples (86% specificity and 93% sensitivity).

**CC Model Differentiates PSC from CC.** For differentiation of PSC and CC, the CC model with 22 peptides (Table 3) was defined with an AUC of 1.0 (95% CI: 0.9-1.0, *P* < 0.0001) on the training set after cross-validation. Figure 3A displays the compiled CE-MS profiles of the peptides in the CC model for the PSC and CC training set.

Applied to the independent set of samples, the CC model exhibited an AUC of 0.87 (95% CI: 0.73-0.95, *P* = 0.0001) in ROC analysis (Fig. 3B) and was able to identify 14 from 18 PSC and 21 from 25 CC patients correctly at 0.008 as the cutoff (78% specificity, 84% sensitivity). Most notably, 8 out of 10 bile samples of patients with CC on top of PSC scored positive for the CC pattern.

Classification at different timepoints in a minimum and maximum time range between the cholangiography dates of 1 week and 22 months, respectively, resulted in repeated correct classification in 7 of 9 cases

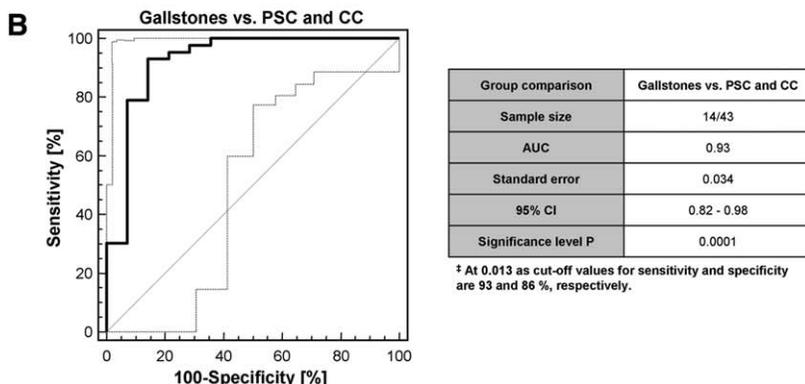
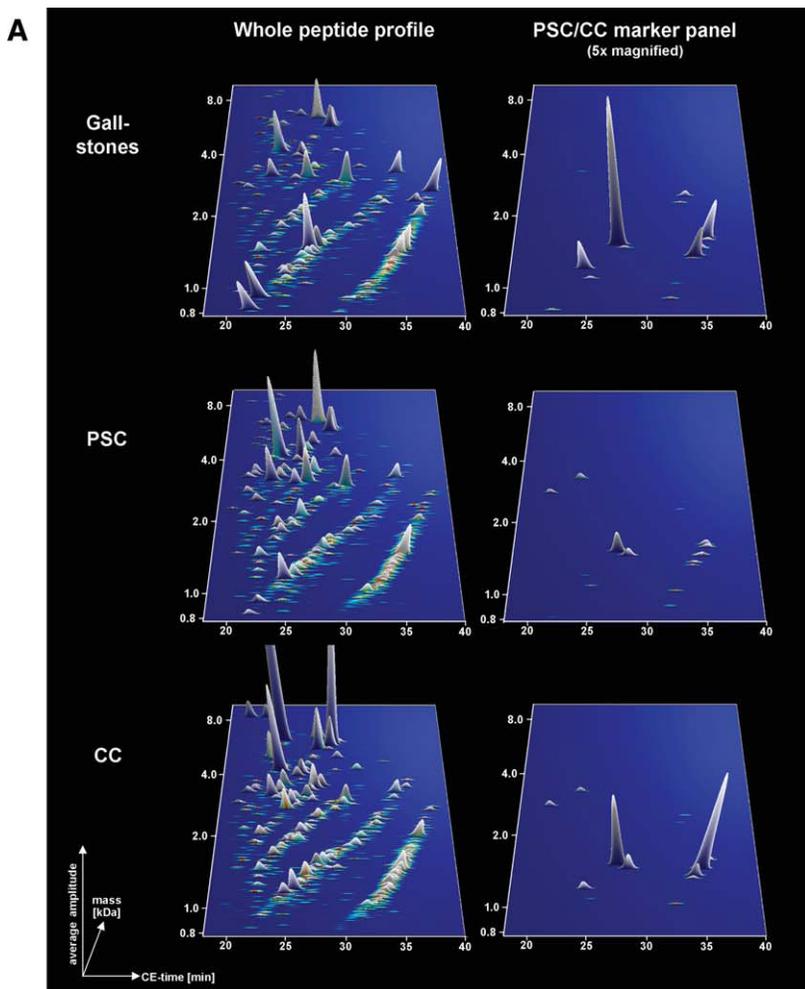


Fig. 2. (A) CE-MS spectra of whole bile peptides in the mass range of 0.8 to 10 kDa (left panels) and peptides included in the PSC/CC specific marker model (right panels) for gallstones (upper row), PSC (middle row), and CC (lower row) samples of the training set. CE migration time (x-axis, 18-40 minutes) is plotted against log molecular weight (y-axis, 0.8-10.0 kDa). Mean signal intensity is given as peak height with a maximum relative intensity of 1.0 (color scale, in arbitrary units) for whole peptide profiles and 0.2 for the PSC/CC specific peptide signature. (B) ROC curve for the differentiation of CC and PSC from gallstones in samples of the validation cohort by the PSC/CC-specific peptide marker model. The table presents values for the respective ROC characteristics such as AUC, 95% CI, and *P*-values. A *P*-value of 0.0001 was regarded as highly significant.

(Supporting Information Table 2). Classification stability of both models was tested by evaluation of three independent measurements of one sample from a patient with CC, one patient with PSC, and one patient with choledocholithiasis. As presented in Supporting Information Table 3, the classification results were not influenced by the number of detected peptides.

**Peptide Sequencing Using Tandem Mass Spectrometry.** To characterize biliary peptides with respect to their amino acid sequence, MS/MS peptide sequencing

was applied.<sup>28,29</sup> The majority of sequence-identified peptides are fragments of hemoglobin alpha and beta chains (Supporting Information Table 1A), followed by peptides of serum albumin, pancreatic triacylglycerol lipase, and cytoplasmic actin 1 (at least 10 different peptides). Other peptides were assigned to structural proteins, i.e., keratins, histones, proteins involved in proteolysis and degradation of lipids and polysaccharides, i.e. proteasome subunits, carboxypeptidases, trypsin, alpha-amylase, bile salt-activated lipase, as well

**Table 3. Characteristics of Bile Peptides Included in the CC Marker Model**

CE-MS Characteristics			Sequence Information			Peptide Distribution in the Training Cohort					
Peptide-ID <sup>†</sup>	Mass [Da]	CE migration time [min]	Sequence <sup>§</sup>	Protein	AA	PSC (n=18)		CCC (n=16)		P <sup>‡</sup>	AUC
						Frequency	Mean amp.	Frequency	Mean amp.		
988	919.4	29.97	n.i.			0.39	88	0.00	0	0.0053	0.69
3085	1,070.6	31.93	n.i.			0.78	437	0.35	185	0.0443	0.69
3169	1,076.6	32.49	n.i.			0.36	455	0.07	20	0.0035	0.73
3456	1,094.6	32.62	AEALERmFL	Hemoglobin subunit alpha	27-35	0.18	24	0.41	463	0.0122	0.71
3700	1,112.6	33.47	n.i.			0.28	369	0.65	1,620	0.0137	0.73
4262	1,159.6	33.56	DKILGATIENS	Keratin, type I cytoskeletal 19	139-149	0.29	2,516	0.03	83	0.0107	0.63
4613	1,190.7	26.59	n.i.			0.61	534	0.35	378	0.0484	0.66
4720	1,200.7	24.98	RQLETLGQEK	Keratin, type II cytoskeletal 8	149-158	0.29	1,101	0.14	94	0.0414	0.59
5103	1,243.7	33.55	LVIENEAGDER	Inter-alpha-trypsin inhibitor heavy chain H2	628-638	0.57	503	0.28	889	0.0383	0.64
6054	1,344.8	31.13	FYAPPELLFFAK	Serum albumin	173-183	0.14	33	0.31	149	0.0099	0.69
6366	1,380.8	26.57	VHLTPPEEKSAVTA	Hemoglobin subunit beta	2-14	0.71	2,257	0.45	1,245	0.0140	0.74
6436	1,388.8	26.57	NELRVAPEEHPV	Actin, cytoplasmic 1	92-103	0.39	855	0.69	21,174	0.0220	0.72
6507	1,397.7	34.85	DLAGRDLTDYlm	Actin, cytoplasmic 1	179-190	0.18	108	0.41	941	0.0336	0.68
7785	1,547.8	35.18	SVTEQGAELSNEER	14-3-3 protein zeta/delta	28-41	0.11	34	0.41	1,878	0.0099	0.69
8084	1,582.8	35.04	n.i.			0.39	575	0.18	6	0.0137	0.64
8386	1,619.8	35.25	n.i.			0.11	29	0.53	209	0.0122	0.71
9032	1,696.9	35.51	ILLTDGDPTVGETNPR	Inter-alpha-trypsin inhibitor heavy chain H4	381-396	0.04	23	0.31	210	0.0090	0.69
10784	1,927.0	36.49	FESFGDLSTPDAVmGNPK	Hemoglobin subunit beta	43-60	0.00	0	0.47	398	0.0013	0.74
11096	1,972.1	25.45	n.i.			0.06	49	0.41	1,753	0.0145	0.68
11106	1,973.1	31.97	n.i.			0.00	0	0.41	315	0.0032	0.71
16759	2,867.5	30.89	VADALTNAVAHVDDMPNA LSALSDLHAH	Hemoglobin subunit alpha	63-90	0.14	289	0.31	1,015	0.0155	0.69
19487	3,468.0	28.42	n.i.			0.39	201	0.06	3	0.0185	0.67

Abbreviations: AA, amino acid; n.i., not identified

<sup>†</sup>Peptide identification numbers; <sup>§</sup>lower case m indicates oxidized methionine; <sup>§</sup>according to UniProt Knowledge Base numbering; <sup>‡</sup>Wilcoxon *P* values

Shown is the peptide identification number (Peptide-ID), mass (in Da) and CE migration time (in min). For all sequence-identified peptides, the amino acid sequence, the name of the protein precursor and the amino acid positions within the protein's primary sequence (according to UniProtKB) are presented. For accurate sequence annotation, only sequences were accepted with a mass tolerance of less than 10 ppm in MS/MS. In addition, the frequency and the mean amplitude in the PSC control and in the CC case group of the training cohort, as well as the *P* and AUC value of the intergroup comparison of peptide distribution is provided.

as proteins involved in immune responses, i.e., complement factors, immunoglobulins. A detailed list of the detected precursor proteins is provided in Supporting Information Table 1B. Correlation of obtained sequence data with the biomarker candidates included in both models revealed differential regulation or proteolysis of hemoglobin alpha and beta chains, cytoplasmic actin, keratins, 14-3-3 zeta/delta, and inter-alpha-trypsin inhibitor heavy chains (Tables 2, 3).

## Discussion

The differentiation between benign and malignant bile duct diseases, particularly strictures, is a very demanding challenge even for specialists in this field.

This study shows, for the first time, that proteomic analysis of bile is able to differentiate malignant bile duct diseases from benign lesions and may become a diagnostic and screening tool in the future.

The analysis of bile to diagnose CC is of particular interest as tumor cells might release and/or shed proteins directly into the bile. Therefore, bile may contain higher levels of secreted or shed markers than serum.<sup>30</sup> Although bile is not easily accessible, it has been shown that bile aspiration during ERC is successful in over 70% of examinations.<sup>21</sup> However, although promising, none of the single markers found in bile has found its way into clinical routine so far.<sup>31,32</sup> A novel approach is the simultaneous analysis of a set of markers that form a specific pattern. The potential of proteomic analysis is obvious: pathological alterations in any organ will result in changes in extracellular proteins. Because such extracellular proteins can be found in the ambient body fluids, proteins in such fluids contain substantial information on the status of organs at the time of sampling. Although blood from first sight appears to be the most appropriate material, its use is associated with several difficulties. The 10 most abundant proteins

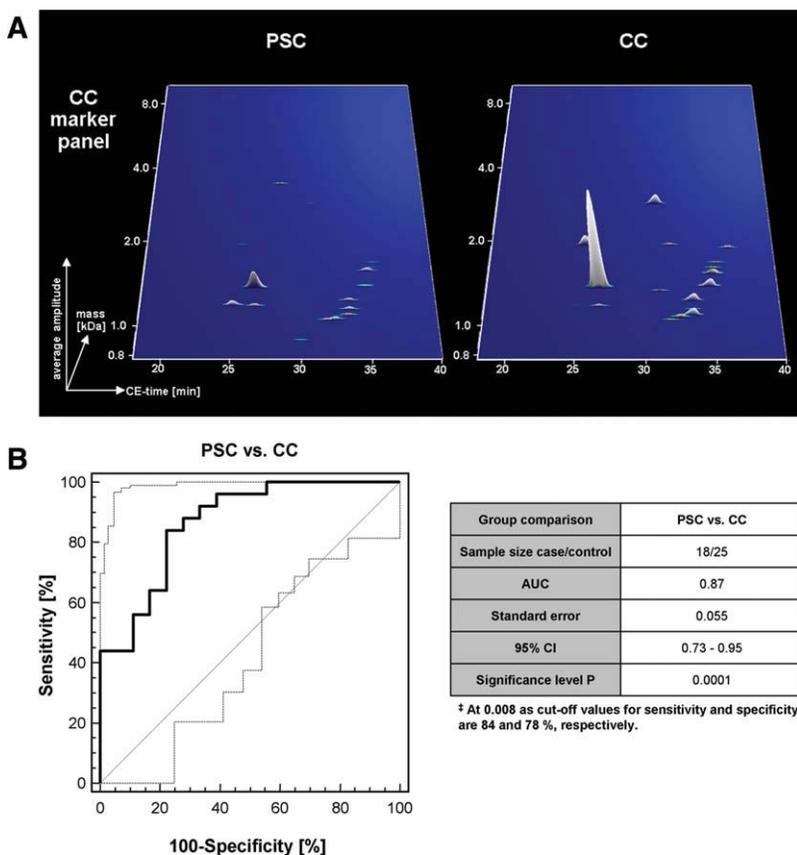


Fig. 3. (A) CE-MS spectra of peptides included in the CC-specific marker panel for PSC (left panel) and CC (right panel) samples of the training set. CE migration time (x-axis, minutes) is plotted against log molecular weight (y-axis, kDa). Mean signal intensity is given as peak height with a maximum relative intensity of 0.2 (color scale, in arbitrary units). (B) ROC curve for the differentiation of CC from PSC in samples of the validation cohort by the CC-specific peptide marker model. The table presents values for the respective ROC characteristics. A *P*-value of 0.0001 was regarded as highly significant.

in blood account for >90% of the total protein content.<sup>33</sup> These most abundant proteins contain little information regarding the status of an organ<sup>33</sup> and greatly inhibit the accurate detection of less abundant proteins that potentially contain more information.<sup>33</sup>

A few preliminary proteomic bile analyses in single patients suggested that this method could be feasible.<sup>13,14,16-18,30,34</sup> Most of the studies applied sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) and liquid chromatography-mass spectrometry (LC-MS). In our study, capillary electrophoresis (CE) coupled to mass spectrometry (MS, CE-MS) was used for the first time to study proteomic profiles in bile.

Although CE and LC possess similar resolution characteristics, the absence of a sieving matrix in CE provides several advantages over LC. No buffer gradients that require ramping of ionization parameters are needed to ensure stable flow; sample migration can be controlled by varying the electric field strength.<sup>35</sup>

The differentiation from PSC and CC was based on a peptide model of 22 peptides. Twelve of those were identified by sequencing and are fragments of hemoglobin subunits, serum albumin, cytoplasmic actin, keratins, inter-alpha-trypsin inhibitor heavy chains, and the 14-3-3 zeta/delta protein (for details, see Table 3). Expression profiles of these peptide markers in PSC

and CC patients may reflect changes in molecular pathways involved in proteolysis/protein catabolism, inflammation, apoptosis, and epithelial cell transformation.<sup>36-38</sup> The increased abundance of a 14-3-3 zeta/delta protein derived peptide fragment (of which the annotated mass spectrum is presented in Supporting Information Fig. 1) in patients with CC is of special interest, because 14-3-3 proteins are involved in many cellular processes, i.e., actin cytoskeletal organization, mitogenesis, cell adhesion, and apoptosis prevention. Hermeking<sup>39</sup> described that cancer-associated down-regulation or loss of 14-3-3 proteins leads to an increased or unscheduled cell-cycle progression. In connection with decreased levels of peptides from cytokeratins and increased levels of actin, 14-3-3 may also be implicated in epithelial-mesenchymal transition,<sup>40,41</sup> with the latter recently associated with CC and invasive CC tumor growth.<sup>42,43</sup> However, further analyses have to be performed to investigate the role of 14-3-3 proteins in the pathogenesis of CC.

A disease-specific biomarker panel to identify different bile-related diseases has never been evaluated so far. Our study focused on clinically diagnostic challenges such as the detection of malignant (CC) and premalignant (PSC) biliary lesions. First, we developed a marker model (PSC/CC model) comparing a PSC

and CC cohort with a choledocholithiasis cohort. For biomarker discovery, the latter was chosen as a control group because the risk of postprocedural pancreatitis or cholangitis ethically bans ERC from application in healthy subjects. However, as the control group consists of patients with choledocholithiasis, proteomic analysis may reflect the difference between a relatively normal biliary tree and liver, and an inflamed, cholestatic liver, which can be expected in patients with PSC and patients with a dominant stenosis due to CC. The PSC/CC model was able to distinguish CC and PSC from nonmalignant lesions with an AUC of 0.93 ( $P = 0.0001$ ), a sensitivity of 93%, and a specificity of 86% as validated in an independent cohort. These findings are of clinical significance, as in patients with suspected CC or PSC endoscopic procedures will be performed and thus bile becomes accessible. Furthermore, even in the presence of large masses suspicious of CC, a definite diagnosis often cannot be made.

The surveillance of patients with PSC is of crucial importance, as those patients have an increased risk to develop CC and curative treatment such as liver transplantation or radical resection can be performed only at an early stage. Therefore, our aim was to distinguish PSC from CC in a second model. This model was established using a training set consisting of 18 patients with PSC and 16 with CC. Applied to an independent validation set (18 PSC, 25 CC) it showed an AUC of 0.87, a specificity of 78%, and a high sensitivity of 84%. Our findings indicate a possible role of proteomic analysis for surveillance in patients with PSC. Nevertheless, PSC-associated CC may be of different origin than sporadic cholangiocarcinoma. Ten patients within the CC group developed CC in addition to PSC. Eight of those patients were identified positive for CC by proteomic analysis.

This proteomic model reaches a high sensitivity compared to single biochemical markers. Direct comparison with the widely used CA 19-9 tumor marker is impossible, because previous studies used different cut-off values in various study populations, leading to enormous range of sensitivity (53%-92%) and specificity (50%-98%).<sup>44</sup> Our proposed model may be of clinical relevance in diagnosing CC in patients with PSC especially if supplementary to other diagnostic methods, as a higher accuracy can be reached by a combination of different diagnostic tools.<sup>45</sup> All in all, proteomic analysis of bile as a diagnostic tool for surveillance of patients with PSC alone or in combination with other methods may provide an early and reliable diagnosis of CC.

In summary, our data indicate a possible role of proteomic analysis of bile to differentiate CC from PSC and benign lesions. Due to the cross-sectional type of

the study with histologically and clinically well-defined CC, no information exists as to what timepoint the CC model is able to detect CC in addition to PSC during its evolution in the PSC patient. This important issue must be addressed in a future prospective multicenter clinical trial by inclusion of patients with PSC in whom ERC is performed regularly to treat strictures and with a clinical follow-up of 1 year for CC diagnosis as requirement for reliability of the longitudinal analysis.

## References

1. Khan SA, Thomas HC, Davidson BR, Taylor-Robinson SD. Cholangiocarcinoma. *Lancet* 2005;366:1303-1314.
2. Demols A, Marechal R, Deviere J, Van Laethem JL. The multidisciplinary management of gastrointestinal cancer. Biliary tract cancers: from pathogenesis to endoscopic treatment. *Best Pract Res Clin Gastroenterol* 2007;21:1015-1029.
3. Weismuller TJ, Wedemeyer J, Kubicka S, Strassburg CP, Manns MP. The challenges in primary sclerosing cholangitis—etiopathogenesis, autoimmunity, management and malignancy. *J Hepatol* 2008;48 Suppl 1:S38-57.
4. Park MS, Kim TK, Kim KW, Park SW, Lee JK, Kim JS, et al. Differentiation of extrahepatic bile duct cholangiocarcinoma from benign stricture: findings at MRCP versus ERCP. *Radiology* 2004;233:234-240.
5. Rosch T, Meining A, Fruhmorgen S, Zillinger C, Schusdziarra V, Hellerhoff K, et al. A prospective comparison of the diagnostic accuracy of ERCP, MRCP, CT, and EUS in biliary strictures. *Gastrointest Endosc* 2002;55:870-876.
6. de Bellis M, Fogel EL, Sherman S, Watkins JL, Chappo J, Younger C, et al. Influence of stricture dilation and repeat brushing on the cancer detection rate of brush cytology in the evaluation of malignant biliary obstruction. *Gastrointest Endosc* 2003;58:176-182.
7. Boberg KM, Jepsen P, Clausen OP, Foss A, Aabakken L, Schrupf E. Diagnostic benefit of biliary brush cytology in cholangiocarcinoma in primary sclerosing cholangitis. *J Hepatol* 2006;45:568-574.
8. Tischendorf JJ, Hecker H, Kruger M, Manns MP, Meier PN. Characterization, outcome, and prognosis in 273 patients with primary sclerosing cholangitis: A single center study. *Am J Gastroenterol* 2007;102:107-114.
9. Bergquist A, Ekbohm A, Olsson R, Kornfeldt D, Loof L, Danielsson A, et al. Hepatic and extrahepatic malignancies in primary sclerosing cholangitis. *J Hepatol* 2002;36:321-327.
10. Walker SL, McCormick PA. Diagnosing cholangiocarcinoma in primary sclerosing cholangitis: an "evidence based radiology" review. *Abdom Imaging* 2008;33:14-17.
11. Ohshio G, Manabe T, Watanabe Y, Endo K, Kudo H, Suzuki T, et al. Comparative studies of DU-PAN-2, carcinoembryonic antigen, and CA19-9 in the serum and bile of patients with pancreatic and biliary tract diseases: evaluation of the influence of obstructive jaundice. *Am J Gastroenterol* 1990;85:1370-1376.
12. Ker CG, Chen JS, Lee KT, Sheen PC, Wu CC. Assessment of serum and bile levels of CA19-9 and CA125 in cholangitis and bile duct carcinoma. *J Gastroenterol Hepatol* 1991;6:505-508.
13. Farina A, Dumonceau JM, Frossard JL, Hadengue A, Hochstrasser DF, Lescuyer P. Proteomic analysis of human bile from malignant biliary stenosis induced by pancreatic cancer. *J Proteome Res* 2009;8:159-169.
14. Farina A, Dumonceau JM, Lescuyer P. Proteomic analysis of human bile and potential applications for cancer diagnosis. *Expert Rev Proteomics* 2009;6:285-301.
15. Gronborg M, Bunkenborg J, Kristiansen TZ, Jensen ON, Yeo CJ, Hruban RH, Maitra A, et al. Comprehensive proteomic analysis of human pancreatic juice. *J Proteome Res* 2004;3:1042-1055.
16. Guerrier L, Claverol S, Finzi L, Paye F, Fortis F, Boschetti E, Housset C. Contribution of solid-phase hexapeptide ligand libraries to the repertoire of human bile proteins. *J Chromatogr A* 2007;1176:192-205.

17. Kristiansen TZ, Bunkenborg J, Gronborg M, Molina H, Thuluvath PJ, Argani P, Goggins MG, et al. A proteomic analysis of human bile. *Mol Cell Proteomics* 2004;3:715-728.
18. Zhou H, Chen B, Li RX, Sheng QH, Li SJ, Zhang L, Li L, et al. Large-scale identification of human biliary proteins from a cholesterol stone patient using a proteomic approach. *Rapid Commun Mass Spectrom* 2005;19:3569-3578.
19. Theodorescu D, Wittke S, Ross MM, Walden M, Conaway M, Just I, Mischak H, et al. Discovery and validation of new protein biomarkers for urothelial cancer: a prospective analysis. *Lancet Oncol* 2006;7:230-240.
20. Metzger J, Schanstra JP, Mischak H. Capillary electrophoresis-mass spectrometry in urinary proteome analysis: current applications and future developments. *Anal Bioanal Chem* 2009;393:1431-1442.
21. Negm AA, Schott A, Vonberg RP, Weismueller TJ, Schneider AS, Kubicka S, Strassburg CP, et al. Routine bile collection for microbiological analysis during cholangiography and its impact on the management of cholangitis. *Gastrointest Endosc*;72:284-291.
22. Kolch W, Neuss C, Pelzing M, Mischak H. Capillary electrophoresis-mass spectrometry as a powerful tool in clinical diagnosis and biomarker discovery. *Mass Spectrom Rev* 2005;24:959-977.
23. Wittke S, Fliser D, Haubitz M, Bartel S, Krebs R, Hausadel F, Hillmann M, et al. Determination of peptides and proteins in human urine with capillary electrophoresis-mass spectrometry, a suitable tool for the establishment of new diagnostic markers. *J Chromatogr A* 2003;1013:173-181.
24. Neuhoﬀ N, Kaiser T, Wittke S, Krebs R, Pitt A, Burchard A, Sundmacher A, et al. Mass spectrometry for the detection of differentially expressed proteins: a comparison of surface-enhanced laser desorption/ionization and capillary electrophoresis/mass spectrometry. *Rapid Commun Mass Spectrom* 2004;18:149-156.
25. DeLeo J. Receiver operating characteristic laboratory (ROCLAB): software for developing decision strategies that account for uncertainty. College Park, MD: Second International Symposium on Uncertainty Modeling and Analysis 1993:318-325.
26. Rappsilber J, Ishihama Y, Mann M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* 2003;75:663-670.
27. Gruhler A, Olsen JV, Mohammed S, Mortensen P, Faergeman NJ, Mann M, et al. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics* 2005;4:310-327.
28. Zurbig P, Renfrow MB, Schiffer E, Novak J, Walden M, Wittke S, et al. Biomarker discovery by CE-MS enables sequence analysis via MS/MS with platform-independent separation. *Electrophoresis* 2006;27:2111-2125.
29. Zimmerli LU, Schiffer E, Zurbig P, Good DM, Kellmann M, Mouis L, et al. Urinary proteomic biomarkers in coronary artery disease. *Mol Cell Proteomics* 2008;7:290-298.
30. Bonney GK, Craven RA, Prasad R, Melcher AF, Selby PJ, Banks RE. Circulating markers of biliary malignancy: opportunities in proteomics? *Lancet Oncol* 2008;9:149-158.
31. Koopmann J, Thuluvath PJ, Zahurak ML, Kristiansen TZ, Pandey A, Schulick R, et al. Mac-2-binding protein is a diagnostic marker for biliary tract carcinoma. *Cancer* 2004;101:1609-1615.
32. Alvaro D, Macarri G, Mancino MG, Marzioni M, Bragazzi M, Onori P, et al. Serum and biliary insulin-like growth factor I and vascular endothelial growth factor in determining the cause of obstructive cholestasis. *Ann Intern Med* 2007;147:451-459.
33. Anderson NL, Anderson NG. The human plasma proteome — history, character, and diagnostic prospects. *Mol Cell Proteom* 2002;1:845-867.
34. He C, Fischer S, Meyer G, Muller I, Jungst D. Two-dimensional electrophoretic analysis of vesicular and micellar proteins of gallbladder bile. *J Chromatogr A* 1997;776:109-115.
35. Gaspar A, Englmann M, Fekete A, Harir M, Schmitt-Kopplin P. Trends in CE-MS 2005-2006. *Electrophoresis* 2008;29:66-79.
36. Correale M, Brunetti ND, De Gennaro L, Di Biase M. Acute phase proteins in atherosclerosis (acute coronary syndrome). *Cardiovasc Hematol Agents Med Chem* 2008;6:272-277.
37. Coutts AS, Weston L, La Thangue NB. Actin nucleation by a transcription co-factor that links cytoskeletal events with the p53 response. *Cell Cycle* 2010;19:9.
38. Zhuo L, Kimata K. Structure and function of inter-alpha-trypsin inhibitor heavy chains. *Connect Tissue Res* 2008;49:311-320.
39. Hermeking H. The 14-3-3 cancer connection. *Nat Rev Cancer* 2003;3:931-943.
40. Hou Z, Peng H, White DE, Wang P, Lieberman PM, Halazonetis T, et al. 14-3-3 binding sites in the snail protein are essential for snail-mediated transcriptional repression and epithelial-mesenchymal differentiation. *Cancer Res*;70:4385-4393.
41. Kalluri R, Neilson EG. Epithelial-mesenchymal transition and its implications for fibrosis. *J Clin Invest* 2003;112:1776-1784.
42. Lee MJ, Yu GR, Yoo HJ, Kim JH, Yoon BI, Choi YK, et al. ANXA8 down-regulation by EGF-FOXO4 signaling is involved in cell scattering and tumor metastasis of cholangiocarcinoma. *Gastroenterology* 2009;137:1138-1150, 1150 e1131-1139.
43. Sato Y, Harada K, Itatsu K, Ikeda H, Kakuda Y, Shimomura S, et al. Epithelial-mesenchymal transition induced by transforming growth factor- $\beta$ 1/snail activation aggravates invasive growth of cholangiocarcinoma. *Am J Pathol* 2010;177:141-152.
44. Alvaro D. Serum and bile biomarkers for cholangiocarcinoma. *Curr Opin Gastroenterol* 2009;25:279-284.
45. Charatcharoenwithaya P, Enders FB, Halling KC, Lindor KD. Utility of serum tumor markers, imaging, and biliary cytology for detecting cholangiocarcinoma in primary sclerosing cholangitis. *Hepatology* 2008;48:1106-1117.