RESEARCH ARTICLE

Editorial Process: Submission:01/27/2022 Acceptance:06/13/2022

Targeted Proteins Reveal Cathepsin D as a Novel Biomarker in Differentiating Hepatocellular Carcinoma from Cirrhosis and Other Liver Cancers

Natthaya Chuaypen[1], Supachaya Sriphoosanaphan[2], Athaya Vorasittha[3], Nutcha Pinjaroen[4], Sakdithep Chaiyarit[5,6], Visith Thongboonkerd[5,6], Pisit Tangkijvanich[1], Pongserath Sirichindakul[3]*

Abstract

Objective: Hepatocellular carcinoma (HCC) represents a global health concern, particularly in Southeast Asia where hepatitis B virus (HBV) infection is common. In this study, we applied tissue-based proteomics to identify novel serological proteins for HCC and validated their performance in serum specimens. **Methods:** In a discovery set, liver tissue specimens of HBV-related HCC, intrahepatic cholangiocarcinoma (iCCA) and colorectal cancer with liver metastasis (CRLM) were analyzed using mass spectrometry (LTQ-Orbitrap-XL). A subset of proteins that showed highly expressed in HCC were then confirmed by Western blotting. Additionally, clinical significance of selected candidate proteins was tested in serum samples of 80 patients with HBV-related HCC, 50 patients with HBV-related liver cirrhosis and 30 healthy controls. **Results:** Based on LTQ-Orbitrap-XL mass spectrometer, various differentially expressed proteins (DEPs) between tumor and adjacent non-tumor tissues were identified. These included 77 DEPs for HCC, 77 DEPs for iCCA and 55 DEPs for CRLM. Among selected candidate proteins, annexin A2 and cathepsin D were confirmed to be overexpressed in HCC tissue by Western blot analysis. In a validate cohort, serum cathepsin D level, but not annexin A2, was significantly higher in HCC compared with the non-HCC groups. Serum cathepsin D level was also positively correlated with tumor size and tumor stage. Additionally, the combined assay of serum cathepsin D and alpha-fetoprotein had a high sensitivity in detecting early HCC (83%) and intermediate/advanced HCC (96%). Moreover, patients with low serum cathepsin D (<305 ng/mL) displayed significantly better overall survival than those whose serum levels were high (≥305 ng/mL). **Conclusions:** Proteomics and subsequent validation revealed cathepsin D as a novel biomarker for HCC. Apart from its diagnostic role, serum cathepsin D might also serve as a prognostic biomarker of HCC. Additional large-scale studies are needed to verify our findings.

Keywords: Cathepsin D- proteomics- biomarker- HCC- iCCA- CRLM

Asian Pac J Cancer Prev, **23 (6)**, 2017-2025

Introduction

Primary liver cancer (PLC) represents one of the most frequent and aggressive malignancies in global scale with its relatively low survival rates resulting from inadequate early diagnosis (Sia et al., 2017). Histologically, PLC is mainly comprised of two types of cancers namely hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (iCCA) (Sia et al., 2017). HCC, originating from the hepatocytes, has a high incidence found in Southeast Asia where hepatitis B virus (HBV) infection is common (Forner et al., 2018). The estimated incidence of HCC in Thailand is 38.6 and 17.2 per 100,000 person-years in males and females, respectively. For iCCA, this adenocarcinoma arises in the intrahepatic bile ducts and remains the second most common PLC after HCC (Sia et al., 2017). Despite its relative scarcity in global scale, iCCA is considerably prevalent in Southeast Asia, especially in the areas of Mekong River Basin countries with widespread liver fluke infestation (Mathema et al., 2015). For instance, Khon Kaen province in Thailand, is considered to be among

*[1]Center of Excellence in Hepatitis and Liver Cancer, Department of Biochemistry, Faculty of Medicine, Chulalongkorn University, Bangkok, 10330, Thailand. [2]Division of Gastroenterology, Department of Medicine, Faculty of Medicine, Chulalongkorn University, Bangkok, 10330, Thailand. [3]Department of Surgery, Faculty of Medicine, Chulalongkorn University, Bangkok, 10330, Thailand. [4]Department of Radiology, Faculty of Medicine, Chulalongkorn University, Bangkok, 10330, Thailand. [5]Medical Proteomics Unit, Office for Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand. [6]Center for Research in Complex Systems Science, Mahidol University, Bangkok, Thailand. *For Correspondence: pongserath.s@gmail.com*

the highest overall age-standardized incidence of iCCA worldwide. Thus, in the geographical areas where chronic viral hepatitis and liver flukes are common, it is essential to differentiate HCC from iCCA as both cancers might present with similar clinical features. Besides PLC, the liver is a common site for tumors spreading from other organs, accounting for 25% of all cases. Among these, colorectal cancer with liver metastasis (CRLM) is the most common metastatic tumors undergoing hepatic resection (Bray et al., 2018). As their prognosis and management differ significantly, distinguishing HCC from iCCA and CRLM remains challenging in clinical practice.

Currently, serum alpha-fetoprotein (AFP), a fetal-specific glycoprotein, is the most commonly used biomarker for HCC (Sauzay et al., 2016). Since approximately 90% of HCC occur within underlying liver cirrhosis (LC), it is recommended that cancer surveillance should be performed in high-risk groups using ultrasound with or without AFP (Heimbach et al., 2018). Nonetheless, AFP has some limitations including lack adequate sensitivity in an early-stage HCC and its concentration could also be elevated in non-malignant liver diseases. Moreover, high serum AFP levels have been reported in some cases with iCCA (Sauzay et al., 2016). Thus, identifying reliable biomarkers participating in molecular pathogenesis of HCC will play a key role in early diagnosis and predicting disease prognosis. Recent evidence in molecular biology have demonstrated that repetitive destruction and regeneration of the hepatocytes are accountable for HCC occurrence (Arzumanyan et al., 2013). Among essential biological mechanisms, aberrant protein expressions are shown to be involved in the multi-step processes of hepatic oncogenesis (Xing et al., 2016). These data indicate the potential use of high-throughput quantitative proteomic techniques to identify differentially expressed proteins as novel biomarkers for diagnosis and monitoring of liver cancers (Beretta 2009; Xing et al., 2016). Currently, less evidence is focusing on new diagnostic biomarkers to distinguish HCC from iCCA and CRLM.

In this study, we aimed at assessing proteomic profiling of HCC in differentiating with the other types of liver cancer. Furthermore, a subset of proteins that showed highly expressed in HCC were confirmed by Western blotting in liver specimens. Additionally, we validated clinical significance of candidate proteins in another set of serum samples obtained from patients with HCC, LC and healthy controls.

Materials and Methods

Patients' tissues and blood collection

To investigate the protein profiling, pair tissues of tumor and adjacent non-tumor from patients with HBV-related HCC, iCCA and CRLM who had undergone surgical resection at the Department of Surgery, Chulalongkorn University from 2017 to 2020 were collected. The diagnosis of tumor type was confirmed by histopathological examination. Liver tissue specimens obtained from surgical resection were immediately frozen in liquid nitrogen until analysis.

For a validate cohort in serum samples of patients with HBV-related HCC, the diagnostic of this cancer was established by imaging studies based on dynamic computed tomography (CT) or magnetic resonance imaging (MRI) according to the American Association for the Study of Liver Diseases (AASLD) guideline (Heimbach et al., 2018). Baseline clinical and laboratory parameters of these patients were collected. The overall survival (OS) of these patients characterized by interval between initial recruitment and death or the last follow-up visit was also recorded. For the control groups, serum samples were taken from healthy controls and patients with HBV-related LC without evidence of HCC. The diagnosis of LC was based on clinical and imaging evidence and confirmed by transient elastography (TE) (F4 >12.5 kPa)(Singh et al., 2017). The diagnosis of chronic HBV infection was documented by the positivity of serum hepatitis B s antigen (HBsAg) for at least 6 months. Patients who were infected with hepatitis C virus (HCV) and/or human immunodeficiency virus (HIV) were excluded. All of the collected serum samples were stored at -70°C until assay.

The study was performed according to the principles of the Declaration of Helsinki and approved by the Institute Ethics Committee of Faculty of Medicine, Chulalongkorn University. All participants provided written informed consent prior to the enrollment.

Protein extraction

Tissue samples from both tumor and adjacent non-tumor of each patient were extracted for proteins. Briefly, the tissue was dissected into small pieces and snap-freeze by immersing in liquid nitrogen then add ice-cold lysis buffer and homogenize using electric homogenizer. After these processes, the tissues were agitated for 2 hours at 4°C. The supernatant was collected following centrifugation at 16,000×g and 4°C for 20 minutes and total protein concentration was determined by Bio-Rad Protein Assay (Bio-Rad Laboratories; Hercules, CA).

In-solution tryptic digestion by filter-aided sample preparation (FASP) method, phosphopeptide enrichment by TiO2 and nanoLC-ESI-LTQ-Orbitrap MS/MS.

Equal amount of total protein derived from each sample was pooled and digested by trypsin according to FASP protocol as described previously (Thanomkitti et al., 2018). Finally, the peptides were resuspended in 0.1% formic acid and then subjected to phosphopeptide enrichment by TiO2 and nanoLC-ESI-LTQ-Orbitrap MS/MS as described previously (Thongboonkerd & Chaiyarit, 2022).

MS/MS spectral interpretation and quantitative analysis.

Data were analyzed using Proteome Discoverer v.1.4.1.14 software (Thermo Scientific). Data were searched against SwissProt database of mammalian using Mascot software version 2.4.0 (Matrix Science; London, UK) with standard Mascot parameters for CID (Enzyme = trypsin, maximal number of missed cleavages = 2, peptide tolerance = ±50 ppm, MS/MS tolerance = ±0.8 Da, fixed modification = carbamidomethyl (C), variable modification = oxidation (M), phosphorylation (S, T and

Y), charge states = 2+ and 3+, and decoy database on FDR <1%). Quantitative data of three most abundant peptides identified from each protein with average value of areas under curve (AUC) was considered. Background was subtracted from all peak areas (Thanomkitti et al., 2018).

Western blot analysis

For Western blot analysis, total protein from liver tissues (adjacent non-tumor and tumor) were extracted and collected in Laemmli lysis buffer. Total protein level was measured by Bio-Rad protein assay (Bio-Rad Laboratories, Inc., USA). Then, protein concentration at 50 ug/lane from each sample was run on 10% SDS-PAGE and transferred to nitrocellulose blotting membranes (GE Healthcare Life science, Germany). Subsequently, the membranes were blocked in 5%BSA solution (0.05% Tween 20) for 1 h at room temperature and then incubated with mouse anti-Cathepsin D (1:1,000, ab6313; Abcam, Cambridge, UK) and mouse anti-Annexin II (1:1,000, sc-28385; Santa Cruz Biotechnology Inc., Santa Cruz, CA, USA) for 16 hours at 4°C. Membranes were washed and incubated with Mouse IgGκ light chain binding protein (m-IgGκ BP) conjugated to horseradish peroxidase (HRP) (1:5,000, sc-516102, Santa Cruz Biotechnology Inc., Santa Cruz, CA, USA) for 2 hours. After that, the membranes were probed with Amersham ECL Prime Western Blotting Detection Reagent (Cytiva, UK). The immunoreactive protein bands were visualized by ChemiDoc MP Imaging System (Bio-Rad Laboratories, Inc., USA). Band intensity data was identified using Image analysin in JAVA (international institute of health, USA). GAPDH was used as the internal control.

Enzyme-linked immunosorbent assay (ELISA)

The quantitative of annexin A2 expression in serum samples was performed according to manufacturer's protocol. Briefly, serum samples were diluted at 1:10 in PBS and added in duplicate into the plate (ELISA Assay Kit for Annexin A2, Product No. SEB944Hu, CLOUD-CLONE CORP, USA) and then incubated for 1 h at 37 °C. After that, the Detection Reagent A and Detection Reagent B was added and incubated at 37 °C for 30 minutes. After washing, TMB Substrate Solution was added and then the reaction was conducted using a spectrophotometer at 450 nm.

For circulating cathepsin D expression, serum samples were diluted at 1:100 in sample diluent buffer and added in duplicate into the plate (ab119586-cathepsin D human Elisa kit, Abcam, Cambridge, UK) following manufacturer's instructions. Briefly, 1X Biotinylated anti-Human Cathepsin D antibody was added into each well and incubate at 37°C for 60 minutes. After each incubation, the plates were washed and added 1X Avidin-Biotin-Peroxidase Complex working solution. In the final step, TMB agent was added into each well and the reaction was stopped with TMB Stop Solution. The results were reported as O.D. using a spectrophotometer at 450 nm.

Global protein network analysis

Protein molecular and biological functions; Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway were analyzed using PANTHER classification system (Mi et al., 2016). Protein–protein interaction of differential upregulated proteins from patients were analyzed using Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) version 10.0 (http://string.embl.de).

Statistical analysis

Data were reported as percentage or mean ± standard deviation, as appropriate. Comparisons between groups were tested by unpaired Student's t-test or Mann-Whitney U test. The receiver-operating characteristics (ROC) curves were generated to calculate the diagnostic performance of serum biomarkers in discriminating HCC from the non-HCC groups. The Pearson correlation coefficient was applied to analyze the correlation between levels of serum biomarkers and related parameters. The assessment of overall survival of patients with HCC was determined by the Kaplan-Meier method and difference between groups was assessed by the log-rank test. P values less than 0.05 were considered statistically significant. Statistical analyses were undertaken by the SPSS software for windows 22.0 (SPSS Inc., Chicago, IL).

Results

Differential protein expression profiles in liver cancers

To investigate the proteomic profile in the discovery set, 34 pairs of tumor and adjacent non-tumor tissues of patients with HCC (12 cases), iCCA (12 cases) and CRLM (10 cases) were used to identify the different protein expression profile. Detailed baseline demographic and characteristics of these patients were shown in Supplementary Table 1.

To identify significant differentially expressed proteins (DEPs) in HCC, iCCA and CRLM tissues, nanoLC-ESI-LTQ-Orbitrap tandem mass spectrometry (MS/MS) was performed with high-stringent criteria. We observed the differential expression patterns of proteins among the studied groups. In particular, significantly DEPs between tumor tissues and adjacent non-tumor tissues including 77 DEPs for HCC, 77 DEPs for iCCA and 55 DEPs for CRLM. Their identities, identification scores, related mass spectral parameters and quantitative data of HCC was summarized in Table 1, while the data regarding iCCA and CRLM were shown in Supplementary Table 2 and 3 respectively. From the discovery sets, 34 proteins were upregulated, and 15 proteins were downregulated in HCC. Additionally, 14 were upregulated and 21 were downregulated in iCCA, and 25 were upregulated and 3 were downregulated in CRLM.

Gene ontology, KEGG pathway and protein network analysis

The upregulated proteins in HCC were subsequently analyzed to identify the function of respective DEPs. From Gene Ontology (GO) and KEGG pathway enrichment analysis, upregulated proteins were involved in biological process, cellular component and molecular function (Figure 1). These proteins were commonly located in cellular process (15, 23.8%), metabolic process (10,

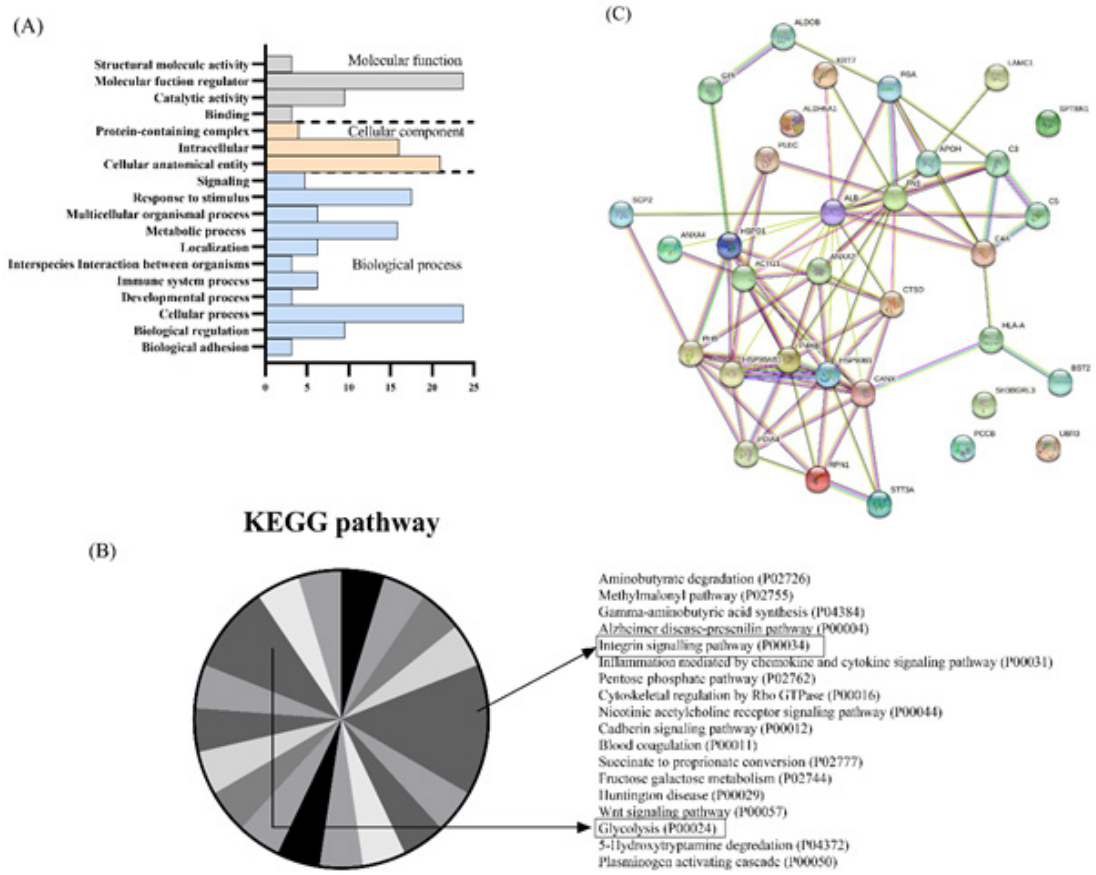

Figure 1. Data Mining of the up-Regulated Proteins in HCC Tissues. (A) GO analysis; (B) KEGG pathway analysis; and (C) STRING analysis.

15.9%), responsive to stimulus (11, 17.5%), cellular anatomical entity (21, 51.2%), intracellular functions (16, 39.0%), protein binding (11, 47.8%), catalytic activity (8, 34.8%) and structural molecular activity (3, 13.0%) (Figure 1A). In addition, the proteins were also enriched in the following pathways, including integrin signaling pathway, glycolysis, aminobutyrate degradation, methyl malonyl pathway, gamma-aminobutyric acid synthesis, Alzheimer disease-presenilin pathway and pentose phosphate pathway (Figure 1B). Protein-protein interactions were identified among all the upregulated proteins (Figure 1C). These significantly enriched GO and KEGG pathway terms could facilitate a better understanding the role of the DEPs in the tumorigenesis of HBV-related HCC.

Assessment of candidate tissue proteins by Western blot analysis

To confirm the results of proteomics, the greatly up-regulated proteins including annexin A2, calnexin,

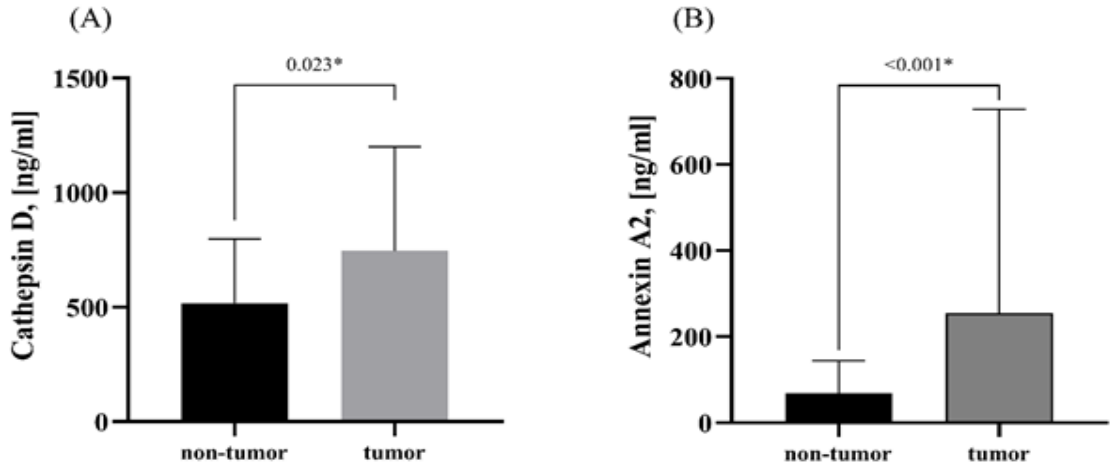

Figure 2. Protein Expression in Liver Tissues of Patients with HCC. (A) Cathepsin D level; (B) Annexin A2 level

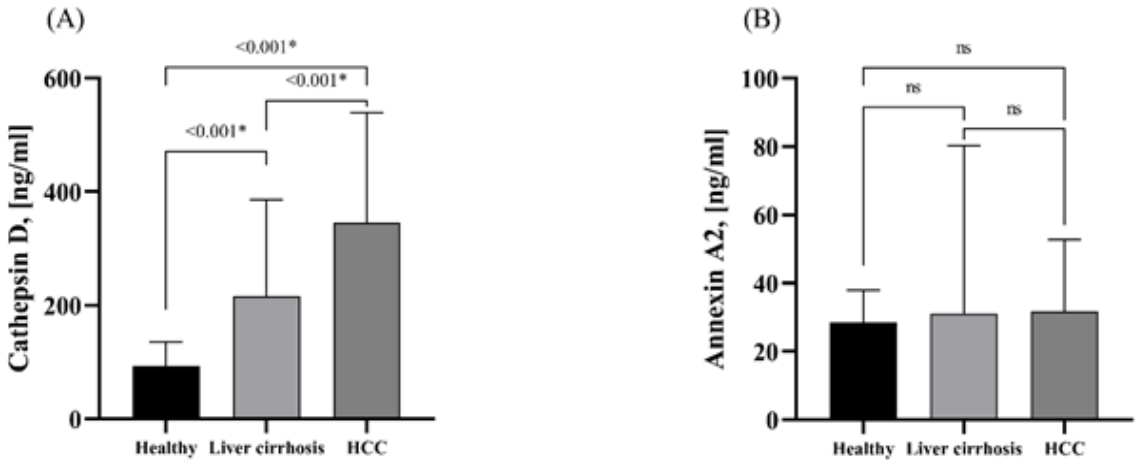

Figure 3. Serum Proteins of Studied Groups. (A) Cathepsin D level; (B) Annexin A2 level

cathepsin D, plectin and prohibitin were selected to verify the protein expression levels in tissue samples using Western blot analysis. Compared with adjacent non-tumor tissues, only annexin A2 and cathepsin D were significantly up regulated in HCC tissues (P<0.001 and P=0.023, respectively) as shown in Figure 2A and 2B. Moreover, there was no significant different between tumor and non-tumor tissues in tissue samples of iCCA and CRLM (data not shown).

Validation of protein expression in serum samples by ELISA

We subsequently validated the expression annexin A2 and cathepsin D in serum samples of 80 patients with HCC, 50 patients with LC and 30 healthy controls using commercial ELISA kits. Our results demonstrated that cathepsin D levels were significantly increased in patients with HCC (345.9±192.8 ng/mL) and LC (216.0±170.2 ng/mL) compared with healthy controls (93.3±41.9 ng/mL) (P<0.001). In addition, significant difference of cathepsin D was also observed between patients with HCC and LC (P<0.001; Figure 3A), which was comparable to the results of protein expression in liver tissues detected by Western blot analysis. Regarding serum annexin A2, there was no significant difference among HCC (31.7±21.0 ng/mL), LC (31.4±48.6 ng/mL) and healthy controls (28.5±9.4 ng/mL)(P> 0.05; Figure 3B).

Diagnostic performance of serum biomarkers

We further conducted the diagnostic performance of serum biomarkers in differentiating HCC from LC. As shown in Figure 4, the area under the curve (AUROC) of HCC and LC was 0.728 [95 % confidence interval (CI); 0.640-0.817, P<0.001] for AFP, 0.780 (95%CI; 0.691-0.869, P<0.001) for cathepsin D and 0.600 (95%CI; 0.488-0.711, P=0.069) for annexin A2. These data confirmed that cathepsin D, but not annexin A2, was a potential biomarker for HCC and its diagnostic performance was further evaluated.

Based on the AUROC, a cut-off level of serum cathepsin D level reflected the highest accuracy for differentiating HCC from LC was 184 ng/mL. At this concentration, the overall sensitivity and specificity of this

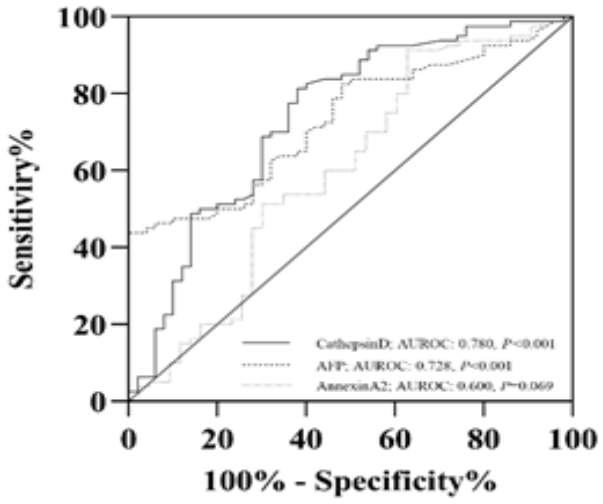

Figure 4. Diagnostic Value of AFP, Cathepsin D and Annexin A2 Using ROC Curve Analysis

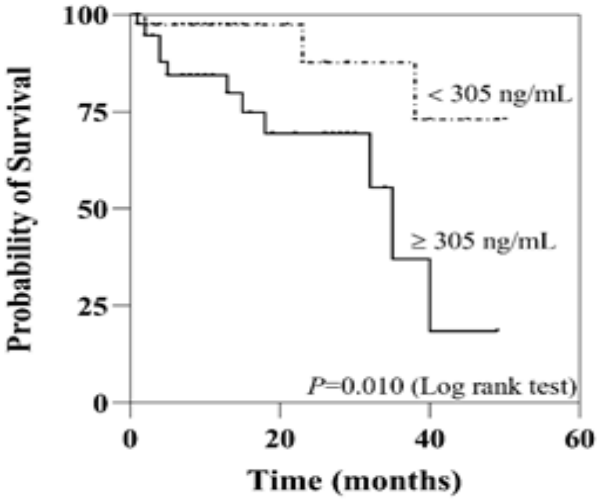

Figure 5. Kaplan–Meier Survival Analysis of Serum Cathepsin D in Patients with HCC

Table 1. Summary of Significantly Altered Proteins Levels in HCC Tissue vs. adjacent Non-Tumor Tissue as Determined by Quantitative Proteomics

Accession no.	Protein name	MS/MS identification score	%Cov	MW (kDa)	Abundance level (×10⁶ arbitrary unit) (Mean±SEM)		Ratio (Cancer/ Normal)	P-value
					Normal	Cancer		
HCC								
Proteins that were newly present in tumor tissues								
P27348	14-3-3 protein theta	135.31	16.73	27.7	0.0000±0.0000	3.3163±0.5077	#DIV/0!	0.0226
Q99714	3-hydroxyacyl-CoA dehydrogenase type-2	45.84	6.51	26.9	0.0000±0.0000	1.0526±0.0294	#DIV/0!	0.0008
P21810	Biglycan	52.48	14.95	41.6	0.0000±0.0000	5.2693±0.7273	#DIV/0!	0.0185
Q9H444	Charged multivesicular body protein 4b	58.46	13.39	24.9	0.0000±0.0000	4.5173±0.1997	#DIV/0!	0.0019
P12109	Collagen alpha-1(VI) chain	56.26	12.65	108.5	0.0000±0.0000	6.6437±0.2223	#DIV/0!	0.0011
P00751	Complement factor B	38.01	15.97	85.5	0.0000±0.0000	2.0210±0.3990	#DIV/0!	0.0368
P27487	Dipeptidyl peptidase 4	52.19	14.62	88.2	0.0000±0.0000	2.2903±0.3483	#DIV/0!	0.0224
Q03001	Dystonin	26.42	9.1	860.1	0.0000±0.0000	46.0600±10.2000	#DIV/0!	0.0457
Q08380	Galectin-3-binding protein	89.82	7.69	65.3	0.0000±0.0000	5.6783±1.2243	#DIV/0!	0.0435
P06396	Gelsolin	133.95	16.62	85.6	0.0000±0.0000	6.1120±0.1580	#DIV/0!	0.0007
P00738	Haptoglobin	36.18	13.05	45.2	0.0000±0.0000	8.1780±1.8220	#DIV/0!	0.0462
P55795	Heterogeneous nuclear ribonucleoprotein H2	64.43	13.59	49.2	0.0000±0.0000	3.1643±0.5243	#DIV/0!	0.0264
P61978	Heterogeneous nuclear ribonucleoprotein K	96.35	10.15	50.9	0.0000±0.0000	2.2270±0.3550	#DIV/0!	0.0245
Q00839	Heterogeneous nuclear ribonucleoprotein U	52.18	16.85	90.5	0.0000±0.0000	6.3940±0.3620	#DIV/0!	0.0032
P02788	Lactotransferrin	49.98	3.66	78.1	0.0000±0.0000	4.0610±0.3320	#DIV/0!	0.0066
Q16787	Laminin subunit alpha-3	93.82	10.5	366.4	0.0000±0.0000	4.8750±0.5340	#DIV/0!	0.0118
P18669	Phosphoglycerate mutase 1	37.98	12.2	28.8	0.0000±0.0000	3.4140±0.7880	#DIV/0!	0.0494
Q13501	Sequestosome-1	51.39	24.32	47.7	0.0000±0.0000	10.4317±1.5483	#DIV/0!	0.0213
Q04837	Single-stranded DNA-binding protein, mitochondrial	103.04	10.14	17.2	0.0000±0.0000	3.5147±0.3177	#DIV/0!	0.0081
Proteins whose levels were significantly increased in tumor tissues								
P10809	60 kDa heat shock protein, mitochondrial	521.45	37.35	61	5.5643±0.6372	25.7167±2.4967	4.6217	0.0084
P63261	Actin, cytoplasmic 2	1937.49	39.2	41.8	373.2667±76.8624	664.0333±33.0333	1.779	0.0323
P07355	Annexin A2	448.13	41	38.6	3.2403±1.4795	18.5067±1.1533	5.7113	0.0081
P09525	Annexin A4	181.27	38.24	35.9	5.1360±2.5769	41.5633±3.2533	8.0925	0.0246
P02749	Beta-2-glycoprotein 1	49.1	8.7	38.3	0.1010±0.1010	5.2723±0.2297	52.2013	0.0015
Q10589	Bone marrow stromal antigen 2	34.42	5.56	19.8	0.2084±0.2084	3.9083±0.2683	18.757	0.0003
P27824	Calnexin	28.89	11.15	67.5	0.2428±0.2428	4.9467±0.7027	20.3706	0.0312
P07339	Cathepsin D	130.84	18.45	44.5	2.3293±1.4404	26.9267±0.4833	11.5598	0.0019
P01024	Complement C3	576.57	22.13	187	5.7683±0.6653	20.3667±0.3033	3.5308	0.0037
P0C0L4	Complement C4-A	113.8	14.68	192.7	3.6213±0.4757	15.8500±0.1800	4.3768	0.0024
P01031	Complement C5	56.12	9.49	188.2	2.6013±2.6013	24.7933±2.9467	9.531	0.0154
P04843	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	55.94	13.67	68.5	0.9681±0.2439	2.8047±0.1477	2.897	0.0426
P46977	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A	36.1	13.19	80.5	1.8350±1.8350	24.2800±0.9500	13.2316	0.005
Q6ZT12	E3 ubiquitin-protein ligase UBR3	49.64	13.98	212.3	3.3583±2.3183	27.5633±0.6933	8.2074	0.0137
P14625	Endoplasmin	121.44	28.27	92.4	1.1410±1.1410	8.6430±0.5600	7.5749	0.0378
P02671	Fibrinogen alpha chain	169.12	20.67	94.9	4.6783±0.1763	25.4200±0.4300	5.4336	0.0009
P02751	Fibronectin	115.69	8.76	262.5	0.4087±0.4087	6.8380±1.1180	16.7325	0.0425
P05062	Fructose-bisphosphate aldolase B	102.01	23.9	39.4	6.0663±0.6697	13.1733±0.8933	2.1715	0.0395
P06744	Glucose-6-phosphate isomerase	42.98	15.05	63.1	0.2564±0.2564	9.4153±0.4197	36.726	0.0016
P08238	Heat shock protein HSP 90-beta	138.42	9.94	83.2	1.8897±0.9472	6.9113±0.5527	3.6574	0.0062
P05534	HLA class I histocompatibility antigen, A-24 alpha chain	68.18	38.36	40.7	2.2293±1.0790	6.3590±0.3550	2.8525	0.0306
P01834	Ig kappa chain C region	1025.95	80.19	11.6	6.8100±1.0788	82.6333±9.0333	12.1341	0.017
P08729	Keratin, type II cytoskeletal 7	201.69	8.96	51.4	3.1607±0.3260	18.5467±0.0967	5.868	0.0004
P11047	Laminin subunit gamma-1	304.94	11.5	177.5	2.8467±0.9277	5.5903±0.4033	1.9638	0.0372
P22307	Non-specific lipid-transfer protein	73.26	18.28	59	3.7180±2.8827	40.1533±5.5333	10.7997	0.0355
Q15149	Plectin	216.53	20.71	531.5	6.0807±1.8479	18.5967±2.8667	3.0583	0.0076

Table 1. Continued

Accession no.	Protein name	MS/MS identification score	%Cov	MW (kDa)	Abundance level (×10⁶ arbitrary unit) (Mean±SEM)		Ratio (Cancer/ Normal)	P-value
					Normal	Cancer		
Proteins whose levels were significantly increased in tumor tissues								
P35232	Prohibitin	258.3	30.88	29.8	8.8697±1.2782	14.7000±2.1800	1.6573	0.0498
P05166	Propionyl-CoA carboxylase beta chain, mitochondrial	157.84	26.53	58.2	5.5760±1.5207	23.6567±1.4333	4.2426	0.0255
P07237	Protein disulfide-isomerase	135.02	15.35	57.1	14.4867±2.1006	20.5133±0.9333	1.416	0.0375
P13667	Protein disulfide-isomerase A4	385.57	19.84	72.9	3.8070±0.6932	5.7597±0.4577	1.5129	0.0197
P02768	Serum albumin	1384.23	45.65	69.3	51.8567±4.2376	151.0667±0.9333	2.9132	0.0023
Q9H299	SH3 domain-binding glutamic acid-rich-like protein 3	36.55	29.03	10.4	2.1060±1.0643	10.9747±1.7077	5.2111	0.0328
Q01082	Spectrin beta chain, non-erythrocytic 1	39.67	11.25	274.4	3.0400±0.9328	18.7433±0.1033	6.1656	0.0041
P51649	Succinate-semialdehyde dehydrogenase, mitochondrial	94.95	17.76	57.2	4.6537±0.3306	7.9290±0.0600	1.7038	0.0135
Proteins whose levels were significantly decreased in tumor tissues								
P42765	3-ketoacyl-CoA thiolase, mitochondrial	333.45	31.49	41.9	18.8333±3.7185	9.0440±2.4780	0.4802	0.0445
P24752	Acetyl-CoA acetyltransferase, mitochondrial	94.33	39.34	45.2	4.7250±0.6243	1.7400±0.4150	0.3683	0.0061
Q8N556	Actin filament-associated protein 1	30.14	10.14	80.7	31.0900±6.0975	8.7400±4.3700	0.2811	0.0386
P00505	Aspartate aminotransferase, mitochondrial	141.22	3.49	47.5	22.9167±2.2488	9.8260±0.9120	0.4288	0.0107
Q93088	Betaine--homocysteine S-methyltransferase 1	499.97	26.85	45	44.7300±2.9310	6.0227±2.0387	0.1346	0.0006
P04040	Catalase	327.97	33.21	59.7	52.9767±3.7150	22.1733±0.2333	0.4185	0.015
P02748	Complement component C9	61.77	10.38	63.1	8.4037±0.5427	5.2840±0.3720	0.6288	0.0325
P33260	Cytochrome P450 2C18	44.23	6.53	55.7	18.1367±2.3993	0.5090±0.5090	0.0281	0.0183
O75891	Cytosolic 10-formyltetrahydrofolate dehydrogenase	229.11	19.18	98.8	10.2643±1.4468	5.9247±0.5117	0.5772	0.0472
O94905	Erlin-2	163.33	34.51	37.8	11.9970±1.8881	2.4373±0.2067	0.2032	0.0411
P00367	Glutamate dehydrogenase 1, mitochondrial	662.8	23.48	61.4	80.7200±13.9885	30.4400±4.9300	0.3771	0.038
Q7Z4W1	L-xylulose reductase	251.82	24.59	25.9	50.8367±1.6628	7.6813±1.2693	0.1511	0.0034
P50897	Palmitoyl-protein thioesterase 1	71.99	15.03	34.2	4.8210±0.0751	2.2910±0.0970	0.4752	0.0045
P16083	Ribosyldihydronicotinamide dehydrogenase	74.47	25.11	25.9	12.8000±1.7481	0.6240±0.6240	0.0488	0.0092
Q16762	Thiosulfate sulfurtransferase	99.98	26.26	33.4	44.3633±6.0569	7.8657±0.3117	0.1773	0.0289
Proteins that were absent in tumor tissues								
Q68CK6	Acyl-coenzyme A synthetase ACSM2B, mitochondrial	47.88	12.82	64.2	4.7023±0.9060	0.0000±0.0000	0	0.0352
Q9BQE5	Apolipoprotein L2	64.32	3.26	37.1	0.7104±0.1493	0.0000±0.0000	0	0.0414
P00966	Argininosuccinate synthase	52.85	28.88	46.5	10.8433±1.6121	0.0000±0.0000	0	0.0214
P17174	Aspartate aminotransferase, cytoplasmic	103.6	10.17	46.2	11.7837±2.3640	0.0000±0.0000	0	0.038
P56385	ATP synthase subunit e, mitochondrial	76.42	18.84	7.9	6.3390±0.6072	0.0000±0.0000	0	0.0091
P0DN79	Cystathionine beta-synthase-like protein	72.75	8.71	60.5	4.0973±0.8011	0.0000±0.0000	0	0.0362
Q14749	Glycine N-methyltransferase	97	6.78	32.7	3.2080±0.3596	0.0000±0.0000	0	0.0123
P30047	GTP cyclohydrolase 1 feedback regulatory protein	92.55	30.95	9.7	21.4433±2.2466	0.0000±0.0000	0	0.0108
P48147	Prolyl endopeptidase	47.32	4.93	80.6	1.0373±0.0799	0.0000±0.0000	0	0.0059

%Cov, %Sequence coverage [(number of the matched residues/total number of residues in the entire sequence) x 100%] #DIV/0! = Divided by zero

biomarker was 81.3% and 67.4%, respectively. Regarding AFP with the current recommendation cut-off point of 20 ng/mL, its sensitivity and specificity was 52.5% and 67.4%, respectively. Among early HCC (BCLC stages 0 and A), 13 of 36 (36.1%) patients had elevated levels of serum AFP ≥20 ng/mL, whereas 26 (72.2%) patients had serum cathepsin D ≥184 ng/mL. For intermediate/ advanced HCC (BCLC stage B and C), the sensitivity of the corresponding biomarkers was 65.9% and 88.6%, respectively. When both serum AFP and cathepsin D were used together, the sensitivity of the combined assay was 83.3% for early HCC and 95.5% for intermediate/ advanced HCC.

Correlation of serum cathepsin D with clinical characteristics and survival

Serum cathepsin D level was positively correlated with serum aspartate aminotransferase (AST) (r=0.407, P<0.001), alanine aminotransferase (ALT) (r=0.258, P=0.021) and AFP (r = 0.323, P=0.003), but had negative correlation with albumin (r=-0.291, P=0.003). The biomarker also had a positive correlation with tumor size

(r = 0.357, P=0.001). Among different tumor stages, serum cathepsin D in patients with early HCC (BCLC 0 and A) (276.3 ± 166.7 ng/mL) was significantly lower than those of patients with BCLC stages B (373.0 ± 194.6 ng/mL) and patients with BCLC stages C (455.2± 192.4 ng/mL) (P = 0.036 and P=0.001, respectively). Serum cathepsin D level in patients BCLC stages B was also lower that in patients with BCLC stage C, although did not reach the statistical significance (P = 0.184).

The potential prognostic value of serum cathepsin D was also analyzed. Among patients with HCC, the median value of serum cathepsin D (305 ng/mL) was used as a cut-off level the compare the overall survival. Our data showed that the overall survival of patients with low levels (<305 ng/mL) was significantly better than that of patients whose serum levels were ≥305 ng/mL (P=0.010 by log rank test) (Figure 5).

Discussion

HCC is regarded as a complex multistep malignant tumor, which arises from a variety of environmental, host genetic and viral factors (Arzumanyan et al., 2013; Forner et al., 2018). Among these etiological factors, chronic HBV infection is the most important risk of HCC development, accounting for at least 50-60% of all cases among Thai and Southeast Asian populations. Thus, the identification of potential diagnostic and prognostic biomarkers for HBV-related HCC is highly needed in clinical practice. In the present study, we applied proteomics strategy to detect and quantitate differentially expressed proteins in various types of liver cancer, including HCC, iCCA and CRLM. In particular, we revealed several HBV-related HCC associated proteins that might be used as novel biomarkers for the diagnosis and prognosis. In the discovery cohort, our data showed that annexin A2, cathepsin D, calnexin, plectin and prohibitin were significantly elevated in HCC tissues. We further confirmed their expression levels in liver tissues using Western blot analysis. However, only annexin A2 and cathepsin D were significantly up-regulated in HCC issues in comparison with adjacent non-tumor tissues. Moreover, these proteins were not found to be any difference in pair tumor and non-tumor tissues of iCCA and CRLM. Thus, measurement of annexin A2 and cathepsin D might represent specific tissue biomarkers for HCC in differentiating from iCCA and CRLM.

In the validated cohort using serum specimens, we observed that circulating annexin A2 level in patients with HCC did not differ from those found in patients with LC and healthy controls, indicating that this protein might not be a reliable biomarker of HCC. Annexin A2, a 36-kDa binding protein belonged to the annexin superfamily, is localized to the extracellular surface of endothelial cells, as well as in several types of tumor cells (Gerke et al., 2002). Indeed, annexin A2 was found to promote various processes related to HCC progression and their resistance to anticancer therapy(Qiu et al., 2020). Additionally, recent data showed that annexin A2 was involved in immune escape of HCC by inducing regulatory T cells and downregulating natural killer cells and dendritic cells(Qiu et al., 2020). Previous data also showed that annexin A2 was up-regulated in HCC tissues compared with non-cancerous tissues and its expression was correlated with aggressive tumor characteristics including tumor differentiation, portal vein thrombus and metastasis(Zhang et al., 2015). Moreover, annexin A2 was shown as an independent serological marker for detecting an early stage of HBV-related HCC(Sun et al., 2013). In contrast, some reports demonstrated that serum annexin A2 was not a good diagnostic and prognostic biomarker for HBV-related HCC, particularly among patients with underlying cirrhosis (Liu et al., 2013). Additionally, a recent report indicated that the prognostic value of annexin A2 was diminished in several clinical subgroups, including advanced HCC or severe impaired liver function(Huang et al., 2021). These inconsistent results in terms of diagnostic and prognostic value of annexin A2 are not known but might be related to the heterogeneity of studied populations, as well as the different in detection methods and distinct functions of intracellular and extracellular of annexin A2 proteins (Hitchcock et al., 2014).

Cathepsin D is a soluble lysosomal aspartic endopeptidase that has main functions involved in protein degradation and precursor activation in normal cells (Masson et al., 2010). It has been shown that cathepsin D is over-expressed and secreted by various human tumor cell types, including breast, ovarian, prostate and bladder cancers (Nicotra et al., 2010; Benes et al., 2008). Previous studies have demonstrated that cathepsin D plays a crucial role in promoting tumorigenesis including tumor growth, angiogenesis, apoptosis and metastatic properties (Masson et al., 2010). In breast cancer, for instance, a recent meta-analysis has revealed that high expression levels of cathepsin D are associated with poor prognosis and this biomarker could also be served as a potential target of anticancer therapy (Kang et al., 2020). Currently, the role of Cathepsin D in chronic liver disease has not been sufficiently explored. Among patients with nonalcoholic fatty liver disease (NAFLD), circulating cathepsin D level was shown to be useful for the assessment of disease severity in some studies (Walenbergh et al., 2015; Walenbergh et al., 2016), although other recent data did not confirm this observation (Kamarajah et al., 2019). In patients with HCC, it was demonstrated that immunoreactivity scores of cathepsin D was significantly elevated in HCC tissue compared with surrounding non-cancerous liver specimens(Lin et al., 2016), which was in line with the findings of our report.

Our data also indicated that serum cathepsin D level was significantly higher in patients with HCC than patients with LC and healthy controls. Additionally, it was shown from the ROC curve analysis that serum cathepsin D could satisfactorily differentiate HCC from LC. In fact, serum cathepsin D might represent a better biomarker for detecting small HCC in comparison with AFP. Currently, serum AFP is regarded as the most commonly-used biomarker for HCC. However, this fetal glycoprotein has some restrictions in clinical use, particularly a relative low sensitivity in early-stage HCC. In our study, combined serum cathepsin D and AFP had the sensitivity of 80% for the diagnosis of early HCC and approached 96% for

detecting intermediate and advanced stages. Thus, the combined usage of AFP and cathepsin D might be more beneficial than using AFP alone for identifying early HCC. However, further studies that include a larger number of patients are necessary to confirm this observation.

Our data also revealed that there was a positive correlation between increased serum cathepsin D and aggressive characteristics of HCC. In particular, enhanced serum cathepsin D level was detected more frequently in patients with larger tumor size and advanced BCLC stages. Indeed, the level of this biomarkers was steadily elevated in patients with more advanced tumor stages in comparison with early HCC. These findings might suggest that cathepsin D involves in the development and progression of HCC, similarly to the observations found in other cancers. Moreover, the Kaplan-Meier analysis indicated that overall survival of patients with high serum level was significantly inferior to that of the low expression group, supporting that serum cathepsin D could be unfavorable predictor of overall survival. To the best of our understanding, this study is the first to demonstrate the diagnostic and prognostic importance of serum cathepsin D in patients with HCC.

In summary, our report showed that tissue-based proteomic profiles in HCC substantially differed from those observed in iCCA and CRLM. Among the selected candidate proteins, cathepsin D was shown to be highly expressed in HCC tissues compared with adjacent non-tumor tissues as assess by Western blotting. Moreover, circulating levels of cathepsin D were significantly elevated in patients with HCC compared to patients with LC and healthy controls. The combined assay of cathepsin D and AFP might also provide a promising tool to better differentiate early HCC from LC. Apart from its diagnostic role, serum cathepsin D might also serve as a prognostic biomarker of HBV-related HCC. Additional large-scale studies are needed to verify our findings and to explore whether this biomarker has a clinical utility in other etiological causes of HCC.

Author Contribution Statement

PS and PT design initiated the study idea and design of the study, NC, SC and VT performed experiment, interpretation, data analysis and visualization. SS, NP and AV collected and provided the clinical data. PS and AV collected and provided specimen. PT and NC drafted manuscript. PT and PS provided revision of scientific content to manuscript. NC re-write the manuscript. PS and PT provided funding. All authors read and approved this manuscript.

Acknowledgements

For Western blot and ELISA assays, we also would like to thank Wirin Bowonsomsarit, Rattanaporn Kiatbumrung and Pornjira Somnark from Center of Excellence in Hepatitis and Liver Cancer, Faculty of Medicine, Chulalongkorn University.

Funding Statement

This study was supported by Ratchadapisek Sompoch Endowment Fund (RA65/008), Faculty of Medicine, Chulalongkorn University. The study was supported by The Thailand Research Fund (TRF) Senior Research Scholar (RTA6280004) and Center of Excellence in Hepatitis and Liver Cancer, Faculty of Medicine, Chulalongkorn University. VT is supported by Mahidol University research grant

Ethical Declaration

The Institute Ethics Committee of Faculty of Medicine, Chulalongkorn University approved the research.

Conflict of Interest

All authors have no conflict of interest.

References

Arzumanyan A, Reis HM, and Feitelson MA (2013). Pathogenic mechanisms in HBV- and HCV-associated hepatocellular carcinoma. *Nat Rev Cancer*, **13**, 123-35.

Benes P, Vetvicka V, Fusek M (2008). Cathepsin D--many functions of one aspartic protease. *Crit Rev Oncol Hematol*, **68**, 12-28.

Beretta L (2009). Comparative analysis of the liver and plasma proteomes as a novel and powerful strategy for hepatocellular carcinoma biomarker discovery. *Cancer Lett*, **286**, 134-9.

Bray F, Ferlay J, Soerjomataram I, et al (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, **68**, 394-424.

Forner A, Reig M, Bruix J (2018). Hepatocellular carcinoma. *Lancet*, **391**, 1301-14.

Gerke V, Moss SE (2002). Annexins: from structure to function. *Physiol Rev*, **82**, 331-71.

Heimbach JK, Kulik LM, Finn RS, et al (2018). AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology*, **67**, 358-80.

Hitchcock JK, Katz AA, Schafer G (2014). Dynamic reciprocity: the role of annexin A2 in tissue integrity. *J Cell Commun Signal*, **8**, 125-33.

Huang SW, Chen YC, Lin YH, Yeh CT (2021). Clinical Limitations of Tissue Annexin A2 Level as a Predictor of Postoperative Overall Survival in Patients with Hepatocellular Carcinoma. *J Clin Med*, **10**.

Kamarajah SK, Khoo S, Chan WK, et al (2019). Limited applicability of cathepsin D for the diagnosis and monitoring of non-alcoholic steatohepatitis. *JGH Open*, **3**, 417-24.

Kang J, Yu Y, Jeong S, et al (2020). Prognostic role of high cathepsin D expression in breast cancer: a systematic review and meta-analysis. *Ther Adv Med Oncol*, **12**, 1758835920927838.

Lin YY, Chen ZW, Lin ZP, et al (2016). Tissue Levels of Stefin A and Stefin B in Hepatocellular Carcinoma. *Anat Rec (Hoboken)*, **299**, 428-38.

Liu Z, Ling Q, Wang J, et al (2013). Annexin A2 is not a good biomarker for hepatocellular carcinoma in cirrhosis. *Oncol Lett*, **6**, 125-9.

Masson O, Bach AS, Derocq D, et al (2010). Pathophysiological functions of cathepsin D: Targeting its catalytic activity versus its protein binding activity?. *Biochimie*, **92**, 1635-43.

Mathema VB, Na-Bangchang K (2015). Current insights on cholangiocarcinoma research: a brief review. *Asian Pac J Cancer Prev*, **16**,1307-13.

Nicotra G, Castino R, Follo C, et al (2010). The dilemma: does tissue expression of cathepsin D reflect tumor malignancy? The question: does the assay truly mirror cathepsin D misfunction in the tumor?. *Cancer Biomark*, **7**, 47-64.
Qiu LW, Liu YF, Cao XQ, et al (2020). Annexin A2 promotion of hepatocellular carcinoma tumorigenesis via the immune microenvironment. *World J Gastroenterol*, **26**, 2126-37.
Sauzay C, Petit A, Bourgeois AM, et al (2016). Alpha-foetoprotein (AFP): A multi-purpose marker in hepatocellular carcinoma. *Clin Chim Acta*, **463**, 39-44.
Sia D, Villanueva A, Friedman SL, Llovet JM (2017). Liver Cancer Cell of Origin, Molecular Class, and Effects on Patient Prognosis. *Gastroenterology*, **152**, 745-61.
Singh S, Muir AJ, Dieterich DT, Falck-Ytter YT (2017). American Gastroenterological Association Institute Technical Review on the Role of Elastography in Chronic Liver Diseases. *Gastroenterology*, **152**, 1544-77.
Sun Y, Gao G, Cai J, et al (2013). Annexin A2 is a discriminative serological candidate in early hepatocellular carcinoma, *Carcinogenesis*, **34**, 595-604.
Thanomkitti K, Kanlaya R, Fong-Ngern K, et al (2018). Differential proteomics of lesional vs. non-lesional biopsies revealed non-immune mechanisms of alopecia areata. *Sci Rep*, **8**, 521. doi:10.1038/s41598-017-18282-1.
Thongboonkerd V, Chaiyarit S (2022). Gel-Based and Gel-Free Phosphoproteomics to Measure and Characterize Mitochondrial Phosphoproteins. *Curr Protoc*, **3**, e390. doi:10.1002/cpz1.390.
Walenbergh SM, Houben T, Hendrikx T, et al (2015). Plasma cathepsin D levels: a novel tool to predict pediatric hepatic inflammation. *Am J Gastroenterol*, **110**, 462-70.
Walenbergh SM, Houben T, Rensen SS, et al (2016). Plasma cathepsin D correlates with histological classifications of fatty liver disease in adults and responds to intervention, *Sci Rep*, **6**, 38278.
Xing X, Liang D, Huang Y, et al (2016). The application of proteomics in different aspects of hepatocellular carcinoma research. *J Proteomics*, **145**, 70-80.
Zhang H, Yao M, Wu W, et al (2015). Up-regulation of annexin A2 expression predicates advanced clinicopathological features and poor prognosis in hepatocellular carcinoma, *Tumour Biol*, **36**, 9373-83.