A proteomics-based approach for the identification of biliary markers of Cholangiocarcinoma

Shahid Farid

BSc (Hons) Med Sc., MBChB, MRCS

Submitted in accordance with the requirements for the degree of

Doctor of Medicine

The University of Leeds School of Medicine

June 2014

Acknowledgements

I would like to thank my supervisors Mr KR Prasad, Professors Rosamonde Banks and Peter Selby for their continued support and guidance.

Special thanks are given to Dr Rachel Craven and Rumana Rafiq for advice, teaching and expert supervision of experiments throughout the study.

Dr Jianhe Peng and Dr Alexander Zougman are thanked for their provision of mass spectrometry expertise. Dr David Cairns and Dr David Jackson are acknowledged for their valuable input to data and statistical analysis.

Personal thanks to all other members of Lab 3 who have shared their experience in facilitating efficient acquisition of laboratory skills and above all making the challenge enjoyable.

The Royal College of Surgeons of England are acknowledged for a Research Fellowship award in 2010 and providing professional support throughout the period of study.

Abstract

Prognosis for patients with cholangiocarcinoma (CCA) continues to be poor as a result of the difficulty in distinguishing malignant from benign bile duct disease, late stage diagnosis and a lack of sufficiently sensitive and specific diagnostic markers. These factors underlie the pressing clinical need for novel disease biomarkers.

The utility of bile as a proximal fluid for biomarker discovery (compared to serum) was investigated using two dimensional difference gel electrophoresis (2D DIGE). Significant differences between the proteomes of bile and serum were identified, supporting the hypothesis that bile offers a potentially enriched microenvironment of proteins shed/secreted by tumour. However, as with serum, a few major abundant proteins dominate the bile proteome, therefore an albumin/IgG depletion technique was optimised to improve biomarker identification.

The bile proteome was initially characterised in samples from patients with hilar CCA. A protein mastermap was generated by two dimensional polyacrylamide gel electrophoresis (2D PAGE) and a catalogue of proteins by liquid chromatography – tandem mass spectrometry (LC-MS/MS), which identified 80 and 813 unique proteins respectively. This represents one of the largest compendiums to date and forms a basis for future proteomic-based biomarker studies.

A comparative analysis of biliary proteins in CCA and benign biliary disease was performed using a label-free proteomic approach to identify potential diagnostic biomarker(s). Comparative analysis of bile protein profile in 5 patients with CCA versus benign biliary disease identified 13 proteins which were at higher levels in malignant disease of which metalloproteinase-9 (MMP-9), Rho GDP-dissociation inhibitor 2 (also known as Ly-GDI), Annexin A3 and pre-B-cell colony-enhancing factor (PBEF) were taken forward in immunoblotting-based validation. MMP-9 was shown to be overexpressed in bile of CCA and represents a potential diagnostic marker. In addition analysis of bile samples showed lipocalin-2 and its complex with MMP-9 were present in greater amounts in CCA compared to benign biliary disease.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Publications arising from this thesis

Shotgun proteomics of human bile in hilar cholangiocarcinoma.

Farid SG, Craven RA, Peng J, Bonney GK, Perkins DN, Selby PJ, Rajendra Prasad

K, Banks RE. Proteomics. 2011 May;11(10):2134-8. PMID: 21500345

' Omics-based' technology applications in Hepatopancreaticobiliary Malignancy.

Farid SG, Dennison AR. ECAB Clinical Update: Surgical Gastroenterology and Liver

Transplantation. 2012. Book Chapter.

Presentations arising from this thesis

A Proteomic Analysis of Human Bile in Patients with Bile Duct Cancer. Shahid G.

Farid, Rachel A. Craven, Jianhe Peng, Glenn K. Bonney, David N. Perkins, Peter J.

Selby, K. Rajendra Prasad, Rosamonde E. Banks. 7th Joint British Society of

Proteomic Research/ European Bioinformatics Institute (BSPR/EBI) Meeting.

Hinxton, Cambridge,UK. July 2010

Shotgun Proteomics of Human Bile in Hilar Cholangiocarcinoma. Shahid G. Farid,

Rachel A. Craven, Jianhe Peng, David N. Perkins, Peter J. Selby, K. Rajendra

Prasad, Rosamonde E. Banks. Society of Academic and Research Surgery Annual

Conference. Dublin, Ireland. January 2011

Identification of Differentially Expressed Proteins in Hilar Cholangiocarcinoma using

a Label Free Proteomic Approach. Shahid G. Farid, Rachel A. Craven, Alexander

Zougman, David Cairns, Peter J. Selby, K. Rajendra Prasad, Rosamonde E. Banks.

Society of Academic and Research Surgery Annual Conference. Dublin, Ireland.

January 2011

1.0 Introduction

1.1.0 Cholangiocarcinoma

Cholangiocarcinoma (CCA) results from the neoplastic transformation of cholangiocytes lining the biliary tree, causing biliary strictures and obstruction and represents 10-15% of all hepatobiliary and 3% of all gastrointestinal malignancies [1], [2]. CCA remains a devastating disease, limited by effective therapies and universal poor prognosis with median survival of 6-12 months without treatment [3].

1.1.1 Classification and clinical presentation

CCA can be classified anatomically into cancers originating in the intrahepatic bile ducts or the extrahepatic bile duct within the hepatoduodenal ligament. The latter is further divided into proximal, middle and distal CCA depending on location within the extrahepatic biliary system. More recent literature has abbreviated extrahepatic CCA into proximal and distal, as middle duct tumours are clinically rare (Figure 1A) [4]. Proximal CCA are also known as hilar CCA or Klatskin tumours and represent the majority of cases (60-70%) followed by distal (20-30%) and intrahepatic CCA (5-10%) [5],[6]. The Bismuth-Corlette classification of hilar CCA describes tumour location and its spread within the biliary tree in more detail and is utilised in the surgical management algorithm (Figure 1B) [7]. In addition to this traditional anatomic classification, a sub-classification based on macroscopic growth identifies tumours as mass-forming, periductal infiltrating or intraductal growing [8]. In intrahepatic disease, tumours develop as solid masses, infiltrate periductal tissue and/or grow within the duct. In contrast

extrahepatic CCAs develop sclerosing strictures (most common), nodular lesions, or more rare papillary growth patterns [9]. Preoperative recognition of such variation can provide information on potential resectability and prognosis. Histologically more than 95% of CCA are adenocarcinomas, and the remainder of squamous cell origin [10].

Although intra and extrahepatic CCA share similar features including being clinically silent and exhibiting non-specific complaints in early stage disease, they have distinct epidemiological, clinical, pathological, and therapeutic features. In intrahepatic CCA, radiological imaging may identify a hepatic mass, often incidentally, which is difficult to distinguish from secondary metastatic disease. Painless progressive jaundice and deranged liver function tests are a presenting feature of extrahepatic CCA. CCA results in a progressive clinical state of biliary sepsis, liver failure and malnutrition which in combination are the leading causes of death.
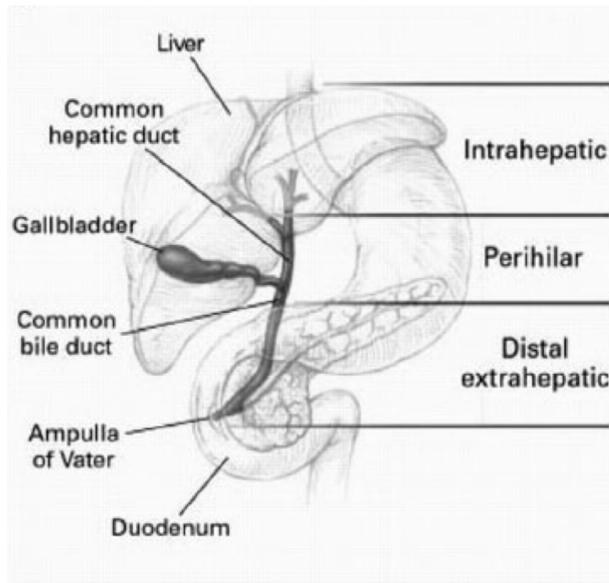
1A
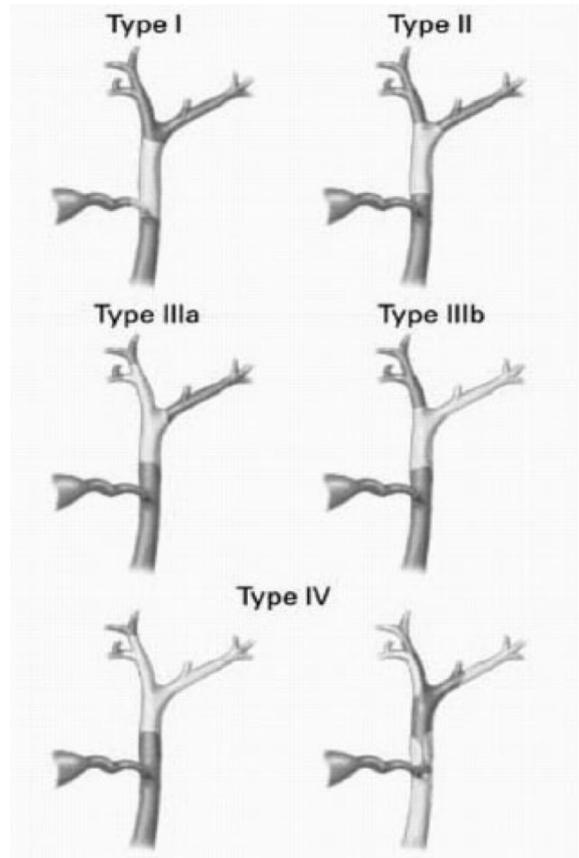
1B

Figure 1. Classification of CCA

(A) <u>Anatomical classification of CCA</u>.

The intrahepatic bile ducts unite draining the right and left parts of the liver to form the common hepatic duct which joins the cystic duct from the gallbladder and is directed to its opening in the duodenum. Intrahepatic CCA by definition originates from the second (segmental) or peripheral branch of the bile duct. Perihilar CCA is located in the extrahepatic biliary tree proximal to the origin of the cystic duct. Distal CCA involves the common bile duct.

(B) <u>Bismuth – Corlette classification of perihilar CCA</u>.

Type I: Limited to the common bile duct; >2cm from the confluence of the right and left hepatic ducts

Type II: <2cm. from the confluence and +/- involving the confluence

Type IIIa: Type II + right hepatic duct involvement

Type IIIb: Type II + left hepatic duct involvement

Type IV: extending to both right and left hepatic ducts or multifocal involvement.

1.1.2 Epidemiology

CCA is the second most common primary liver tumour after hepatocellular carcinoma (HCC) and is now the main cause of death from a primary liver tumour [11]. Significant variation exists in the worldwide incidence and prevalence of CCA with the highest proportional incidence in north-east Thailand (96 per 100,000 men), 1-2 per 100,000 in the UK and USA and the lowest rates observed in Australia [12]. It has a male to female ratio of 1.5 and individuals present most commonly in their seventh decade [12].

Several epidemiological studies have shown a significant rise in the incidence and mortality of intrahepatic CCA in the western world [13]. From 1968 to 2001, the age-standardized mortality rate (ASMR per 100,000 population) for intrahepatic CCA increased from 0.10 to 1.49 in males and 0.05 to 1.24 in females and the annual number of deaths increased 30-fold, from 36 in 1968 to 1003 in 2004 in the UK [14]. Data from the World Health Organisation and others also demonstrate an almost universal increase in ASMR for intrahepatic CCA in both sexes across US, Europe and Australasia between 1979 – 1998 [15-16]. In the US this observation is seen to be highest in black followed by white men, white and black women respectively [17].

The cause of increase in intrahepatic CCA remains unclear. Even with improvements in diagnostic tools, correction for previous errors in misclassification of hilar tumours as intrahepatic and with up to 40% of primary liver tumours as adenocarcinoma instead of more specifically CCA or HCC, the phenomenon is considered genuine [2],[18-19]. Data from the Far East and the US has implicated the increasing prevalence of hepatitis B & C related liver cirrhosis and cirrhosis in general to explain in part the rising incidence [20-21].Evidence of environmental carcinogens particularly related to rural/ agricultural areas have been considered to account for some of the geographical variation in the

UK but specific agents have yet to be characterised [14]. Proquinazid (6-iodo-2-propoxy-3-propyl-3*H*-quinazolin-4-one) is a new fungicide intended for use in agriculture and control of powdery mildew in cereals and grapes has been noted to carry a mutagenic risk and higher CCA incidence in rats (http://www.iacom.org.uk/ statements /COM05S4.htm, accessed 12/07/12).

The trend for extrahepatic CCA is in marked contrast, with incidence and mortality rates being stable or in decline [12]. The ASMR has been shown to decrease from 0.6 to 0.3 over twenty years from 1979 to 1998 in the US and 0.7-0.8 to 0.23 in the UK [15], [22].

1.1.3 Risk Factors

Most CCA arise in the absence of any known predisposing factors [23]. In the western world primary sclerosing cholangitis (PSC) confers the highest risk of developing CCA, with an annual risk between 0.6-1.5% and lifetime risk of 20% [24]. PSC is an autoimmune disease causing progressive bile duct inflammation and strictures leading to chronic cholestatic liver disease and the requirement for regular surveillance for transformation of malignant biliary strictures [25-26]. However up to 36% of explanted livers from PSC patients undergoing liver transplantation exhibit occult pathological features of CCA and highlight the current limitations in surveillance strategies [27]. The majority of PSC patients who subsequently develop CCA will have done so within 2.5 years of diagnosis and overall individuals will present at a younger age (30-50 years) compared to the general population [28-29]. Additional risk of CCA in PSC is conferred by concomitant PSC-associated inflammatory bowel disease, advanced age at time of PSC diagnosis, previous colonic carcinoma, history of smoking and alcohol consumption >80 g a day [24],[30].

A number of other risk factors for CCA exist and these are summarised in Table 1.0. The exact mechanism in each case is not fully understood but in principle they revolve around the creation of a chronic inflammatory environment, release of inflammatory signalling molecules, cell stress and subsequent DNA damage and increased malignant potential [31-33]. Factors exhibiting such features include viral (Hepatitis B & C), and parasitic (Opisthorchis viverrini, Clonorchis sinensis) infections and account for increased risk of CCA in endemic areas such as the Far East [34-36].

Obesity and type II diabetes are considered to confer increased risk of CCA [37]. Since both conditions are linked with insulin resistance it is of interest to note that investigators have also shown polymorphism of selected genes relating to insulin sensitivity and elevated risk of biliary cancer in a population-based case-control study in Shanghai, China [38].

In hepatolithiasis, bile stasis and recurrent secondary infection result in inflammatory bile duct changes, promoting further stone formation, chronic inflammation and CCA risk. Similarly in 5-15% of patients with the congenital anomaly of choledochal cyst and its associated pancreaticobiliary mal-junction, bile stasis and pancreatic fluid reflux activate bile acids, resulting in chronic inflammation and a "hyperplasia – carcinoma sequence" [39]. Thorotrast containing thorium dioxide was used as a common radiological contrast agent until 1950-60 and has since then been implicated in several cancers including biliary, pancreatic, gastric, hematological and renal [40-42].

Table 1.0 Known risk Factors for CCA.

Risk Factor
General
Obesity
Type II Diabetes
Infection
Viral Hepatitis B
Viral Hepatitis C
Opisthorchis viverrini
Clonorchis sinensis
Inflammation
Primary sclerosing cholangitis (PSC)
Hepatolithiasis
Liver cirrhosis
Genetic - Polymorphisms
cytochrome P450 1A2,
arylamine N-acetyltransferase 2
Environmental
Organic solvents (aromatic, alicyclic & chlorinated hydrocarbons)
Thorotrast
Others
Choledochal cysts
Caroli's syndrome/congenital hepatic fibrosis
Bilio-enteric anastomosis

1.1.4 Staging of CCA

Accurate staging can direct optimal therapeutic strategies and allow for accurate prognostication. The staging for intra and extrahepatic CCA are different and several systems have been proposed. The Tumour Node Metastasis (TNM) staging is a common system applied to cancer proposed by the International Union against Cancer (IUCC) and the American Joint Committee on Cancer (AJCC) [43] and is currently used internationally in CCA. It collectively describes the tumour and its degree of lateral extension, related lymph node involvement and spread to distant sites. The results are combined to formulate a clinical stage from 0 – IV. The TNM classification and stage for intra and extrahepatic CCA is shown in Tables 1.1 & 1.2 respectively. Node and metastasis status classification are identical in both intra and extrahepatic CCA.

The Memorial Sloane Kettering Cancer Centre (MSKCC) has advocated the T-stage criteria for hilar CCA and corresponds to what can be assessed by evaluating the longitudinal and lateral spread pattern based on the Bismuth-Corlette classification and TNM staging system but irrespective of lymph node or metastasis status (Table 1.3) [44]. A more detailed staging system for extrahepatic CCA has been used by the Japanese Society of Biliary Surgery (JSBS) since 1981 (Table 1.4) with the final score calculated by including surgical and pathological findings [45].

However the late presentation of CCA, sensitivity of preoperative radiological imaging and challenges in practical application of current systems has resulted in no single system being ideal in determining resectability and predicting survival. Resection rates show considerable variation in range of between 25-85% in surgical centers and unresectable disease is often only confirmed at laparotomy [46]. Both the AJCCS/UICC and JSBS classifications are formed with significant data reliant upon histopathological

features and limit preoperative applicability and potential postoperative adjunct therapies. The Bismuth – Corlette system (described in Figure 1.2) stratifies patients based on the extent of duct involvement and while it can facilitate surgical planning has not been shown to be indicative of survival [47]. To date no biomarkers exist or have been evaluated in integration with current staging systems to improve accuracy and the focus still remains on the refinement of existing clinical and pathological based systems.

Table 1.1 TNM staging of intrahepatic CCA.

Primary tumour (T)

TX	The primary tumour cannot be assessed
T0	There is no evidence of a primary tumour
T1	The tumour is only a single tumour and does not involve adjacent blood vessels
T2	Either of these conditions: Any tumour that involves adjacent blood vessels is present Multiple tumours, none larger than 5 cm are present
T3	Either of these conditions: More than one tumour larger than 5 cm is present The tumour involves the major veins within the liver
T4	Either of these conditions: The tumour has spread to the organs near the liver (except gallbladder) The tumour is present with perforation of the visceral peritoneum

Regional lymph nodes (N)

NX	The regional lymph nodes cannot be assessed
N0	No regional lymph node involvement
N1	Regional lymph node involvement

Distant metastasis (M)

MX	The tumour cannot be assessed
M0	No distant metastases
M1	Distant metastases

Table 1.1 TNM staging of intrahepatic CCA (continued)

Stage grouping

Stage I	T1	N0	M0
Stage II	T2	N0	M0
Stage IIIA	T3	N0	M0
Stage IIIB	T4	N0	M0
Stage IIIB	Any T	N1	M0
Stage IV	Any T	Any N	M1

Table 1.2 TNM staging of extrahepatic CCA (AJCC 6th Edition).

Primary tumour (T)

TX	The primary tumour cannot be assessed
T0	There is no evidence of a primary tumour
Tis	Carcinoma in situ
T1	The tumour is confined to the bile duct
T2	The tumour has spread beyond the wall of the bile duct
T3	The tumour has spread to the liver, gallbladder, pancreas, and/or an unilateral branch of the veins and/or arteries within the liver
T4	The tumour has spread bilaterally (both sides) to the veins or arteries within the liver and/or adjacent structures, such as the colon, stomach, duodenum, or abdominal wall

Table 1.2 TNM staging of extrahepatic CCA (continued)

Stage grouping			
Stage 0	Tis	N0	M0
Stage IA	T1	N0	M0
Stage IB	T2	N0	M0
Stage IIA	T3	N0	M0
Stage IIB	T1	N1	M0
	T2	N1	M0
	T3	N1	M0
Stage III	T4	Any N	M0
Stage IV	Any T	Any N	M1

Table 1.3 T Stage criteria (MSSK) for hilar CCA.

Stage grouping	
T1	Tumour confined to confluence and/or right or left hepatic duct without portal vein involvement or liver atrophy
T2	Tumour confined to confluence and/or right or left hepatic duct with ipsilateral liver atrophy. No portal vein involvement demonstrated
T3	Tumour confined to confluence and/or right or left hepatic duct with ipsilateral portal venous branch involvement with/without associated ipsilateral lobar liver atrophy. No main portal vein involvement (occlusion, invasion or encasement)
T4	Any of the following: i) Tumour involving both right and left hepatic ducts up to secondary radicles bilaterally ii) Main portal vein encasement

Table 1.4 JSBS (2nd Edition) staging classification of extrahepatic CCA.

Primary tumour (T)

T1	S_0	$Hinf_0$	$Panc_0$	PV_0	A_0
T2	S_1	$Hinf_1$	$Panc_1$	PV_0	A_0
T3	$S_{2,3}$	$Hinf_1$	$Panc_2$	PV_0	A_0
T4	Any	$Hinf_{2,3}$	$Pancc_{2,3}$	$PV_{1,2,3}$	$A_{1,2,3}$

Key:

(S) Serosa found in a part of the anterior and right posterior wall of portal, superior and middle bile duct

S0 No invasion of the serosa

S1 Doubtful invasion of the serosa and/or serosal surface

S2 Definite invasion of the serosa

S3 Invasion of other organs or structures: the stomach, abdominal wall, colon, and inferior vena cava

(H) Hepatic

$Hinf_0$ No direct invasion of the liver

$Hinf_1$ Doubtful direct invasion of the liver

$Hinf_2$ Definite direct invasion of portal bile ducts

$Hinf_3$ Definite direct invasion beyond portal bile ducts

(Panc) Pancreas

Panc 0 No invasion of the pancreas

Panc 1 Doubtful invasion of the pancreas

Panc 2 Definite invasion of bile duct around the pancreas

Panc 3 Severe invasion of the pancreas

(DU) Duodenum

DU 0 No invasion of the duodenum

DU1 Doubtful invasion of the duodenum

DU 2 invasion of duodenum around bile duct

DU 3 Definite invasion of the duodenum

Key (continued): (PV) Portal vein

PV 0 No invasion of the portal vein

PV 1 Doubtful invasion of the portal vein

PV 2 Definite invasion of the portal vein

PV 3 Severe invasion of portal veins (narrowing or obstruction)

(HA) Hepatic artery

HA 0 No invasion of the hepatic artery

HA 1 Doubtful invasion of the hepatic artery

HA 2 Definite invasion of the hepatic artery

HA 3 Severe invasion of the hepatic artery (narrowing or obstruction)

Lymph Node - JSBS classification of extrahepatic CCA

N0	No evidence of lymph node metastasis
N1	Metastasis to group 1 lymph nodes, but no metastasis to group 2 and 3 lymph nodes
N2	Metastasis to group 2 lymph nodes, but no metastasis to group 3 lymph nodes
N3	Metastasis to group 3 lymph nodes

Metastasis - JSBS classification of extrahepatic CCA.

Liver metastasis	
H0	No evidence of liver metastasis
H1	Metastasis limited to one lobe
H1(r)	Metastasis limited to the right lobe
H1(l)	Metastasis limited to the left lobe
H2	A few metastases to both lobes
H3	Numerous metastases to both lobes
Peritoneal metastasis	
P0	No evidence of peritoneal metastasis
P1	Metastasis to the peritoneum adjacent to extrahepatic bile ducts
P2	A few metastases to the distant peritoneum
P3	Numerous metastases to the distant peritoneum
Distant metastasis	
M(-)	No evidence of distant metastasis other than peritoneal and/or liver metastases
M(+)	Distant metastases other than peritoneal and/or liver metastases

JSBS stage classification of extrahepatic CCA.

	H0, P0, M (-)				H1,2,3, P1,2,3, M (+)
	N0	N1	N2	N3	
T1	I	II		Iva	
T2	II	III			IVb
T3			Iva		
T4	Iva				

1.1.5 Molecular pathogenesis of CCA

Exact molecular mechanisms remain to be clearly defined in the multi-modal development and growth of CCA but remain important in directing studies focused on diagnostics and therapeutics. Nevertheless inflammation and its mediators in response to biliary duct epithelium damage, bile flow dysfunction, increased cholangiocyte turnover and accumulation of DNA damage ultimately leading to malignant transformation are important common features [31]. In addition aberrations in cell cycle regulation mediators and factors implicated in malignant transformation including enhanced proliferative signalling, evasion of apoptosis, angiogenesis, invasion and metastasis have all been described.

The release of key pro-inflammatory cytokines such as interleukin-6 (IL-6) and tumour necrosis factor alpha (TNFα) inducing activation of reactive nitrogen species (RNS) such as nitric oxide (NO), reactive oxygen species (ROS) and increased malignant potential are well cited in human cancers [49],[50] including CCA [31],[51-52]. During this process cholangiocytes secrete mitogens that activate local cellular receptors and several intracellular pathways resulting in prolonged cellular stress, increased cell turnover, accumulation of DNA damage and failure of DNA repair.

Overproduction of NO by inducible nitric oxide synthase (iNOS) and ROS results in DNA damage recognisable by several markers including the generation of 8-nitroguanosine and 8-oxo-7,8-dihydro-2-deoxyguanosine (8-oxodG), deamination of guanine and guanine to tyrosine transversions [53]. Overexpression of both 8-oxodG and 8-nitroguanosine have been demonstrated in the biliary epithelium of intrahepatic CCA and correlated to known adverse prognostic factors of perineural and lymphatic invasion [54].

Tumour growth can lead to surrounding tissue hypoxia and is associated with up regulation of hypoxia inducible factor (HIF)α subunits, which can mediate transcription of various genes including iNOS [55]. The over expression and co-localisation of HIF1α and iNOS in CCA tissue has raised interest in a reciprocal link [54]. In this hypothesis model, NO accumulation results in DNA damage and HIF1α over expression. Subsequent tumour growth-related oxidative stress leads to a positive feedback loop resulting in further HIF1α and iNOS activation.

NO can activate cyclooxygenase-2 (COX-2), a molecule implicated in a number of gastrointestinal cancers [56]. COX-2 over expression has been demonstrated in CCA cell lines exposed to bile acids and subsequent immunohistochemical studies comparing CCA and normal tissue showed a stepwise increase in levels as cells progressed from inflammation to dysplasia to carcinoma [19],[57-58]. Its role in cholangiocarcinogenesis is thought to involve Fas-mediated inhibition of apoptosis and proliferation via induction of p21$^{waf/cip}$ and p27^{kip1} pathways [59-60].

Under normal states cholangiocytes produce little IL-6 but significant increases are seen in conditions associated with inflammation such as PSC and in CCA [61-62]. The autocrine release of IL-6 by CCA tumour cells has been shown to enhance tumour growth by activating pro-survival p38 mitogen-activated protein kinase and causing evasion of apoptosis via upregulation of myeloid cell leukaemia-1 (MCL1) through STAT3 and AKT related signalling pathways [63-64]. COX-2 and phospholipase A2 inhibitors have been shown to reduce IL-6 stimulated tumour growth [65].

TNFα, in addition to having an inflammatory role, has been shown to be overexpressed in CCA associated with hepatolithiasis and promote invasiveness and migration of malignant cells via interaction with its receptor TNFR2, tumour expressed CXCR4 and

activation of matrix metalloproteinase-9 (MMP-9) [66-67]. More recently, TNFα induced aberrant production of activation-induced cytidine deaminase in CCA and PSC was shown to result in mutations in tumour suppressor genes (p53, c-myc), providing further evidence linking bile duct inflammation and enhanced predisposition to CCA [68].

CCA cells are highly infiltrative and are supported by neovascularisation and angiogenesis networks [69] as has been recently reviewed [70]. Increased levels of vascular endothelial vascular growth factor (VEGF) and VEGF-C (a factor associated with lymphogenesis) have been detected in CCA tumour tissues and cell lines and shown to be important in cholangiocyte proliferation during bile stasis in animal models [71-72]. More recently VEGF has been shown to have an important role in mediating the proliferative effects of oestrogens on human CCA [73]. Several angiogenic-associated factors including angiopoeitin 1 & 2 and thrombospondin-1 have been investigated for prognostic utility in CCA but no relationship with clinico-pathological factors has been detected [74].

Effective tumour invasion and neovascularisation requires the surrounding stromal matrix to be degraded. The family of MMPs play an important role here and have been studied in cholangiocarcingenesis with MMP-2, MMP-7 and MMP-9 being shown to enhance breakdown of basement membrane during tumour invasion [75-76]. Overexpression of human aspartyl (asparaginyl) B – hydroxylase (HAAH) – an enzyme that catalyzes the hydroxylation of b carbons of specific residues on EGF like domains on keratin proteins - has been shown to be overexpressed in CCA cell lines and confer increased motility and invasion properties [77].

Perineural invasion is a poor prognostic feature in CCA and its importance has been reviewed [78]. The proliferation of CCA through perineural invasion is a process with

multiple factors and pathways and includes the overexpression of molecules such as nerve growth factor (NGF) [79], ErbB2 and MMP-9 [80] and Neural Cell Adhesion Molecule (NCAM) [81]. A summary of the key mechanisms regulating carcinogenesis is illustrated in Figure 2.
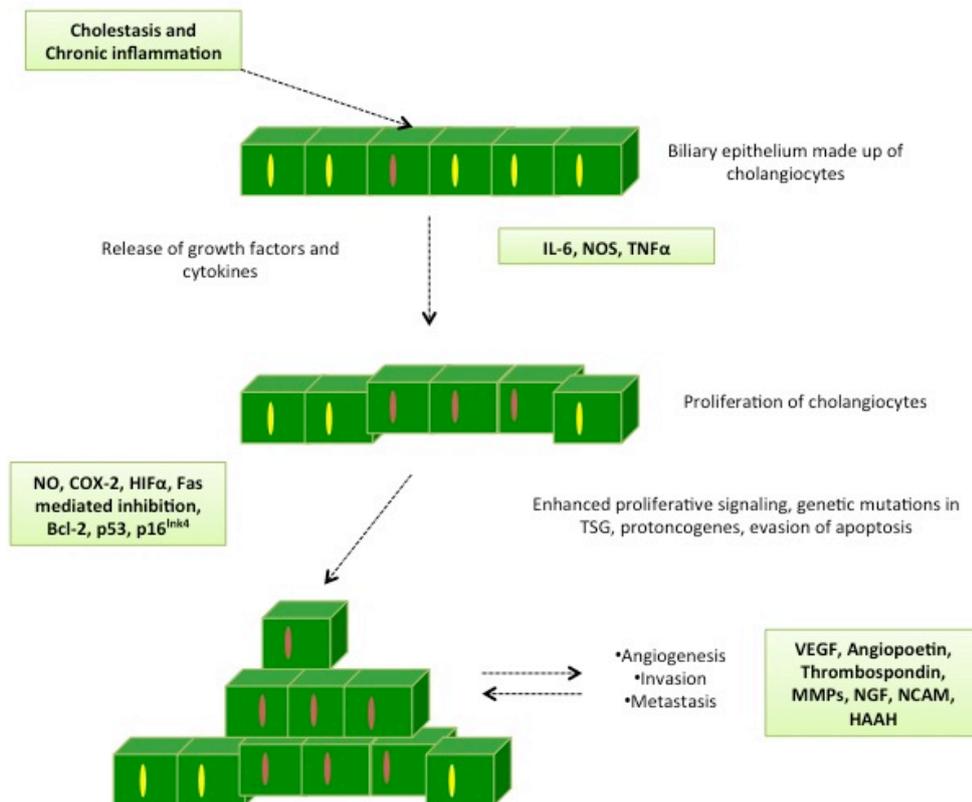
Figure 2: Summary of key mechanisms regulating carcinogenesis in CCA.

A common and important factor in the malignant transformation of cholangiocytes is chronic inflammation. Injury of the bile duct epithelium and bile stasis results in proliferation of cholangiocytes. Highlighted are proteins and genes implicated in the autologous promotion of proliferative signalling, evasion of apoptosis, angiogenesis, stromal invasion, and metastasis observed in CCA.

1.1.6 Treatment and Surgical Resection Outcomes

At present, surgical resection or transplantation in selected cases remains the only potential cure for patients with CCA. In those assessed as eligible for surgical resection based on co-morbid status and absence of extrahepatic disease, current BASG/BASL guidelines recommend *en bloc* resection of the extrahepatic bile ducts and gallbladder, regional lymphadenectomy and roux en y hepaticojejenostomy for perihilar type I & II tumours (Bismuth-Corlette classification), and in addition a right or left hepatectomy and extended right or left hepatectomy respectively in type III and IV patients [82]. At St James University Hospital, Leeds, all patients with type II-IV disease will undergo caudate (segment 1) lobectomy in addition to the above guidelines [83]. Intrahepepatic and distal CCA are treated by resection of the involved hepatic segment and pancreaticoduodenectomy respectively [82].

Despite significant progress improving the operative morbidity and mortality of aggressive liver surgery in recent years, survival outcomes for CCA remain poor. In hilar CCA outcomes have not changed significantly in the last 30 years and even after R0 resection (microscopic clear resection margin) 5-year survival is around 20-40% [84]. Analyses of the St James University Hospital, Leeds surgical experience in resection outcomes show overall 5-year survival of 20%. As in extrahepatic CCA, achievement of R0 resection is paramount to confer long-term survival in intrahepatic CCA. However resectability and 5-year survival rates are still low and vary considerably between 18-70% and 20-40% respectively [85]. In distal CCA, pancreaticoduodenectomy is reported to achieve 5-year survival rates of 20-30% [86]. Chemo and radiotherapy have to date failed to confer significant survival advantage and no standard adjuvant protocols have been established in CCA. However results of the first UK randomised clinical trial

evaluating adjuvant chemotherapy with capecitabine compared to expectant treatment alone (BILCAP) following surgery for CCA are still pending.

CCA in PSC has previously been considered a contraindication for liver transplantation due to universally poor results. However recent experimental protocols using chemo/radiotherapy prior to transplantation have reported survival rates comparable to patients without the diagnosis of CCA [87]. Furthermore, data from patients undergoing transplantation for PSC found to have incidental early CCA were associated with 70-80% survival at 5-years, providing proof of principle that early detection and surgery can be associated with improved outcomes [88].

However more than two-thirds of patients are not amenable to resection due to late presentation and stage of disease with median survival between 9-12 months, reflecting the inability to diagnose disease early [3]. More recent outcomes of photodynamic therapy in combination with biliary stenting for unresectable disease have shown improvement in quality of life and overall survival [89].

1.1.7 A Diagnostic Challenge

Current diagnostic modalities lack the sensitivity and specificity to facilitate early identification of CCA and distinguish between benign and malignant biliary strictures (particularly in PSC). At present, clinical suspicion together with biochemical abnormalities in liver function, tumour biomarkers like CA19-9, radiological imaging features (ultrasound, computed tomography and magnetic resonance cholangiopancreaticography), endoscopic retrograde cholangiopancreatography (ERCP) and combinations of endoscopic ultrasound guided fine needle aspiration (FNA), tissue biopsy, immunohistochemistry, and cytology help support a diagnosis of

CCA. Each of the described modalities has limitations and these have been reviewed [22], [90-95].

In brief, pathological diagnosis remains a considerable challenge as a result of tumour location, size and associated desmoplastic reaction. FNA is not utilised in hilar CCA due to close proximity of vascular structures, and representative tissue obtained at endoscopy is limited by the intense fibrotic reaction in and around the tumour resulting in poor diagnostic accuracy. Cytological assessments of bile and endobiliary brushings of the main duct have a sensitivity of 33-56% and 62% respectively [96-97]. Indeed 15% of patients who go onto surgical resection are subsequently found to have non-neoplastic lesions such as chronic fibrosing or erosive inflammation, sclerosing cholangitis, or a granular cell tumour [98]. In patients with stable asymptomatic PSC, routine screening by ERCP is not recommended due to the risk of iatrogenic pancreatitis [99].

At present no single or panel of biomarkers have sufficient accuracy to independently confirm a diagnosis of CCA or distinguish malignant and benign biliary strictures. Potential serum and bile biomarkers in CCA have been reviewed [100-102]. The use of the most common circulating tumour biomarker, carbohydrate antigen 19-9 (CA19-9) and several biliary markers are discussed below.

1.1.8 Carbohydrate Antigen (CA) 19-9 and Circulating Biomarkers

CA19-9 is a carbohydrate tumour-associated antigen, also known as sialylated Lewis A, originally isolated from mice immunised with a human colorectal cancer cell line [103]. It is elevated in cancers of the gastrointestinal tract, including biliary, pancreatic,

hepatocellular, gastric and colorectal. In patients without PSC, a value greater than 100U/l has a sensitivity of 53% and specificity of 75-90% for CCA (90). However its role is limited by several factors. The Lewis phenotype is absent in 7% of the population, CA19-9 is elevated in benign conditions of the biliary tree (cholethiasis, pancreatitis, cholangitis), and particularly relevant in extrahepatic CCA, is raised in benign causes of jaundice [104-106]. Indeed the British Association for the Study of the Liver, British Society of Gastroenterology and American Society of Clinical Oncology do not advocate the use of CA19-9 for screening, evaluation of resectability or disease follow-up [82], [107].

CA19-9 is still measured in patients with stable, asymptomatic PSC undergoing annual non-invasive surveillance. A value greater than 100U/l has a sensitivity of 75-89% and specificity between 80-86% for diagnosis for CCA and greater accuracy can be achieved in combination with radiological imaging [90]. However less than 25% of patients are actually amenable to surgical resection using such cut off values and ultimately this limits the clinical utility of such a biomarker in CCA [108]. Furthermore, there are no significant outcome studies or cost effectiveness models advocating this surveillance approach [109].

Combination of CA19-9 with other tumour markers including carbohydrate antigen 125 (CA 125), carcinoembryonic antigen (CEA), IL-6, apolipoprotein AII, MUC5AC and MAC-2-binding protein (identified by a proteomic approach and discussed below) have been promising but await large-scale validation studies.

1.1.9 Biliary markers of CCA

Efforts have been made to incorporate knowledge of the pathogenesis of CCA into the analysis of markers in bile as it represents a proximal biofluid favouring the enrichment of tumour-derived products. Bile insulin-like growth factor-1 (IGF-1) measured in patients with biliary obstruction undergoing ERCP with extrahepatic CCA was 15-20 fold greater than from pancreatic cancer and benign biliary disease [48]. The pancreatic elastase/amylase ratio, CA19-9, CEA, minichromosome maintenance (MCM) 2 & 5 proteins have all been assessed but no single marker has been subjected to large-scale validation [100].

Albin *et al* have demonstrated differential (1) H magnetic resonance spectrometry profiles of phosphatidylcholine, lipid, cholesterol and bile acid in patients with CCA, PSC and benign biliary disease [110]. However the role played by factors such as degree of biliary obstruction, liver dysfunction was not accounted for and limit its use in clinical practice. Early proteomic approaches to bile biomarker discovery have also identified relevant molecules and are discussed below.

Hence there exists a pressing clinical need for biomarkers enabling the early diagnosis of CCA and distinguishing between benign and malignant biliary strictures, distinguishing CCA from metastatic adenocarcinoma (especially pancreatic carcinoma) and poorly differentiated HCC with pseudo-glandular growth patterns and monitoring therapeutic response if current survival outcomes are to improve.

1.2 Proteomics

1.2.1 Biomarker Discovery and Proteomics

Biomarkers are defined as measurable molecular indicators of normal, pathological processes, or pharmacologic responses to therapeutic intervention [111]. In addition they may have potential as therapeutic targets or offer options for individualised treatment for cancer patients. Current clinical examples include Prostate Specific Antigen (PSA) and precursor PSA isoforms in identifying those at risk of prostate cancer and recurrence after treatment, and KRAS mutation detection in colon cancer patients to select optimal chemotherapies [112-113]. Efforts still continue for the development of more accurate and cost efficient biomarkers for clinical translation.

Proteomics refers to the study of the entire protein complement expressed by the genome of a tissue/cell type, biological fluid or organism [114]. It has several advantages over genomic approaches for identification of biomarkers or pathways altered in cancer progression. Proteomics allows the measurement of changes in protein level occurring as a result of post-transcriptional regulation of gene expression, including protein turnover which are not always apparent at the mRNA level and facilitates the study of posttranslational modifications (PTM) altering the form but not the level of protein. PTMs such as glycosylation have been implicated in HCC [115], and phosphorylation plays a key role in EGF/EGFR downstream signaling cascades relevant to several cancer pathways [116]. The proteome is highly dynamic and responsive to subtle changes in the intra and extracellular environment induced by normal physiological and pathological states.

Proteomic approaches to the discovery of biomarkers in human body fluids and tissue have received increasing interest in the last decade, raising hope for the early detection of malignancy, recurrence and prognostication [117]. To date more than several hundred potential cancer biomarkers have been proposed but very few have been translated into clinical practice. The recent FDA approval of the ovarian tumour triage test (OVA1) identified using surface-enhanced laser desorption/ionisation (SELDI) is one exception. Investigators have shown that this test, in combination with clinical assessment, has >90% sensitivity and negative predictive value in patients with cancer and can facilitate earlier referral and potentially improve treatment outcomes [118].

1.2.2 Proteomic techniques

Proteomics can be applied to cell lines, tissues and/or biological fluids and the proteome can be analysed using a range of complementary proteomic techniques for biomarker identification. In simple terms the approaches can be divided into 'top down' where separation and quantitation are carried out at the protein level followed by protein identification by mass spectrometry (MS) and 'bottom up' where protein mixtures are digested with a protease (most often trypsin) and the resultant peptides subjected to initial LC separation and MS/MS sequencing. Both these approaches are discussed in more detail in 1.2.2.3 & 1.2.2.4. Mass spectrometry has become a central driving tool in proteomic based experimental design and to identify peptide sequences from MS spectra various search engines exist (MASCOT, Sequest, MAXQUANT) which can compare observed fragment ions against theoretical ion masses and/or present a calculated score predicting the chance of generating the observed fragmented ions spectra against known sequences in a protein database. To estimate

the chance of false identification, decoy bases are used allowing searches to be given

a false positive rate. Typical proteomic experimental designs are shown in Figure 3.
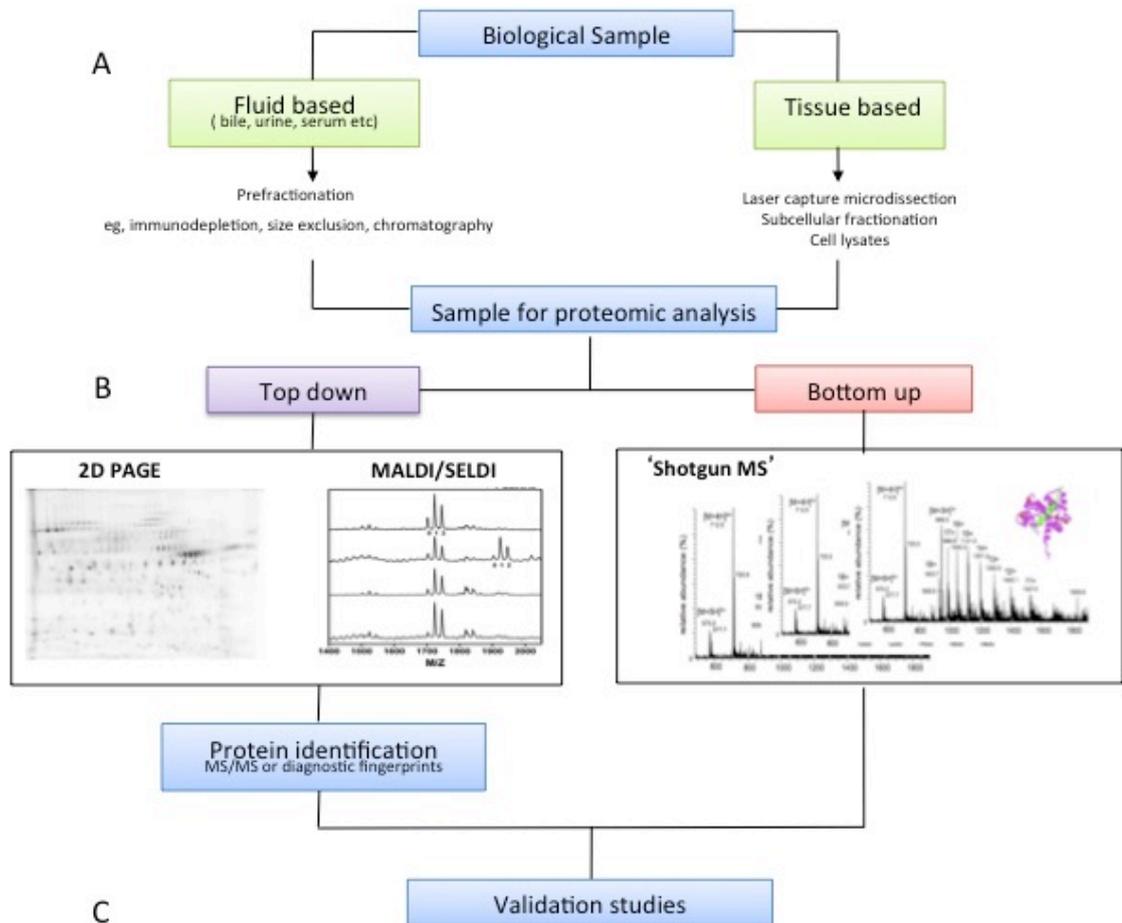
Figure 3. Typical proteomic analysis work streams.

A. Samples are prepared to reduce sample complexity and enrich potentially interesting parts of the proteome before being subjected to two main forms of proteomic approaches **B**. Top down refers to the analysis of proteins in their intact form and can be studied using 2D PAGE or MALDI/SELDI. In contrast bottom up approaches involves initial protease digestion (usually by trypsin) of protein samples into peptide fractions and separation by high performance HPLC before MS. This is a powerful approach used in current proteomic studies able to detail thousands of proteins, quantify difference in abundance and extend the proteome coverage beyond traditional 2D PAGE limits of resolution. MS remains the central driving tool to identify protein/peptides in combination with vast protein databases. **C**. After identification of proteins from data sets preliminary validation is undertaken using several techniques include ELISA, immunoblotting and immunohistochemistry.

1.2.2.1 Sample Collection and Preparation

Irrespective of the technique chosen, a crucial aspect of any proteomic investigation is attention to pre-analytical factors, sample integrity and handling to ensure experiment validity and reproducibility [119-121]. No one set of conditions is ideal for all proteins in tissue and biological fluids but attention to consistency in sample banking, with the use of standard operating procedures, careful handling of specimens to avoid degradation and artefactual bias is crucial. Important considerations include the duration and temperature of storage of sample, the type of collection tube, delays in specimen processing and heterogeneity in patient characteristics [122-123].

Tissue can be contaminated by blood and other cells at time of collection and this can be a challenge in quantitative analysis. Laser capture microdissection can be used to overcome tissue heterogeneity by isolating specific cell types and has been employed in a recent study in combination with accurate mass tags to identify several differentially expressed proteins in CCA including in particular vimentin, annexin A2, and actin binding proteins (cofilin-1, profilin-1, and transgelin-2) [124]. Cell lines are an alternative source of homogenous enriched populations of normal and cancer cells used in proteomic studies including all main hepatopancreaticobilary malignancies [125-127]. However one main limitation of their use is that not all *in vitro* changes are consistently reflected *in vivo*.

With respect to bile a number of factors have limited biomarker discovery to date including the collection of suitably large sample banks from patients with CCA and appropriate controls such as patients with PSC and benign disease and the invasive method (ERCP/surgery) required obtaining bile. Furthermore the protein content of bile represents only 7% of the total solute and a large component is represented by bile

salts, phospholipids, bilirubin, fatty acids and cholesterol [128]. These non-protein components (particularly salts) can hinder analysis based on electrophoretic protein separation and require centrifugation, desalting and delipidation of samples to improve improved resolution in gel-based approaches.

1.2.2.2 Sample Prefractionation

An extension to sample preparation is prefractionation to reduce complexity in an effort to improve proteome coverage. In the case of biological fluids such as plasma/serum, the range of protein levels exceeds at least 10 orders of magnitude with a small number of proteins (approx. 40 main plasma proteins) dominating the profile. Serum albumin and immunoglobulins are the most abundant circulating proteins accounting for more than 85-90% of the plasma profile. The removal of the most abundant proteins using commercially available products such as Multiple Affinity Removal System (MARS) Human-14 (Agilent Technologies) and ProteoPrep®20 (Sigma) has been adopted as one strategy but as yet failed to show a paradigm shift in the detection of lower abundant proteins [129]. An alternative approach is the purification of particular protein groups such as glycoproteins or phoshoproteins from samples of interest to reduce sample complexity prior to proteomic analysis [130].

1.2.2.3 Top-down proteomic approaches

The classical technique used in top down approaches is two dimensional polyacrylamide gel electrophoresis (2-D PAGE) [130]. Proteins are separated based on isoelectric point, *pI* (pH at which a particular molecule or surface carries no net electrical charge) in the first dimension by focussing with immobilised pH gradients and subsequently based on separation by molecular weight (MW) by sodium

dodecylsulphate (SDS) PAGE in the second dimension. A single format gel can resolve between several hundred and more than two thousand protein species and detect <1ng of protein per spot [130]. After separation, proteins are visualised by various staining techniques (Coomassie Blue, silver or fluorescent dyes such as SyproRuby) and image analysis carried out to identify proteins that differ in intensity between sample groups. The majority of 2D PAGE studies currently adopt difference in gel electrophoresis (DIGE) where proteins are labelled with spectrally resolvable, size and charge-matched fluorescent lysine reactive cyanine (Cy) dyes (Cy2, Cy3 and Cy5), and simultaneously resolved by 2D PAGE [131]. Incorporation of a pooled internal standard labelled with one of the Cy dyes and run on every gel improves gel matching and quantitation, thereby improving the identification of differentially expressed proteins [132]. Standard DIGE experiments useminimal labelling, where each protein carries a maximum of one dye molecule and a total of <5% proteins are labelled in the reaction. An alternative is saturation labelling where all of all the cysteine residues are coupled with dye, which can be used in cases requiring increased sensitivity. To increase proteome coverage zoom gel electrophoresis with narrow range immobilised pH gradients can be used although in some cases this may only identify further isoforms of the same proteins rather than reveal novel proteins [133]. A major advantage of 2D gels is their ability to detect posttranslational modifications that alter spot position whilst limitations include sample throughput and under-representation of proteins of extreme pI or MW and hydrophobic membrane proteins [134-135].

Alternative top down approaches include the use of other (usually two dimensional) separation techniques such as the ProteomeLab™ PF 2D system (Beckman Coulter) which separates proteins by chromatofocussing and followed by reverse phase chromatography. Matrix-assisted laser desorption/ionisation (MALDI)-based peptide profiling, often following chromatographic separations for example using ProteinChip

arrays with SELDI technology or magnetic beads coated with different chromatographic surfaces such as the CLINPROT system is an alternative approach [136]. An extension of this is imaging MS where tissue sections are directly analysed by MALDI allowing spectral profiles to be linked to histology [137].

1.2.2.4 Bottom-up proteomic approaches

In bottom-up – often termed shotgun or multidimensional protein identification technology (MudPIT) – approaches, protein mixtures are initially digested with a protease generating peptides that are separated by (multidimensional) liquid chromatography (LC) and analysed by tandem MS generating a catalogue of protein identities. In addition to simply profiling proteins in a biological sample, the field of quantitative proteomics has emerged as an important speciality in biomarker studies to demonstrate differential protein levels between disease and non-disease. Broadly speaking these can be separated into the use of stable isotope labelling or label-free techniques. Common labelling techniques include stable isotope labelling of amino acids in culture (SILAC), which involves metabolically labeling with light/heavy versions of amino acids lysine and arginine, generating peptide pairs - the intensity of which indicates relative abundance [138]. Other approaches for introduction of stable isotope labels include digesting in ^{18}O labeled H_2O [139] and iTRAQ (isobaric tags for relative and absolute quantification) which uses families of 4 (or 8) tags that are indistinguishable in MS but fragment to give different reporter ions in MS/MS allowing comparative analysis of up to 8 samples [140]. After labelling different proteins, samples are combined and pooled before analysis by LC-MS/MS. Labelling strategies are considered to be more accurate in quantifying protein abundance however this is at the expense of cost, complex sample preparation, and runs the risk of incomplete labelling.

In response to these challenges interest has developed in 'label-free' approaches for quantitation in shotgun investigations [141]. Like all other proteomic techniques label free methods involve all the principal steps of protein extraction, reduction, and digestion followed by sample separation using LC before analysis by MS. The data output allows for identification of proteins, their relative abundance and statistical analysis between samples. However in contrast to labelling methods each sample is separately prepared before individual LC-MS/MS runs (i.e. not pooled). Protein quantification is generally based on two main techniques: spectral counting or ion intensity (peptide peaks areas or heights in chromatography).

In the spectral counting approach, the relative protein quantification is achieved by comparing the number of identified MS/MS spectra from the same protein in each of the multiple LC-MS/MS runs. It relies on the basis that an increase in protein level will result in a greater number of proteolytic peptides and hence greater protein sequence coverage, unique peptides and the number of identified total MS/MS spectra (spectral count) for each protein. Examples of relevant protein biomarker studies utilising this technique include identification of serum biomarkers for colorectal cancer metastasis, in the differential analysis of the A549 lung carcinoma cell line and in identifying elevated levels of neutrophil gelatinase-associated lipocalin (NGAL; lipocalin-2) in bile from patients with malignant pancreaticobiliary disease [142-144].

The fundamental concept in using peptide peak intensity approach relies on an ion with a particular mass to charge ratio to be recorded in an LC-MS run and its intensity (as measured by the height or the area under the peak) being correlated with ion concentration [145]. The technique is subject to variation between different runs of samples as a result of any difference in sample preparation and injection. As a result normalisation methods are employed together with complex computer algorithms to

enable more accurate matching and quantification [146]. Such an approach has demonstrated differences between control and radiated human colon cancer cells and proteins associated with metastasis in melanoma [147-148]. However the spectral counting approach is reported to be more reproducible and with a larger dynamic range than peptide ion based quantification [149].

1.3 Proteomics in CCA

1.3.1 Tissue and Serum-Based Approaches

Kawase and colleagues identified 38 differentially expressed proteins in two paired cancer and normal bile duct tissues using a label-free LC-MS/MS approach and validated findings in further tissue samples [150]. A number of proteins previously reported in CCA including MUC5AC [151], moesin [152], galectin 1 [153] and keratin 903 [154] were increased in CCA tissues. Western blotting and immunohistochemical validation of four novel proteins (actinin-1, actinin-4, protein DJ-1 and cathepsin B) in the original samples and four additional CCA cases showed all were overexpressed compared with normal bile duct cells.

Actinin 1 and 4 play an important role in the actin cytoskeleton and cell mobility and confer cell invasion and metastatic properties [155] and studies confirmed overexpression to be associated with aggressive phenotype in colorectal and pancreatic cancer [156-157]. Protein DJ-1 is an oncogene that can promote cell proliferation and carcinogenesis in several cancers (esophageal, pancreatic, glottic) through activation of AKT, mTOR and HIF1 pathways and inhibition of p53 – mediated apoptosis [158]. Cathepsin B belongs to a family of lysosomal proteases and is active at the normal physiological pH range of bile (7-7.7). Its role in cancer is multifaceted

including extracellular matrix invasion; promotion of angiogenesis, cell migration and metastatic potential [159], and overexpression has been demonstrated previously in chemically induced CCA animal models [160].

In an effort to focus on more lower abundance and secreted proteins Kristiansen et al applied a membrane enrichment strategy coupled with a ^{18}O labeling based quantitative proteomic approach in two CCA patients [152]. In addition to several proteins previously identified in CCA, MUC-5AC, golgi membrane protein 1, annexin IV and EGFR pathway substrate 8 (EPS8) were shown to be overexpressed. Golgi membrane protein 1 is a transmembrane protein located on epithelial cells (including cholangiocytes) of unclear function. It has relevance to various inflammatory and chronic liver diseases and has been identified in urine of patients with prostate cancer and shown to be superior to the use of AFP alone in HCC surveillance in cirrhotic patients [161-163]. Annexin IV belongs to a family of Ca^{2+} regulated phospholipid binding proteins and believed to be involved in exocytosis and regulation of epithelial chloride ion secretion [164]. In cancer its role is not clearly established, although proteomic data from RCC and colorectal cancer implicate changes in cell migration by way of loss of cell to cell adhesion [165-166]. EPS8 forms a complex with Sos 1 and Abi1 facilitates the activation of specific proteinases involved in actin remodeling and increasing metastatic potential [167].

Malignant cholangiocytes tend to be embedded in and around desmoplastic tissue [168] and can render identification of tumour-derived protein signatures difficult in proteomic studies. However, using laser microcapture dissection and accurate mass and time tag approaches, Santos and colleagues produced enriched cancer cell lysates from 4 patients with intrahepatic CCA and compared them to control bile duct tissue [124]. Proteins were separated by SDS PAGE and subjected to in-gel tryptic digestion and

LC-MS/MS to generate a list of more than 500 proteins of which 39 were differentially expressed and related to cancer-associated pathways. Many of these differentially expressed proteins were involved in glycolysis and regulation of the cytoskeleton and specifically, Mac 2 binding protein and vimentin (previously identified by proteomic approaches in CCA) were increased in intrahepatic CCA [169-170]. Other elevated proteins included actin binding proteins (profilin-1, cofilin-1 and transgelin) associated with cell transformation and motility, S100A11 (calgizarin) involved in Ca^{2+} signalling, cell regulation and motility and TGFβ which can promote metastatic potential by inducing epithelial to mesenchymal transition [171].

Studies in serum have mainly used SELDI to identify a number of differentially expressed peptides in CCA [172-173]. Liu et al analysed 427 serum samples from patients with CCA (n=56), benign hepatobiliary diseases (n=49), other cancer controls (n=269), and healthy individuals (n=53). Peaks with m/z 13.760, 13.880, and 14.040 were significantly decreased in CCA compared with the control groups and identified as native transthyretin and its two variants. Further studies incorporating enzyme-linked immunosorbent assay (ELISA) of transthyretin and CA19-9 demonstrated significantly down-regulated levels of transthyretin in sera of CCA patients.

Wang *et al* analysed serum samples from 60 patients with CCA, 60 with benign hepatobiliary disease and 53 normal individuals and detected apolipoprotein A-I as being significantly decreased in CCA and able to discriminate between the groups [173]. However, no single serum marker identified by a proteomic approach has been subjected to a large validation experiment.

1.3.2 Proteomics studies in CCA – Cell lines

The CCA cell line HuCCA-1 was established in 1991 [174] and only a few studies have been reported using it in proteomic studies [126], [170], [175-176]. Srisomap et al compared the protein pattern of the HuCCA cell line to the human HCC (HepG2 and HCC-S102), and breast (MCF-7) cancer cell lines using a 2D PAGE and LC-MS-MS approach. They established the first proteomic map of a HuCCA-1 cell lysate and identified 43 major proteins. Of these cytokeratin 7 and 19 were present only in CCA and although there was one patient in the study, the differential expression of cytokeratin 19 had been reported previously in immunohistochemical studies [177].

To determine the effects of *C. sinensis* infection (a known risk factor for CCA) and their excretory secretory products (ESP) on protein expression in host bile duct epithelium, Pak *et al* analysed differences in HuCCT1 cells treated with ESP [176]. Using a combination of 2D PAGE and MALDI-TOF MS analysis, 83 proteins were differentially expressed in response to ESP, of which 49 were up-regulated and 34 down-regulated. Specifically ESP induced the expression of redox-regulating proteins, including peroxiredoxins (Prdx 2, 3, and 6) and thioredoxin 1 (Trx 1), potentially via intracellular ROS generation.

In an effort to provide a comprehensive proteome-wide analysis of tumour surface membrane proteins, Yonglitthipagon et al utilised a sequential protein extraction method to study four CCA cell lines with different malignant potential and a non-malignant H69 biliary cell line control [175]. Using 2D PAGE and MALDI-TOF-MS differentially expressed proteins were identified. Annexin A2 upregulation was validated by western blotting and tissue microarray in 300 CCA patients and found to be associated with lymphatic invasion, metastasis and prognosis.

1.4 Proteomics approaches in Bile

Only a small number of studies have reported the proteomic investigation of human bile and these have been reviewed [128], [178]. One of the earliest studies described the successful application of 2D PAGE to bile samples from individuals with cholethiasis and identified 77 proteins [179]. The same investigators subsequently reported a more in depth analysis of vesicular and micellar protein profiles in gallbladder bile [180]. Bile was analysed by 2D PAGE after purification steps involving ultracentrifugation, delipidation and gel filtration chromatography resulting in identification of nearly 500 (mainly plasma) proteins. Efforts to identify hydrophobic proteins in bile in subsequent studies using extraction with organic solvents and subsequent chromatography and MALDI-TOF MS resulted in the discovery of three previously unknown biliary proteins: MAC-2 binding protein, CD14 and ATP synthase lipid binding protein [181].

Kristiansen *et al* produced a comprehensive catalogue of proteins from bile of a patient with CCA, sampled in the common bile duct by ERCP [182]. Bile was centrifuged, delipidated and concentrated by 3 kDa size exclusion filter prior to analysis by SDS PAGE and LC-MS/MS directly or following lectin affinity chromatography -/+ immunoglobulin depletion. Overall 87 unique proteins were identified, the majority of which were proteins derived from hepatocytes or of pancreatic origin. In addition to several known "cancer-associated proteins" such as CA125, MAC-2 binding protein and MUC 2, two proteins not described in the context of CCA were reported: lipocalin-2 and deleted in malignant brain tumours 1. MAC-2 binding protein was evaluated as a potential diagnostic marker using ELISA in bile and serum from patients with CCA, PSC and benign biliary diseases [169]. Results demonstrated a 3-fold increase in biliary MAC-2 binding protein in patients with CCA compared to non-neoplastic diseases, and giving a similar receiver operator characteristic value (0.7) as biliary CA19-9 and

greater diagnostic accuracy when used in combination with CA19-9. However MAC-2 binding protein was not elevated in serum from patients with CCA.

Zhou *et al* used the complementary techniques of 2D PAGE and shotgun proteomics to analyse gallbladder bile from an individual with cholethiasis [183]. Following dialysis, precipitation and delipidation, samples underwent 2D PAGE or direct protein digestion, separation of peptides by strong cation exchange and reversed-phase chromatography prior to 2D-LC-MS/MS. A total of 222 (mainly plasma) proteins were identified (48 and 218 from 2D PAGE and shotgun approaches respectively), 44 of which were common to both techniques and 27 proteins in common with the Kristiansen study.

In the only other study comparing bile from an individual with CCA to benign biliary disease (control), a 2D PAGE approach resolved 250 and 216 spots on pH 3-10 and 182 and 175 spots on pH 4-7 IPG strips respectively [184]. There were 16 and 23 spots differentially expressed between malignant and non-malignant disease using the different IPG strips. The study did not provide MS identification of proteins of interest but highlighted a reproducible methodology for high-resolution separation of biliary proteins and formation of 2D bile maps.

In contrast, Guerrier *et al* studied the pooled gallbladder bile of nine patients with no biliary tract disease. In addition to ultracentrifugation and delipidation, samples were treated with immobilised hexapeptide ligand libraries in an effort to concentrate and reduce levels of low and high abundance proteins respectively [185]. Samples were then subjected to three distinct elution steps using solutions with different ionic strengths and pH before comparison with unfractionated bile using LC-MS/MS or SELDI-MS. Including proteins identified with a single peptide, 141 and 222 proteins were identified in unfractionated and fractionated bile respectively, of which 143 had not been

previously reported in proteomic studies, and 81 proteins were a result of application of ligand libraries.

The most recent large-scale proteomic investigation of bile sampled from the bile duct of two patients with pancreatic cancer by ERCP identified more unreported biliary proteins [186]. Samples were subjected to ultracentrifugation, non-ionic adsorbents and size exclusion filtration before analysis by SDS PAGE and LC-MS/MS. Some samples underwent further purification steps involving acetone/ethanol precipitation, methanol/chloroform extraction or albumin/immunoglobulin depletion. Furthermore pellets after ultracentrifugation were included in the analysis revealing proteins that would have otherwise been lost. Overall 127 proteins, of which 34 were identified by two or more significant peptides and 38 by a single significant peptide, were not previously described in bile.

In summary, studies to date have demonstrated the proof of concept of proteomic analysis of samples including bile in CCA but have yet to include a significant number of patient samples in investigations, malignant bile obtained from within the biliary ducts, perform more intensive prefractionation strategies including the concentration of low abundance proteins, produce comprehensive protein maps and investigate protein expression profiles between intra and extrahepatic CCA.

1.5 Aims of Study

Principal Aim

To identify potential circulating tumour derived biomarkers utilising proteomic techniques applied to bile.

Work streams

1. To investigate the differences in protein profiles between serum and bile using 2D PAGE.

2. To develop and optimise antibody-based depletion strategies of highly abundant proteins from bile to facilitate the discovery of lower abundance molecules more likely to represent potential biomarkers.

3. To characterise the malignant proteome in bile of four patients with hilar CCA using complementary approaches of 2D DIGE and shotgun GeLC-MS/MS.

4. To perform a quantitative comparative analysis of prefractionated bile samples from patients with CCA, and benign biliary disease using a label free proteomic approach.

5. To analyse molecules of interest using antibody-based approaches (Western blotting) in validation studies on larger bile sample groups.

2.0 Materials and Methods

2.1 Materials

Milli-Q water was used throughout. General chemicals (AnalaR grade or equivalent) were purchased from Sigma (Poole, UK), ICN Biomedicals (Cambridge, UK), and VWR (Poole, UK). IPG strips, Bromophenol Blue, Albumin/IgG Removal Kit, CyDye DIGE fluor Cy3 and Cy5 minimal dyes, Hybond-C super nitrocellulose membrane (GE Healthcare, Buckinghamshire, UK), Criterion Precast Gels 10.5-14%, Bio-Rad protein assay (Bio-Rad, Hemel Hempstead, UK), ProteoSilver™ Silver Stain, Protease Inhibitor cocktail, Coomassie Colloidal Blue Stain (Sigma, Poole, UK), InstantBlue™ Coomassie Colloidal Stain (Expedeon, Cambridgeshire, UK), PAG backing (Lonza, UK), Trypsin (Promega, Southampton, UK), Acrylamide (National Diagnostics, Hessle UK), Agarose LMP (Invitrogen, Paisley, UK), CHAPS (Merck, Nottingham, UK), Swell Gel Blue Albumin Removal Kit, Supersignal West Extended Dura Substrate, BCA assay (Pierce, Tattenhall, UK), silicone thin walled tubes (Bioquote, York, UK), MMP-9/Lipocalin Complex (Calbiochem, Darmstadt, Germany) and Omnipaque dye (Amersham Health, Cork, Ireland), anti-mouse and anti-rabbit HRP– conjugated Envision + reagents (Dako, Ely, UK). Full details of suppliers and constituents of all solutions and buffers are listed in Appendix 1 and 2 respectively.

2.2 Clinical Samples

All bile and serum samples were obtained following informed consent and ethical approval (REC 06/Q1206/136 Appendix 5) and included samples from the following groups of patients: (a) Malignant disease (CCA), (b) predisposed disease (PSC), (c) benign disease (cholethiasis, strictures), and (d) 'normal' (transplant organ donors).

Specifically consent was taken by specialist donor coordinators in cases where bile was obtained from organ donors. No bile was collected from the gallbladder or via percutaneous transheptic drains to ensure bile samples were obtained in a homogenous pattern. Appendix 3 lists all samples collected including basic demographics and group status.

2.2.1 Bile Sample Collection – Surgery

For patients undergoing surgery for CCA, or during multiorgan procurement for transplantation the common bile duct (CBD) was isolated and ligated at its distal end to allow accumulation of bile for 5-15 minutes before collection through a 22 FG silastic catheter into a 10 ml syringe and transferred to a Falcon tube under sterile conditions. In cases of resection and organ donors bile was obtained before division of the CBD. Only donors who are classified as 'heart beating – brain dead' donors were used. Donation after cardiac death donors were excluded from sample collection. All attempts were made to prevent the contamination of bile samples with blood and tissue. Bile was transported to the processing laboratory on ice within 15 minutes.. To ensure samples reached the laboratory in 15 minutes, bile obtained from multiorgan procurement operations only occurred at St James Universoty Hospital or Leeds General Infirmary. Only bile from donors in which the liver was implanted and the status of the parenchyma known was used.

2.2.2 Bile Sample Collection – ERCP

Bile was collected from patients undergoing ERCP diagnostic/therapeutic interventions for biliary disease (benign and malignant) and without a pre-existing CBD stent *in situ*.

After cannulation and placement of catheter, 5-20ml Omnipaque dye was inserted to confirm its position within the bile duct. An equal volume of fluid to dye injected was discarded to minimise contamination, before collection of bile into a sterile syringe and transfer to a Falcon tube. Bile was transported to the processing laboratory on ice within 15 minutes. In those with ERCP for CBD stones liver function tests were also recorded to allow optimal matching of representative samples in comparative proteomic analysis.

2.2.3 Blood Sample Collection

Blood was obtained prior to surgery and ERCP from a peripheral site via standard technique into a 6 ml Vacuette serum clot activator and two 4ml Vacuette EDTA tubes. Blood was transported on ice to the laboratory within 15 minutes.

2.2.4 Sample Processing

Aliquots of bile in 1.5 ml eppendorf tubes were subjected to centrifugation at 4°C for 15 minutes at 13,000 x g, aliquotted and stored at -80°C. Venous blood samples were left to clot at room temperature for 60 minutes before centrifugation at 20°C for 10 minutes at 2,000 x g. Serum was removed using a fine point pastette, aliquotted and stored at -80°C.

2.3 Protein Assay

Protein concentration was determined using the modified Bradford assay following the manufacturer's instructions or alternatively by densitometric analysis of Coomassie

stained gels using a standard curve (0-5mg/ml) formed with a serial dilution of albumin depleted serum (ADS). ADS was prepared using Swell Gel Blue Albumin Removal Kit following the manufacturers instructions.

2.4 Sample Purification and Prefractionation

2.4.1 Trichloroacetic Acid (TCA) Precipitation of Bile

Desalting and concentration of bile samples was carried out by TCA precipitation. In brief, an equal volume of 20% w/v TCA was added to bile and the mixture allowed to precipitate on ice for 30 minutes before centrifugation for 10 minutes at $4°C$ and 18000 x g. The supernatant was removed without disturbing the pellet, 1 ml of ice-cold acetone wash was applied to the pellet and centrifugation repeated. The acetone was removed, the pellet was left to stand for 2 minutes to allow evaporation of any residual acetone then resuspended in DIGE lysis buffer (7M urea, 2M thiourea, 4% w/v CHAPS).

2.4.2 Albumin and IgG Depletion of Serum and Bile

To improve the detection/coverage of lower abundance proteins, two components predominating the serum and biliary proteome - albumin and immunoglobulin G -were removed using the Albumin/IgG Removal Kit. The kit was originally designed for albumin/IgG depletion from serum, and optimisation experiments were undertaken to assess the efficiency of depletion in bile. SDS PAGE and Coomassie staining and 2D DIGE were used to compare the depleted samples to unprocessed bile to assess removal of albumin/IgG. More detail for experimental technique is described in 3.2.2.

2.5 1D SDS-PAGE

Serum and bile samples were adjusted to 1X Laemmli buffer (62.5mM Tris-HCl pH 6.8, 10% v/v glycerol, 5% v/v β-mercaptoethanol, 2% w/v SDS, 0.0025% w/v bromophenol blue (BPB)). Protein (typically 5-7.5 µg) was separated on 10% polyacrylamide gels using the Bio-Rad Mini Protean II electrophoresis system or using Bio-Rad Criterion Gel System with precast 18 well 10.5-14% Tris-HCl gels and SDS PAGE running buffer (24mM Tris, 192mM glycine, 0.1% w/v SDS).

2.6 Coomassie Blue and Silver Staining

Gels were fixed in 40% v/v methanol, 7% v/v acetic acid for 30 minutes before staining with Coomassie Colloidal Blue Stain for 2-24 hours. Gels were destained in 25% v/v methanol, 10% v/v acetic acid for 1 minute then in 25% w/v methanol for up to 6 hours. Mini 2D gels were fixed and silver stained using the ProteoSilver™ Silver Stain kit according to manufacturers instructions. Gels were scanned as 12 bit images using a personal densitometer SI (GE Healthcare).

2.7 2D PAGE

2.7.1 First dimension

First dimension isoelectric focusing was carried out using the IPGphor system (GE Healthcare). Samples were prepared in DIGE lysis buffer (7M urea, 2M thiourea, 4% w/v CHAPS). 5 µg of sample was loaded on to 7cm pH 4-7 IPG strips in a total volume of 125µl of reswell buffer (7M urea, 2M thiourea, 4% w/v CHAPS, 0.46% w/v (30mM) DTT, 0.2% v/v pharmalyte pH3-10, trace BPB and applied by in-gel rehydration (30 V

14 hours). Isoelectric focussing was carried out to give a total of 9.1 kVh (200 V 1 hour, 500 V 1 hour, 500-3500 V 30 minutes, and 3500 V to end) at 20°C with a current limitation of 50 µA/strip.

2.7.2 Second dimension

Strips were reduced in 5ml equilibration buffer (6M urea, 30% w/v glycerol, 2% w/v SDS, 50mM Tris-HCl pH6.8) containing 1% w/v DTT for 15 minutes, and alkylated in 5ml equilibration buffer containing 4% w/v iodoacetamide for 10 minutes. Strips were placed on 10% T polyacrylamide gels and SDS PAGE was carried out using the Bio-Rad Mini Protean II electrophoresis system as described in section 2.5.

2.8 2D DIGE

Bile samples were resuspended in DIGE lysis buffer (7M urea, 2M thiourea, 4% w/v CHAPS). Samples were then diluted to 1 mg/ml and a sample volume equivalent to 50µg was used for DIGE labelling. The pH of the sample was adjusted to pH 8.5 using Tris-HCl pH 8.5, giving a final concentration of 50 mM, and then samples were vortexed, pulsed in a microfuge and placed on ice.

1 mM stock cyanine dye (CyDye DIGE fluor Cy3 and Cy5 minimal dyes) was diluted 1 in 5 with fresh anhydrous DMF to give a dye solution of 200 µM. Samples were labelled with Cy3 or Cy5 in the ratio 50 µg protein: 200 pmol dye and left on ice for 30 minutes in the dark. The reaction was stopped by the addition of 1µl 10mM lysine per 200 pmol dye used, samples were vortexed and left on ice for 10 minutes in the dark.

An equal volume of 2x DIGE buffer (7M urea, 2M thiourea, 4% w/v CHAPS, 2% w/v DTT, 1.6% v/v pharmalyte pH3-10) was added to restore DTT and pharmalyte levels to those required for IEF. Samples were left at room temperature for 15 minutes, covered in foil. Samples were either stored at -80°C until use or used immediately to rehydrate immobilised pH gradient (IPG) strips.

2.8.1 First dimension isoelectric focusing

Samples to be run together (usually 50 μg of each of labelled Cy3 internal standard and Cy5 labelled individual sample) were combined and then made up to a total volume of 450μl with reswell buffer (7M urea, 2M thiourea, 4% w/v CHAPS, 0.46% w/v (30mM) DTT, 0.2% v/v pharmalyte pH3-10, trace BPB). Isoelectric focussing was carried out on pH4-7 IPG strips using the Ettan™ IPGphor™ 3 IEF system (GE Healthcare). Protein was applied by in-gel rehydration at 30V for 18 hours then focussing was carried out using the IPGphor Manifold Ceramic Tray (GE Healthcare) as per the manufacturer's instructions (Table 3). Strips were stored at -80°C until use.

Table 3.0 Voltage and Duration Times for IEF

Step		VOLTAGE	DURATION
S1	Step	500V	0:01Hr
S2	Gradient	3500V	1:30Hr
S3	Gradient	3500V	72000VHr
S4	Gradient	8000V	0:10Hr
S5	Step	8000V	1:00Hr
S6	Step	100V	10:00Hr
			To 85,000VHr

2.8.2 Second dimension 2D PAGE

SDS PAGE was carried out using the Ettan DALTt*welve* Separation Unit (GE Healthcare) with 8-15% gradient gels cast in low fluorescence plates using a 2DEoptimizer (NextGen Sciences Ltd, Huntington, UK). Strips were reduced in 10 ml equilibration buffer (6 M urea, 30% w/v glycerol, 2% w/v SDS, 50 mM Tris-HCl pH 6.8) containing 1% w/v DTT for 15 minutes then alkylated in 10 ml equilibration buffer containing 4% w/v iodoacetamide for 10 minutes. Strips were rinsed with 1x running buffer (25 mM Tris, 192 mM glycine, 0.1% w/v SDS, pH8.3), applied to the top of gels and sealed in place with 1% w/v LMP Agarose in 1x running buffer containing a trace of BPB. Gels were run overnight at ~1-1.5W/gel at 15°C.

2.9 Gel scanning and Analysis

Gels were scanned using a Typhoon Trio (GE Healthcare). Initial scans at 1000µm were done to optimise the PMT voltage to give a maximum intensity of 80 000 - 90 000 for the second most abundant spot after albumin or the most abundant spot following albumin depletion. Final scans were done at 100µm. Gels were analysed using Progenesis SameSpots version 3.3 (NonLinear Dynamics, Newcastle, UK).

2.10 Protein Identification

2.10.1 For 2D PAGE Mastermap Experiments

Preparative gels were generated using 500 µg protein pooled from 4 bile samples from patients with hilar CCA. Gels were backed with polyester backing and were silver stained using a modified version of the ProteoSilver™ protocol using lower percentage alcohol to reduce gel volume variations and to avoid gels detaching from the PAG backing. The following amendments were made to the protocol: Gels were fixed with 30% v/v ethanol, 5% v/v acetic acid for a minimum of 40 minutes, and the 15% v/v ethanol wash was carried out for 20 minutes. Pick lists were created and data were imported from Progenesis SameSpots into the Ettan Spot Picker V1.2 software. Gel spots with a diameter of 1.4mm were collected directly into in a 96-well plate in water using the Ettan Spot Picker(GE Healthcare).

2.10.2 Tryptic digestion

Gel pieces were destained in 50 mM sodium thiosulphate/15 mM potassium ferricyanide, washed in water and equilibrated in 25 mM ammonium bicarbonate. They were then dehydrated with acetonitrile, rehydrated in 25 mM ammonium bicarbonate, further dehydrated with acetonitrile and dried in a SpeedVac. Gel pieces were incubated on ice for 45 minutes in 5 µl (200ng) sequencing grade modified trypsin solution (0.1mg/ml trypsin, 16000U/mg prepared in ice-cold resuspension buffer supplied with the enzyme then diluted with 25 mM ammonium bicarbonate, to 40 µg/ml), followed by the addition of 30 µl ammonium bicarbonate. After digestion for 4 hours at 37°C, the supernatant was transferred in to a 0.2 ml siliconised thin walled

tube. Extraction of tryptic peptides was carried out twice through incubation in 5% (v/v) formic acid for 15 minutes in a sonicating water bath, the supernatants combined and dried in a SpeedVac. Peptides were resuspended in water and dried in a SpeedVac twice and finally reconstituted in 5 µl of 50 % (v/v) acetonitrile / 0.1% (v/v) trifluoroacetic acid. Samples were stored at -80°C until analysis.

2.10.3 Samples and Mass Spectrometry for GeLC-MS/MS 'Shotgun' Based Experiments

Protein (100 µg) from each bile sample was separated on 10% SDS PAGE using the Hoefer SE600X Chroma Deluxe Electrophoresis Unit. Gels were stained with InstantBlue Coomassie® and each lane was divided into 51 gel slices by hand using a sterile scalpel. The gel pieces were sliced into smaller pieces washed with water then with 25 mM ammonium bicarbonate and dehydrated with ACN. Samples were then reduced with DTT, alkylated with iodoacetamide and digested with trypsin as described in the Section 2.10.2.

Tryptic digests were analysed by LC-MS/MS using a nano-HPLC (Agilent, USA) and a QSTAR-XL quadrupole time-of-flight hybrid mass spectrometer (Applied Biosystems, UK) as previously described (Aggelis et al., 2009) which was performed by Dr Jianhe Peng, University of Leeds.

The mass spectrometer was operated in a data dependent mode: a MS scan from 400–1800 m/z was performed for 1 s, the three most abundant doubly- and triply-charged ions (m/z 400–1000) with intensities over 40 counts were selected for MS/MS analysis which was acquired from 80–1800 m/z for 1 s in the Enhance All mode and

precursors were then excluded for 200 s. The MS/MS data was processed by Analyst (version 2.0, Applied Biosystems) and searched using a local Mascot search engine (version 2.3, Matrix Science, London) with the following parameters – database: IPI human (89652 sequences, version 3.74); enzyme: trypsin; fixed modification: carbamindomethyl (c); variable modification: oxidation(M); peptide mass tolerance: +/- 0.15 Da; fragment mass tolerance: +/- 0.1Da; maximum missed cleavages: 1; instrument type: ESI-QUAD-TOF. Peptides were considered identified with scores above identity level, p< 0.05.   For each patient, the intermediate files produced by Mascot were combined, peptides with probability scores with p>0.05 were excluded and redundancy then removed. Furthermore peptides were filtered to remove trypsin autolysis peptides and keratins. Proteins required at least one unique significant peptide to be considered identified. For proteins identified with a single peptide, spectra were inspected manually and only those passing this quality control were included.

2.10.4 Filter Aided Sample Preparation and Tryptic Digestion

Following TCA precipitation and albumin/IgG depletion, samples for experiments in a label-free quantitative 'shotgun' approach to compare bile from patients with hilar CCA and benign biliary disease were subjected to tryptic digestion using filter-aided sample preparation (FASP) prior to HPLC-MS/MS.

In brief, 25 µl of bile sample (50 µg protein) was mixed with 225 µl of UA (8 M urea in 0.1 M Tris/HCl pH 8.5) and added to a filter unit (Amicon Ultra-0.5, Ultracel-30 membrane, 30 kDa) and centrifuged at 14,000 x g for 10 minutes. The flow through in the collection tube was discarded and the remaining sample washed twice with 250 µl of UA with centrifugation 10 minutes at 14,000 x g and the flow through again

discarded. 100 µl IAA solution (0.04 M iodoacetamide in UA) was added to the filter unit and mixed at 500 rpm in a thermo-mixer for 1 minute and incubated without mixing for 10 minutes prior to centrifugation. The flow through was again discarded and a further 250 µl of UA added and centrifuged at 14,000 x g for 10 minutes. The sample was subjected to a further four cycles of centrifugation at 14,000 x g for 10 minutes with 250 µl of ABC (0.05M NH_4HCO_3) added to the filter unit.

The filter units were transferred to new collection tubes, 45 µl of ABC and trypsin (16000U/mg prepared in ice-cold resuspension buffer supplied with the enzyme then diluted with 25 mM ammonium bicarbonate, to 40 µg/ml, enzyme to protein ratio of 1:100) was added to the samples, which were mixed at 600 rpm in a thermomixer for 1 minute. The filter units were incubated in a wet chamber at 37 ^{o}C for 18 hours and peptides collected by centrifugation at 14,000 x g for 10 minutes. Finally 100 µl of H_2O was added to the filter unit and centrifuged at 14,000 x g for 15 minutes and this flow through was combined with the initial elute. Concentration of peptides was determined using a NanoDrop 8000 Spectrophotometer (Thermo Scientific, Stafford House, Hemel Hempstead, UK) assuming that a 1 mg/ml protein solution produces an absorbance of 1.0 at 280nm when the path length is 1cm.

2.10.5 Samples and Mass Spectrometry for Label-Free Comparative Proteomic Analysis

Protein loading of bile for experiments was determined by parallel Coomassie. All mass spectrometry in the experiments in the comparative label free approach of analysis of malignant vs. benign bile were performed by Dr Alexandre Zougman, and biostatistics by Dr David Cairns, both of University of Leeds.

Peptide mixtures (3 μg) were separated by online reversed-phase capillary liquid chromatography and analyzed by electrospray tandem mass spectrometry. Each sample was analysed in triplicate. Samples were injected onto a 20 cm reversed phase capillary emitter column made in-house (inner diameter 75 μm, packed with 3.5 μm Kromasil C_{18} media) using an UltiMate 3000 RSLCnano nanoflow system (Dionex). The LC setup was connected to a linear quadrupole ion trap-orbitrap (LTQ-Orbitrap) Velos mass spectrometer (Thermo Scientific) equipped with a nanoelectrospray ion source (Proxeon). The total acquisition time was 240 min, the major part of the gradient (from 10 to 220 min) being 4–25 % ACN in 0.1% formic acid at a flow rate of 400 nl/min. Survey MS scans (scan range 300-1500 amu) were acquired in the orbitrap with the resolution set to 60000. Up to 20 most intense ions per scan were fragmented and analysed in the linear trap.

2.10.6 Protein Identification and Statistical Analysis

The acquired data were searched against an International Protein Index (IPI 3.68) human protein sequence database with MaxQuant 1.1.1.14 [187]. The maximum protein and peptide false discovery rates were set to 0.01. The subsequent complete data output was analysed using a Wilcoxon Mann Whitney test to compare the two groups of 5 samples with a permutation strategy used to estimate the false discovery rate [188-189].

2.11 Western Blotting

After 1D SDS-PAGE, gels were rinsed in Towbin buffer (25 mM Tris, 192 mM glycine, 10% v/v methanol, pH 8.3) for 15 minutes before protein transfer to Hybond- C Super

nitrocellulose membrane in Towbin buffer using the BioRad Mini Trans-Blot® Electrophoretic Transfer Cell (100 V, 1 hour). Blots were allowed to air dry for 1 - 24 hours at room temperature before being blocked in TBS-T (TBS, 0.1% v/v Tween 20), 10% w/v dried skimmed milk for 1 hour at room temperature. Blots were then incubated in primary antibody (Table 4) diluted in TBS-T 1 % w/v dried skimmed milk for 1 hour at room temperature, subjected to 4 x 5 minute washes in TBS-T before final incubation with anti-mouse or rabbit horseradish peroxidase (HRP) - conjugated Envision[+] reagent diluted 1:100 in TBS-T, 5% dried skimmed milk for 1 hour at room temperature. Repeated washes (4 x 5 minutes) of blots in TBS-T were made prior to being developed using Supersignal West Dura Extended Substrate and exposed to Kodak Biomax MS film. Western blots were normalised using parallel Coomassie stained gels and probing with a Complement-3a antibody. Negative control blots were probed with matched irrelevant antibodies used at the same concentration. In order to determine antibody concentrations, a sample of processed bile was probed with serial dilutions of primary antibody and linearity was shown by probing a serial dilution of bile. Films were scanned using a Personal Densitometer SI and analysed with ImageQuant software.

Table 4. Details of antibodies used in Western Blotting.

Antigen	Species/clone	Concentration	Source
MMP-9	mouse clone 56-2A4	1:1000 (1ug/ml)	Abcam, Cambridge UK
Lipocalin-2	mouse clone 5G5	1:80000 (12.5ng/ml)	Abcam, Cambridge UK
Complement 3a	mouse clone H13	1:1000(0.1ug/ml)	Abcam, Cambridge UK
PBEF	rabbit (polyclonal)	1:500	Bethyl Laboratories
Annexin A3	rabbit clone 693510	1ug/ml	R&D systems Abingdon UK
Lygdi	mouse clone 97A1015	1:4000	Abcam, Cambridge UK

3.0 The Utility of Bile in Biomarker Discovery

3.1 Introduction and Aims

3.1.1 Biliary Proteins

Bile has multiple functions including fat and mineral absorption and excretion of a wide range of metabolic breakdown products. Its chemical composition is complex and varies with nutritional status but comprises of in the main; bile salts (61%), fatty acids (12%), cholesterol (9%), phospholipids (3%), bilirubin (3%), inorganic salts and metals (5%) [190]. Proteins account for only 7% of bile and its profile has only recently been investigated in detail [128]. Protein concentration of bile obtained from the gallbladder ranges from 0.2 - 31 mg/ml compared to 0.34 – 13 mg/ml from the CBD [191-192].

The majority of protein content in bile is derived from plasma or secretion by the hepatobiliary system [190]. Under physiological conditions major fractions of serum proteins found in bile are thought to cross the tight junctions separating the sinusoids from the bile caniculi [193]. However, since the liver is responsible for the production of the large proportion of serum proteins, it can be difficult to determine whether a bile protein is derived from the serum crossing the tight junctions or active secretions from hepatocytes or cholangiocytes. Bile, like serum, is dominated by a small number of highly abundant proteins including albumin, α_2 macroglobulin, apolipoproteins, and ceruloplasmin [194-195]. Examples of other common proteins found in high abundance in serum and bile are shown in Table 5. Prior to the present study the largest catalogue of bile proteins to be derived from a proteomic approach in CCA was 87, in normal gallbladder bile was 222 and in pancreatic cancer was 127 [182], [185-186].

Table 5. Abundant proteins found in serum and bile.

Protein
Albumin
Orosomucoid
$\alpha 1$ – antitrypsin
$\alpha 2$ – macroglobulin
Apolipoprotein A-I, A-II, B, C-II, C-III
Ceruloplasmin
Haptoglobin
Hemoglobin
Immunoglobulin (IgG, IgM, IgA)
Transferrin
Mucin
Alkaline phosphatase
Alkaline phosphodiesterase I
β- galactosidase
β- glucoronidase

Since bile is stored, concentrated and transported in the biliary tract after being secreted by hepatocytes and cholangiocytes, it represents a "proximal" fluid to bile duct cancer. It potentially contains higher levels of marker proteins compared to those in serum, which is predominated by abundant proteins, limiting the ability to detect proteins of interest present in lower abundance. To investigate the potential utility of bile for biomarker discovery, a preliminary analysis was performed to detect if differences exist between bile protein profiles and serum, which confer an advantage to bile analysis and justify its more invasive collection methods.

Several commercially available products exist that can deplete abundant proteins to improve the profiling of proteins of lower abundance of potential biological relevance to disease. Examples include Cibacron Blue, Protein A or G, Multiple Affinity Removal System (MARS), Human-14 (Agilent Technologies) and ProteoPrep 20 (Sigma). Here we utilised an antibody affinity resin-based Albumin/IgG Removal Kit (GE Healthcare) commonly applied to serum, and optimised its use with bile. The impact on the 2D PAGE protein profiles was tested in bile samples from 4 patients with CCA to provide evidence for its application in subsequent bile biomarker discovery investigations.

3.1.2 Aims

In summary the aims of this part of the work were:

(a) To compare the 2D DIGE profiles of bile and serum from 4 patients with CCA

(b) To optimise an albumin/IgG depletion strategy for use with bile

(c) To analyse the effect of albumin/IgG depletion on the protein profile bile from 4 patients with CCA

3.2 Results

3.2.1 2D DIGE Profiles in Bile and Serum

Samples from 4 patients with CCA (8628, 8662, 9037, 9369) were analysed by 2D DIGE to compare the protein profiles of bile and serum. Serum was processed both in its 'neat' form and after TCA precipitation to simulate the sample processing of bile. The experimental workflow is illustrated in Figure 4. In brief, TCA precipitated protein was resuspended in DIGE lysis buffer and 50 µg of protein from each patient was labelled with Cy5. The incorporation of a pooled internal standard labelled with Cy3 that was run on every gel allowed for more accurate gel matching and quantitation. Samples were separated using pH 4-7 IPG strips for IEF and 5-18% gels for SDS PAGE.

Representative gels of bile and serum following TCA precipitation are shown in Figure 5. In this figure, major abundant serum proteins including albumin, serotransferrin, and haptoglobin, which are clearly present in bile gels, are indicated. However there are very significant differences in the protein profiles of bile and serum, which are also highlighted. The number of spots detected following visual confirmation and manual editing (deleted/merged/split/redrawn) in each bile and serum analytical gel ranged between 287 - 392 and 308 – 452 respectively.
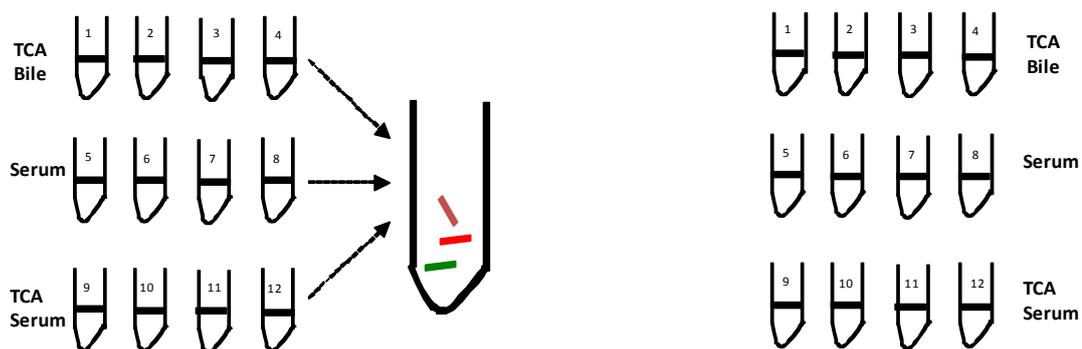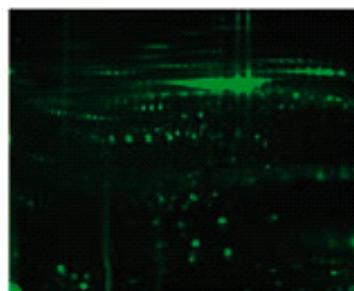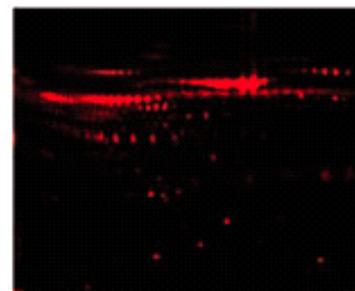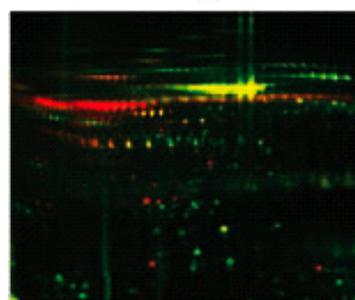
Figure 4. Summary of 2D DIGE experiment.

50 µg of pooled internal standard and individual sample were labeled with Cy 3 and Cy 5 respectively and separated on pH 4-7 IPG strips and 8-15% gradient gels. Number of gels = 12.
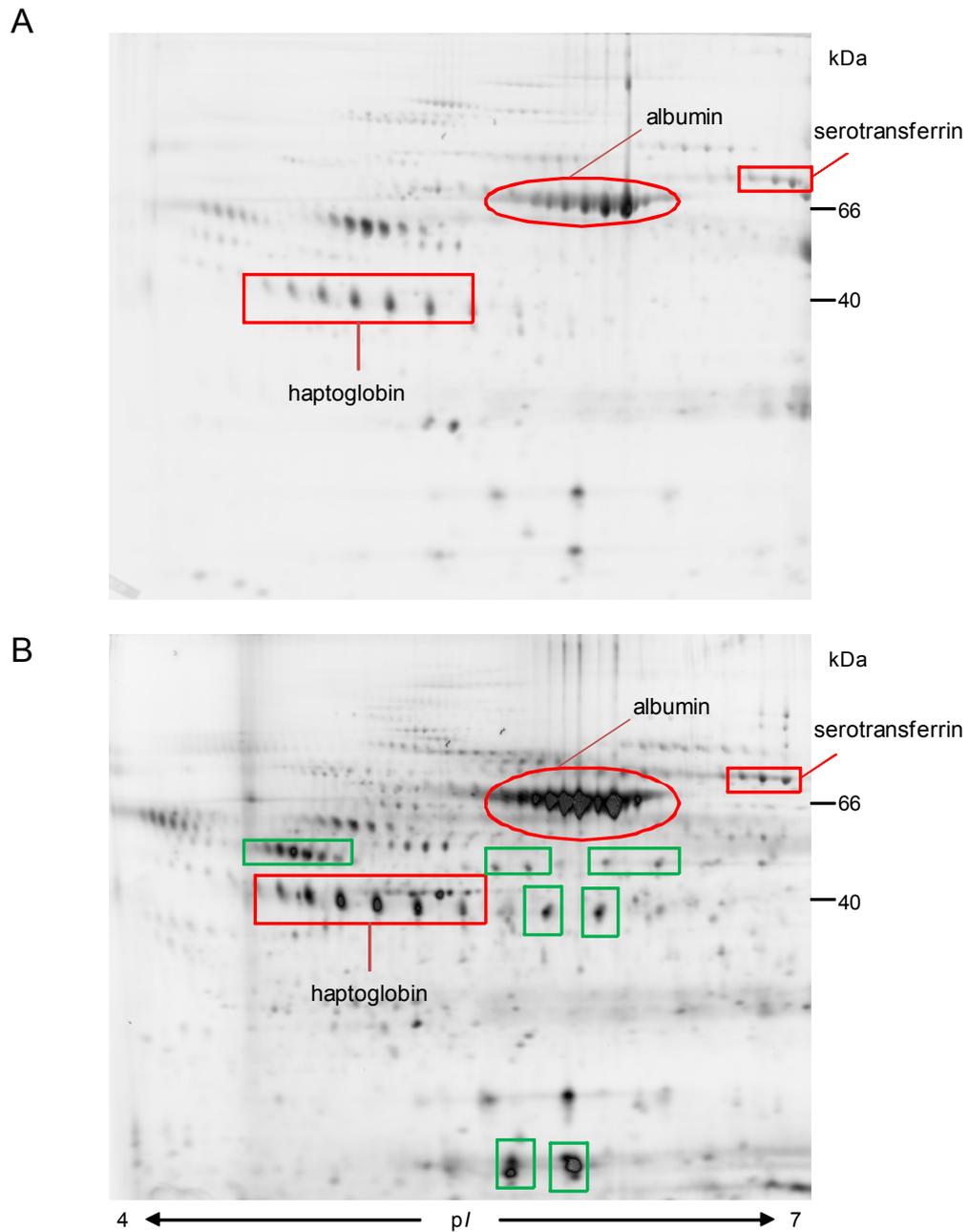
Figure 5. Comparison of bile vs. TCA serum in CCA.

Typical gel profiles following separation of 50µg of TCA precipitated serum (A) and bile (B) using 24 cm, pH 4-7 IPG strips and 8-15% gels for SDS PAGE. Major abundant proteins present in both gels are shown in red and areas of difference in bile in green.

An 'unedited' analysis of the gel images using Progenesis SameSpots™ software detected 1742 features of which 713 were found to have a statistically significant difference in abundance between bile and serum (p ≤ 0.05, ANOVA). Principal component analysis (PCA) was used to provide a visual representation of relationships within the data i.e. between samples or disease groups. The separation shown between bile and serum gels are derived from an unsupervised multivariate analysis in which the PCA accounts for 75.74% of the variation in the dataset (PC 1 = 68.56%; PC 2 = 7.18%, Figure 6). There was little separation across PC 2 indicating similarity of spot representation within the bile and serum gel sets. TCA precipitation of serum induced little difference compared to the untreated serum.
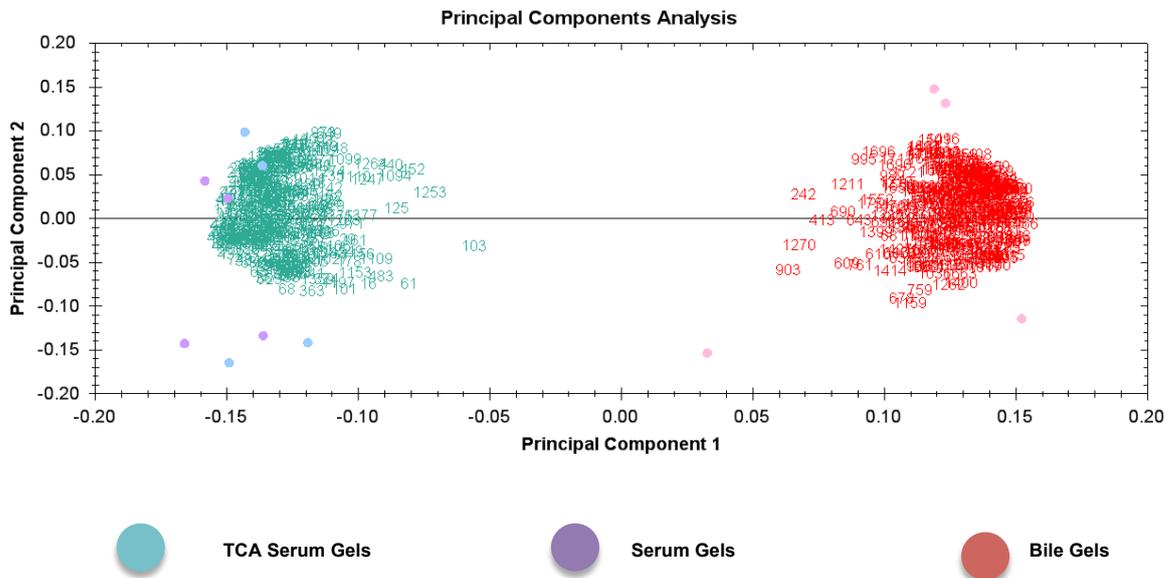
Figure 6. Principal Component Analysis (PCA) of Gel Profiles using Progenesis Same Spots Software.

An 'unedited' PCA of all 713 statistically significant spots (p ≤ 0.05) is shown. Clear separation of the gels was seen with serum gels clustering together away from the bile gels. Green indicates spots of decreased abundance in bile and red indicates spots of increased abundance in bile.

In Figures 7 and 8, representative examples of spots of increased and decreased abundance in bile and serum are illustrated.
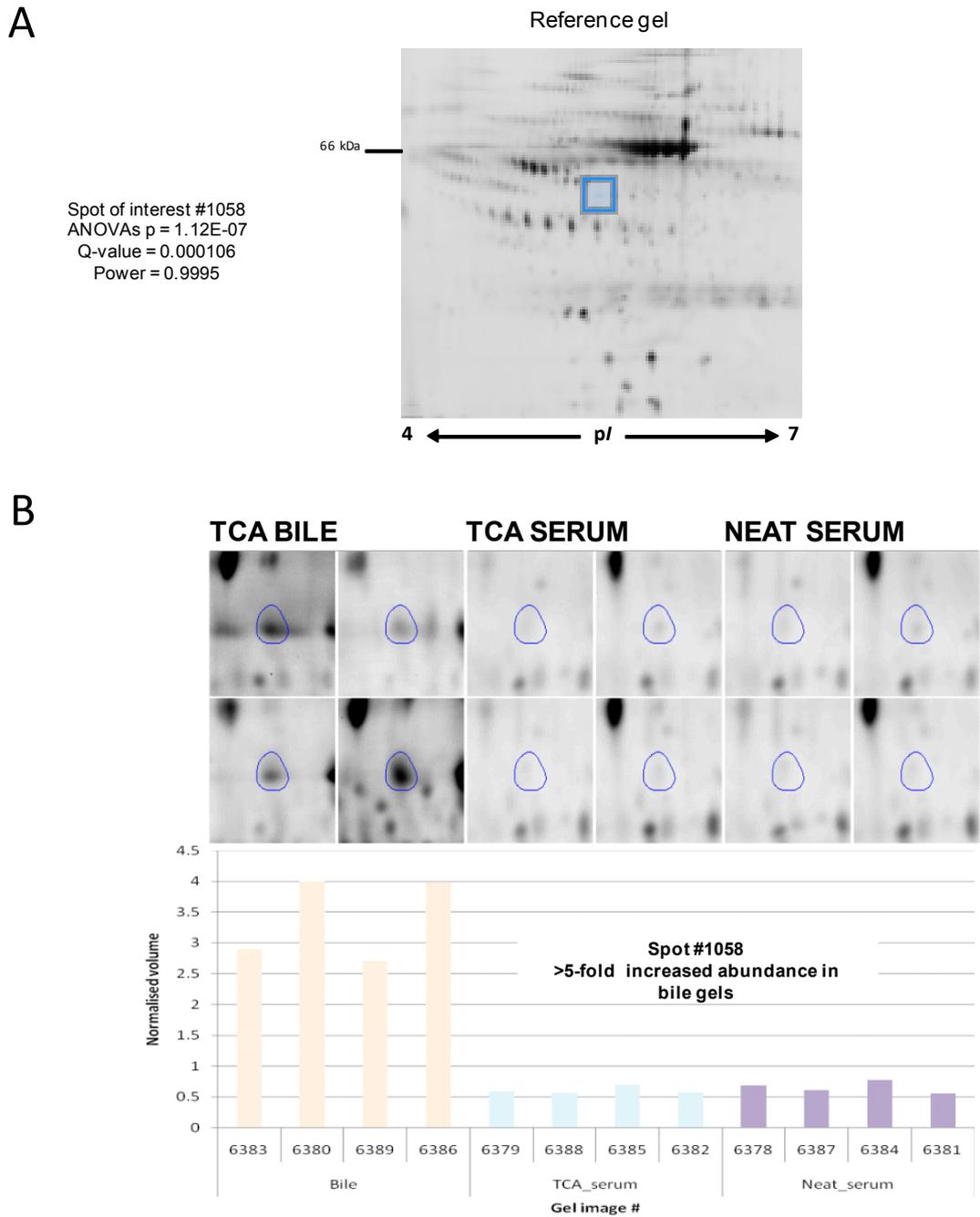
Figure 7. Protein spot 1058 and its relative abundance in bile and serum.

A. The position of spot of interest (1058) on the reference gel profile is indicated. **B**. Relative level of spot 1058 in the gels of serum and bile displaying higher abundance in bile gels compared to serum in the same patient with CCA.
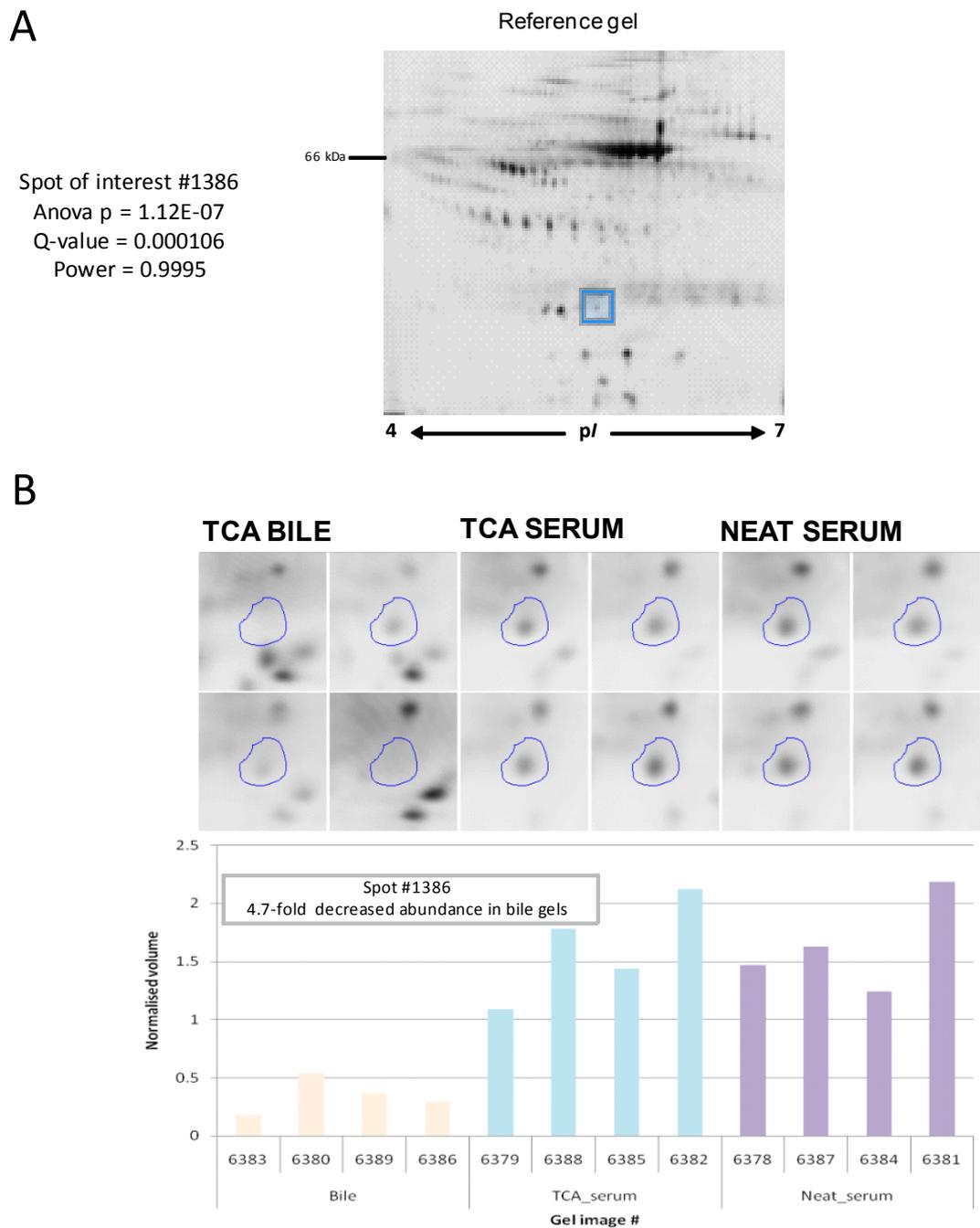
Figure 8. Protein spot 1386 and its relative abundance in bile and serum.

A. The position of spot of interest (1386) on the reference gel profile is indicated. B. Relative level of spot 1386 in the gels of serum and bile displaying higher abundance in serum gels compared to bile in the same patient with CCA.

3.2.2 Albumin and Immunoglobulin Depletion of Bile and 2D DIGE Profile to Facilitate Visualisation of Lower Abundance Proteins

The bile and serum proteomes contain proteins with a wide dynamic range and efforts to identify potential markers are facilitated by removal of major abundant proteins. Removal of albumin and immunoglobulin is one strategy used in serum studies in an effort to allow deeper proteome coverage but its application in bile is yet to be reported. The Albumin/IgG Removal Kit was optimised for application to bile. Given the difference in protein concentration between bile and serum, an assessment of the maximum concentrations of bile that could be used with the kit and the effectiveness in removing albumin and IgG was investigated. Increasing amounts of bile (9089 CCA - 2.7µg/µL) were subjected to depletion according to manufacture instructions utilising 750 µl of slurry (Figure 9). The maximum volume that can be applied to the depletion column including slurry is 870 µl. Successful depletion was achieved in a bile sample ranging from 60 - 120 µl (162 - 324 µg). Relative densitometry after depletion demonstrated a range of protein recovery of 48 -51%.
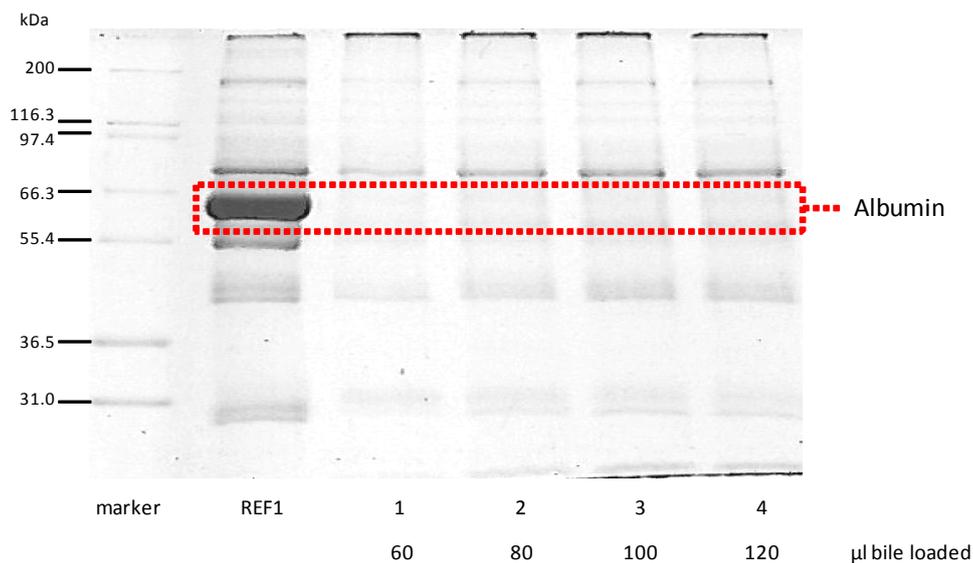
Figure 9. 1D PAGE of Bile Subjected to Albumin/IgG depletion.

Increasing amounts of bile (60, 80,100, 120 µl) were subjected to albumin/IgG depletion and compared to 10 µL of a non-depleted sample (Ref 1 - 9089 CCA). Samples were separated by SDS PAGE and visualised by Coomassie staining. The albumin band is observed to be depleted.

Given the effectiveness of depletion of bile samples containing up to 324 µg of protein in 750 µl of slurry, further experiments were undertaken to determine the maximum amount of protein that could be applied to a depletion column containing 500 µl of slurry (Figure 10). 1D PAGE illustrates successful depletion of albumin up to 540 µg (200 µl) of bile protein after which albumin reappeared. All subsequent experiments in the study utilised 500 µl of slurry with no more than 540 µg of bile protein. Effective depletion of heavy and light chain immunoglobulins was shown using 'mini' 2D PAGE gels (Figure 11).
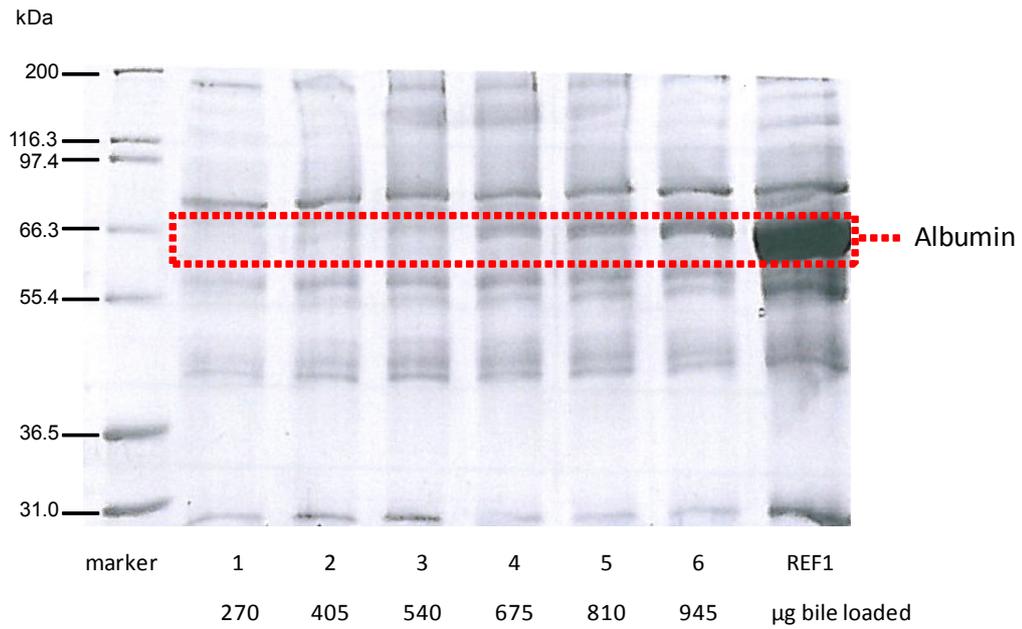
Figure 10. 1D PAGE of increasing amounts of bile subjected to albumin/IgG depletion.

1D PAGE of (REF1-9089 CCA) increasing amounts of bile subjected to albumin/IgG depletion using 500µl slurry. 10 µg of original bile sample and equivalent volume of depleted bile were separated by SDS PAGE and visualised by Coomassie staining. The predominant albumin band begins to appear with protein load > 540 µg.
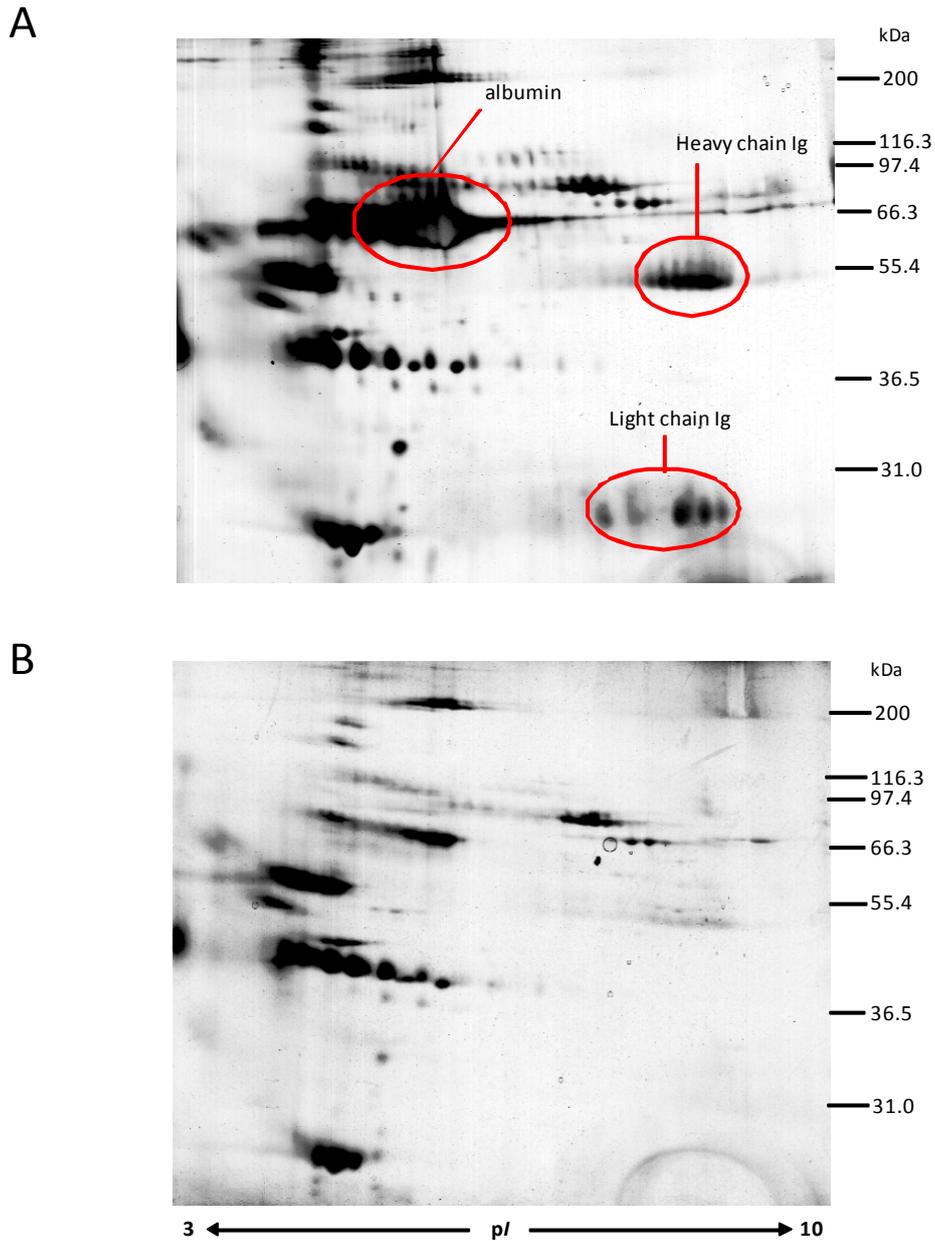
Figure 11. 2D PAGE of bile with and without albumin/IgG depletion.

5 μg of original sample (9089 – 2.7μg/μl) and equivalent volume of depleted bile were separated on to 7cm pH 3-10 IPG strips , separated by SDS PAGE and visualised by ProteoSilver™ Silver Stain. Albumin, heavy and light chain immunoglobulin shown in (**A**) are depleted in (**B**).

3.2.3 2D DIGE Profiles of Bile with and without Albumin/IgG Depletion

540 µg of bile samples from 4 patients (9369, 8397, 10312, 10319) were subjected to albumin/IgG depletion using 500 µl of slurry and 2D DIGE profiles compared to samples without depletion (Figure 12 A-D). Average protein recovery after depletion was 47.5% (45.2-49.2), consistent with optimisation results. On visual inspection, depletion of albumin was apparent with no gross overall change in appearance of the rest of the protein profile.

Progenesis SameSpots™ software detected 1434 features of which only 68 spots were found to change by greater than two-fold ($p \leq 0.05$, ANOVA) between gels with and without albumin/IgG depletion in an unsupervised multivariate analysis. In PCA of the entire profile little separation existed between samples of individual patients subjected to albumin/IgG depletion (Figure 13). However there was clear separation between different patients indicating that patient variability exerted a greater effect than depletion.
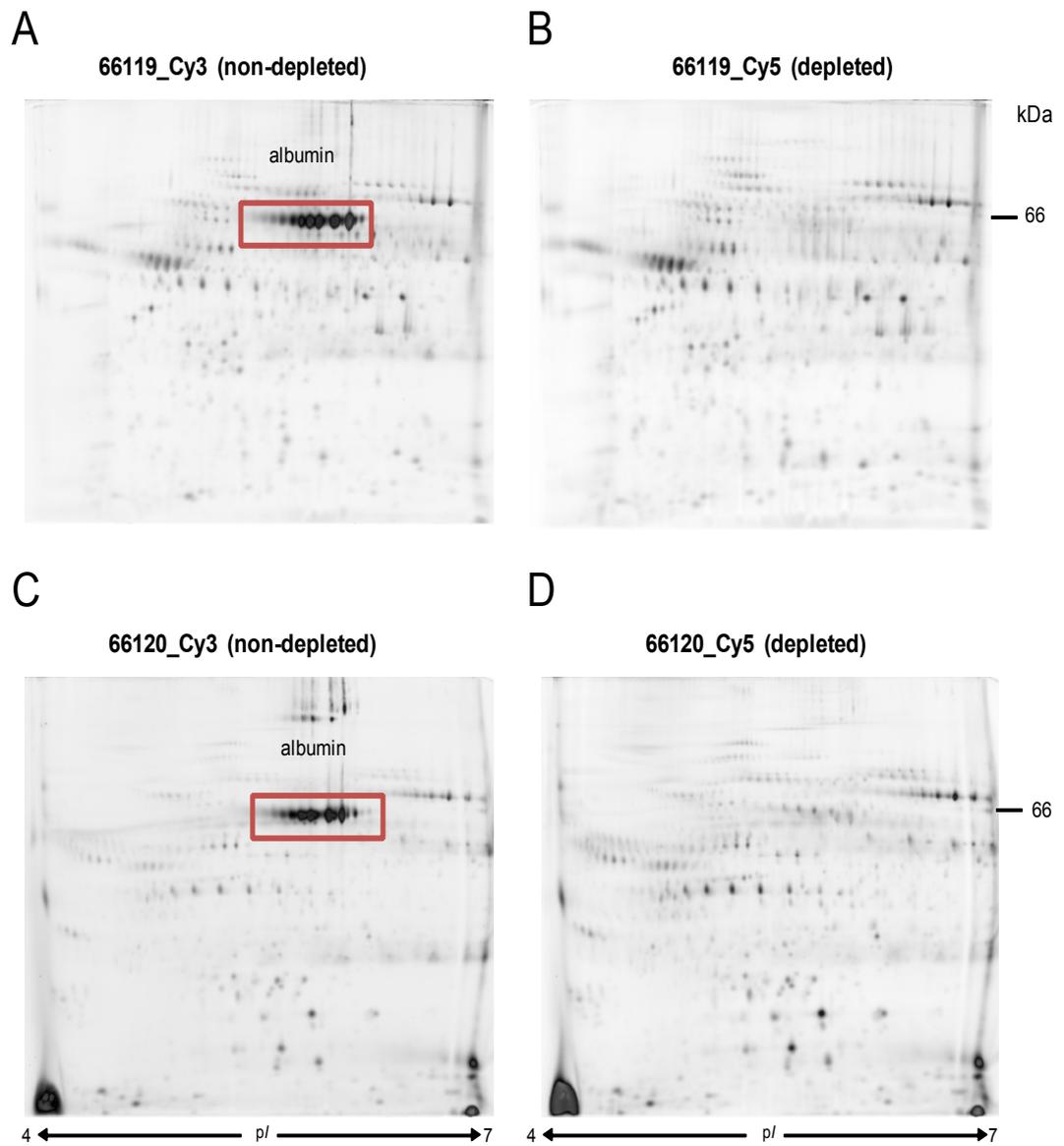
Figure 12 A - D. 2D gel profiles of bile with and without albumin/IgG depletion.

50 μg of bile protein (A) & (C) and albumin/IgG depleted bile (B) & (D) were labelled with Cy 3 and Cy 5 respectively and separated using pH4-7 IPG strips. Removal of albumin is clearly seen without significant change in the overall gel.
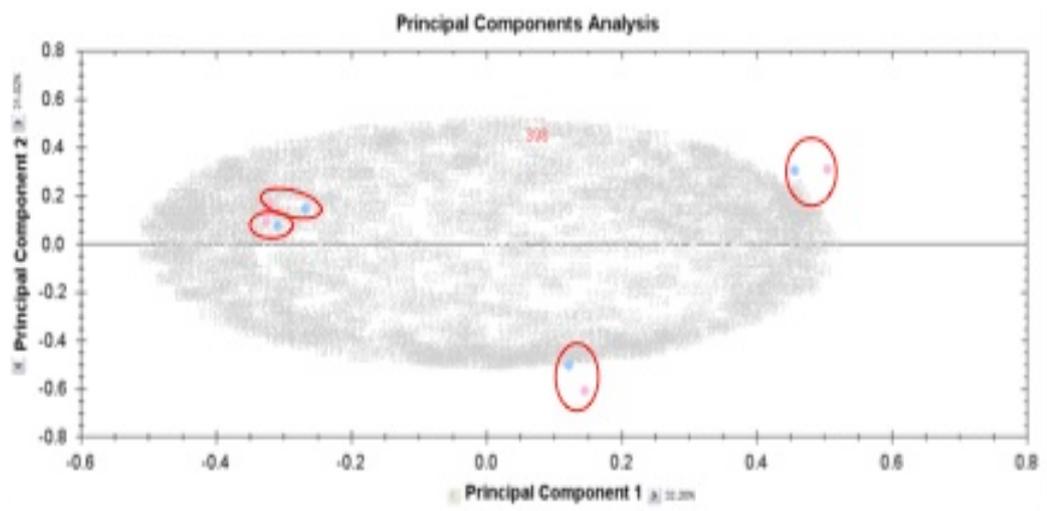
Figure 13.   Principal Component Analysis (PCA) of Gel Profiles with and without Albumin/IgG depletion using Progenesis Same Spots Software.

A PCA of 1434 detected features in the dataset of 4 patients with and without albumin/IgG depletion. The PCA accounts for 63.3% of the variation in the dataset (PC 1=32.2%; PC 2=31%). Individual patients are shown circled in red and highlight minimal change with sample treatment. Clear separation was observed between patients.

A preliminary correlation analysis between the normalised volumes of detected features in each patient before and after depletion provided further evidence that little change was conferred on the overall profiles. Results using the Spearman Rank correlation test for all patients are shown in Table 6 and a representative graph in Figure 14.

Patient	N	R^2	P value
1	1434	0.972	4.51e-264
2	1434	0.933	5.51e-274
3	1434	0.961	3.45e-290
4	1434	0.915	4.194e-264

Table 6 Spearman Rank Correlation Analysis.

Correlation analysis of normalised volumes of each feature in gels with and without albumin/IgG depletion. N= no of features, R= Spearman Rank correlation coefficient.

Normalised spot volumes
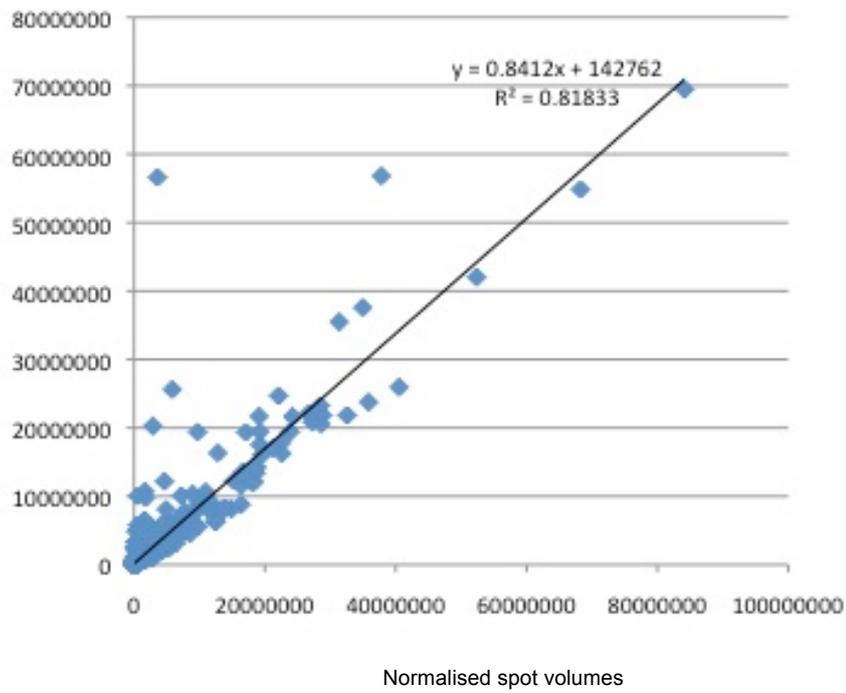
Figure 14. Correlation plot of normalized spot volumes.

A graph of normalised volumes for each feature detected in samples with and without albumin/IgG depleted bile.

3.3 Discussion

In this chapter experiments were undertaken to provide information regarding:

    (a) The difference in 2D DIGE profiles of patient-matched bile and serum

    (b) The application of an albumin/IgG depletion strategy to bile

    (c) The effect of this depletion on the overall protein profile

with a view to optimising a methodology for the indepth analysis of the CCA bile

proteome and a comparative analysis of bile in malignant and benign disease.

3.3.1 Performance of 2D DIGE in Bile and Serum

In preliminary work carried out prior to this study, a protocol for analysis of bile using proteomic platforms including 2D DIGE highlighted the requirement for removal of contaminants by techniques such as TCA precipitation to prevent interference with subsequent electrophoresis, facilitating better gel resolution and reproducibility. Indeed all bile/serum gel profiles in this study display good resolution and quality compared to recent published studies [196-198].

The results confirmed a difference in the profiles of bile and serum with differential expression of 713 potential spots highlighted in an unedited automated analysis and clear separation of groups in PCA. TCA treatment of serum induced little difference in profile compared to non-treated serum and discounts potential bias in comparing bile and serum in our study. While the proteins of increased abundance in bile were not identified by way of gel excision and MS formally in this series of experiments, the differentially displayed proteomes between bile and serum indicate potential utility of bile over serum to identify disease-associated biomarkers. To date no other investigations using DIGE labelling has been applied to proteomic investigation of bile

in CCA. Chen et al used 2D PAGE to compare proteins profiles from a patient with CCA and benign biliary disease and separated in 2D maps, with means of 182 and 176 spots on pH 4-7 IPG strips and detected 16 spots unique to the malignant group but did not incorporate Cy dye labelling [184]. While the identity of these proteins was not reported it does provide proof of concept in the ability to identify potential biomarker proteins. Indeed proteins of interest detected in bile could be taken forward in larger scale validation studies in serum requiring less invasive collection methods and facilitate more applicable clinical diagnostic tests.

3.3.2 An Analytical Challenge: Detecting Low Abundance Proteins

Albumin and immunoglobulins comprise around 80% of the total plasma proteome and the presence of a few high abundance proteins can effectively mask the detection of more relevant lower abundance proteins which presents a challenge in proteomic based approaches in biomarker discovery [199]. The depletion of highly abundant proteins is one effective method used to reduce the dynamic range of the proteome and shift the focus towards proteins of lower abundance with potential biomarker relevance. Attempts have ranged from the removal of albumin to more than 20 of the most abundant proteins accounting for more than 85-90% of the serum proteome. Cibacron Blue is a chlorotriazine dye with affinity for albumin but has been also shown to co-deplete other proteins [200]. Similarly Protein A and G which have affinity for immunoglobulin do not bind all of the subgroups. Comparative studies indicate that using antibody affinity ligands for HSA and IgG result in more specific depletion compared to the traditional Cibacron blue/Protein A or G depletion methods [201-202]. More developments in mono/polyclonal antibodies have improved the sensitivity of

depletion compared to these methods and in this study we elected to use an antibody affinity ligand based Albumin/IgG depletion kit [203].

Prior to this no protocol for application in bile had been reported and the manufacturer's protocol was established only for serum studies. Our results showed an effective and consistent depletion of albumin/IgG in bile and established a maximum value of 540 µg of bile protein that could be applied to a depletion column. The above process was repeated in further samples to ensure reproducibility. Relative densitometry after depletion demonstrated a range of protein recovery of between 48-51% using 1D PAGE.

Initially depletion was assessed by 1D PAGE where albumin is clearly visible. To check the effectiveness of IgG depletion, 2D PAGE was employed which showed shown effective depletion of both heavy and light IgG chains. No significant change was observed in a preliminary analysis of the rest of the profile. These initial experiments confirmed the ability of Albumin/IgG removal kit to be effectively applied to bile, provided limits regarding the amount of bile protein that could be applied and showed the method does not significantly increase the time required for sample preparation.

However despite the efficiency of the immunoaffinity-based depletion of albumin and IgG studies suggest that removal of more abundant proteins (6, 14, 20 and 60) is required to enhance the detection of very low abundance proteins [204-205]. An alternative and complimentary method to depletion strategies is to enrich low abundance proteins by the use of solid phase hexapeptides ligand libraries [206]. Indeed such an approach by Guerrier et al in bile identified 81 proteins not reported in previous proteomic studies [185] and in plasma of individuals with liver diseases allowed for the

identification of gelsolin as possible candidate biomarker for hepatitis B-associated liver cirrhosis [207].

3.3.3 2D DIGE Profiles of Bile with and without Albumin/IgG Depletion

Visual analysis of each of the large-format 2D DIGE bile gels with and without albumin/IgG depletion showed consistent depletion of albumin without significant change in the overall profile between the four samples.

An 'unedited' analysis using Progenesis SameSpots™ software detected 1434 features of which only 68 spots were nominally found to change by greater than two-fold between the gels. PCA of the entire profile showed little separation between samples of individual patients resulting from albumin/depletion but did highlight separation between different patients indicating that patient (biological) variability exerted a greater effect than depletion. Furthermore a correlation analysis between the normalised volumes between detected features in each patient before and after depletion provided further evidence that little change is conferred to the overall profiles in gels by the treatment (R^2=0.9).

Limitations of the 2D approach include investigation of only a subset of proteins in the proteome. Furthermore depletion strategies applied to biological samples can lead to potential for other non-specific protein loss – the so-called 'sponge effect'. Here small proteins and peptides can bind to larger carrier proteins that have been specifically depleted and result in losses of proteins of interest such as cytokines [208-209]. The magnitude of this non-targeted loss also does not take in to account potential losses

due to non-specific binding to depletion matrices and few qualitative studies exist in serum and bile.

Our evidence supports the use of albumin/IgG depletion in bile prior to proteomic investigation, allowing for removal of a significant proportion of the proteome comprised by highly abundant proteins without significantly altering the protein gel profiles. The question now is whether such an approach results in an increased detection of proteins not yet reported in the bile proteome and is in part the subject of the next chapter.

3.3.5 Conclusion

In conclusion, we have shown that despite the presence of highly abundant plasma proteins in bile, there are significant differences between the bile and serum proteomes supporting the use of bile as a proximal fluid in biomarker discovery. The dynamic protein range remains an issue and we have optimised a cost effective, time efficient and reproducible method for the depletion of two of the most abundant proteins in bile (albumin and IgG) without significant changes in the overall profile which has the potential to allow for the investigation of lower abundance of interest in biomarker discovery.

4.0 Characterisation of the Biliary Proteome in Hilar Cholangiocarcinoma

4.1. Introduction

Protein biomarker discovery in biliary disease is facilitated by the availability of comprehensive descriptions of the malignant bile proteome. A small number of studies using different strategies, including 2D gel based analysis and LC-MS/MS have started to attempt such characterisation, either in the form of gel mastermaps or catalogues of proteins, generating a body of data that may direct subsequent comparative studies using bile [128], [183]. However prior to this study only a limited number of biliary proteins had been identified.

In this chapter the complementary approaches of 2D PAGE and GeLC MS/MS were utilised to provide a bile protein mastermap and a catalogue of proteins in malignant bile from four patients with hilar CCA. Demographics and liver function tests for each of the patients are summarised in Table 7. Bile obtained from four patients undergoing diagnostic or therapeutic ERCP was subjected to albumin/IgG depletion and TCA precipitation as previously described.

Patient	Age/Sex	Bilirubin (µmol/l)	ALP (µmol/l)	ALT (µmol/l)	Liver Parencymal Status
1 9416	72 F	121	878	71	Mildly fatty liver, evidence of steatohepatitis
2 10938	64 F	76	513	89	Normal
3 8445	73 M	121	776	67	Moderate degree of steatosis
4 9089	72 F	88	591	59	Normal

Table 7. Patient demographics and liver function tests for each patient in bile protein catalogue.

4.2 2D PAGE Analytical Gels of Bile in Hilar CCA

Bile samples from 4 patients with hilar CCA were subjected to albumin/IgG depletion, labelled with Cy5 and protein separated by 2D PAGE as previously described. Examples of one analytical gel per patient are shown in Figure 15.
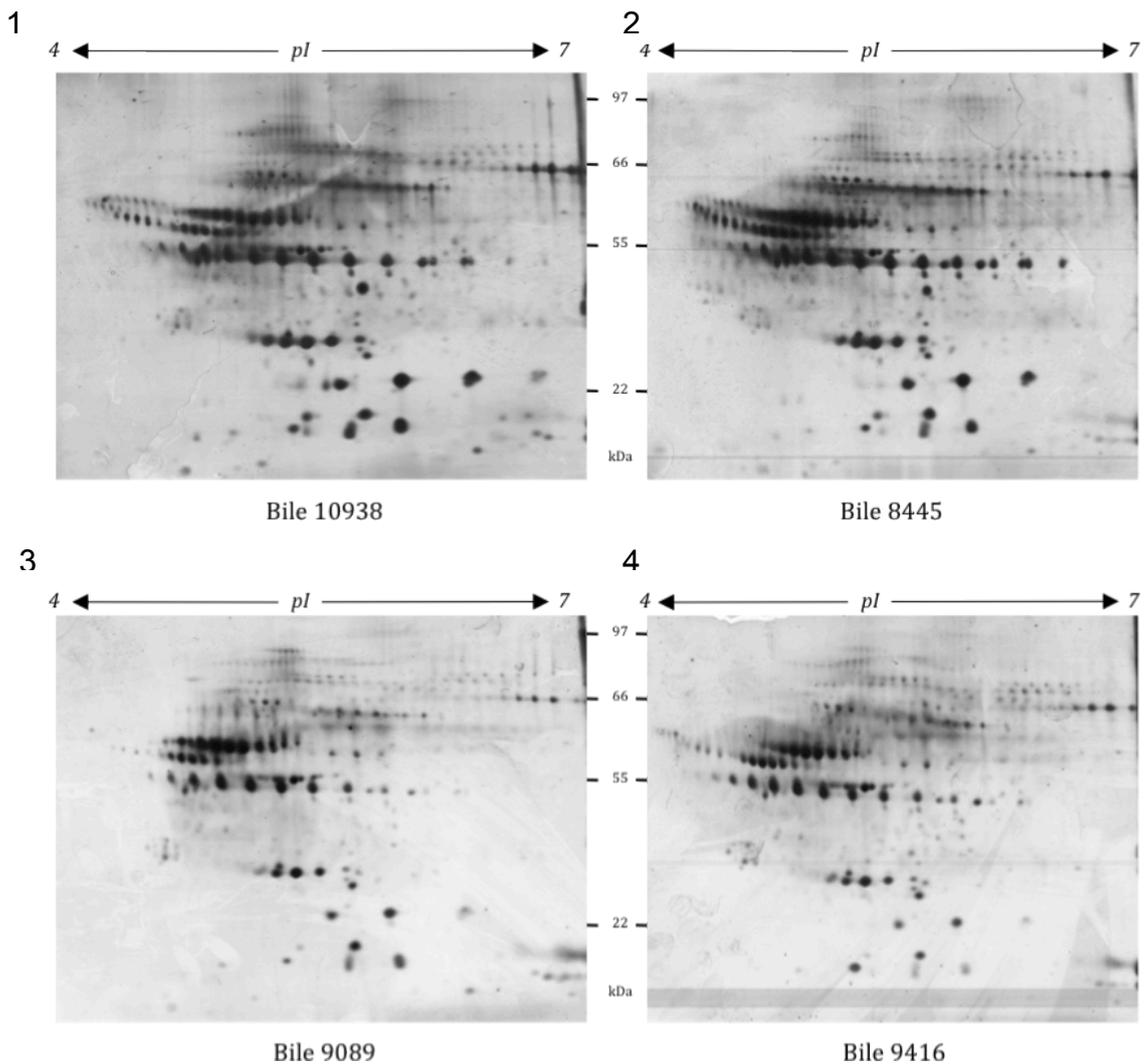
Figure 15. Analytical 2D PAGE gels in hilar CCA.

50µg of TCA precipitated, albumin/IgG depleted bile was labelled with Cy5 and separated by 2D PAGE with 24cm pH 4-7 IPG strips for IEF and 8-15% gradient gels for each patient 1-4.

On visual inspection there was notable similarity between the bile profiles in the group of patients with hilar CCA selected with similar age, biochemical parameters and histological grade. There were 748 spots common to all 4 patients. Using the Spearman Rank correlation test for pairwise comparison between patients, R^2 values of between 0.59 and 0.80 were obtained (Table 8).

Patient	N	Pairwise comparison between patients	R^2	P value
1	1336	1-2	0.80	2.31e-189
		1-3	0.61	1.08e-111
		1-4	0.69	1.34e-148
2	1395	2-3	0.59	1.44e-97
		2-4	0.64	1.67e-158
3	1205	3-4	0.68	2.01e-195
4	1317	-	-	-

Table 8 Spearman Rank Correlation Analysis

Correlation analysis of normalised volumes of each feature in gels in pairwise comparisons of gel profiles.

N= no of features, R= Spearman Rank correlation coefficient.

4.2.1 Bile Mastermap in Hilar CCA

To create a mastermap of bile from patients with hilar CCA a parallel preparative gels were generated using 500 µg of pooled bile protein from four patients of proteins visualised with silver staining and digital images analysed with Progenesis Same Spots. Spots of interest were imported to the Ettan Spot Picker V1.2. and excised directly into a 96 well plate. Tryptic digestion and MS analysis was performed as described in section 2.10.2 & 2.11.1.

4.3 Results

There were 632 distinct spots detected by SameSpots software overall. Of these 323 spots were suitable for selection by the Ettan Spot Picker V1.2 for gel excision, tryptic digestion and MS/MS analysis. Overall 254 spots (representing 80 unique proteins) gave valid identification based on two or more significant peptides (Table 9 and Figure 16). Multiple proteins were identified in the same spot in 10 cases. Of the 80 proteins listed, 72 were annotated for molecular and biological processes using the Panther Classification system (www.pantherdb.org). The most common molecular processes were as follows: 48% protein binding; 19% for enzyme regulatory activity; 17% catalytic activity and 14% structural activity. The biological processes observed were as follows: metabolic 19%, immune system 14%, cellular processes 14%; transport 13%, response to stimulus 10% and cell communication 8.6%.

Table 9. Proteins identified by MS.

Spot number, gene name, protein name, protein coverage and number of peptides are given. Where the same protein was found in other spots, these are included in smaller font. The spot numbers relate to the gel master map in Figure 16. N= number of spots.

N	Spot	Gene Name	Protein Name	Protein Coverage (%)	No of Peptides	
1	1158	Q4VJB6	14-3-3 protein	9.4	2	
2	2155	Q13707	ACTA2 protein	24.5	5	
3	2156	ACTA	Actin, aortic smooth muscle	18.6	6	
	932			18.6		4
	2352			15.9		4
4	2155	ACTB	Actin, cytoplasmic 1	37.1	15	
	932			25.3		5
	972			15.5		2
	1476			10.1		2
	877			12.8		3
	2156			25.3		9
	2351			22.7		5
	2352			22.7		6
5	2446	ACBP	Acyl-CoA-binding protein	32.2	2	
6	2344	APMAP	Adipocyte plasma membrane-associated protein	7.2	2	
7	2321	AACT	Alpha-1-antichymotrypsin	31.4	10	
	770			22.9		8
	775			14.7		4
	786			18.9		8
	789			7.1		2
	796			26.5		8
	840			20.6		7
	1752			27.9		8
	2074			16.5		6
	2075			21.5		7
	766			14.4		3
	2323			28.8		8
8	2293	A1AT	Alpha-1-antitrypsin	43.8	15	
	523			18.2		5
	738			18.4		5
	821			20.3		5
	868			24.2		3
	874			26.6		7
	877			29.7		9
	882			22.2		6
	888			26.8		5
	891			31.1		8
	1476			16.5		4
	2094			20.3		5
	2188			6.2		3
	2190			20.1		3
	2252			32.3		15
	2293			43.8		15
	2302			20.6		4
	2319			41.6		12
	2352			24.4		8
	2364			30.4		10
	2387			18.4		4
	2394			16.5		4
9	705	A1BG	Alpha-1B-glycoprotein	34.7	13	

	710			17.2		5
	2307			30.9		11
	2314			34.7		11
10	**840**	**FETUA**	**Alpha-2-HS-glycoprotein**	**13.9**	**5**	
	828			8.4		3
	861			8.4		3
11	**620**	**A2MG**	**Alpha-2-macroglobulin**	**12.1**	**9**	
	618			4.8		6
	574			5.6		5
	627			7.3		4
	631			3.9		3
	634			2.3		3
	2256			2.9		4
	2430			10.2		5
12	**1411**	**SYUA**	**Alpha-synuclein**	**35**	**3**	
13	**2445**	**ANXA1**	**Annexin A1**	**6.9**	**2**	
14	**1086**	**B2R9W6**	**Annexin A3**	**39**	**12**	
15	**1089**	**ANXA4**	**Annexin A4**	**23.8**	**6**	
16	**2183**	**ANT3**	**Antithrombin-III**	**22.8**	**6**	
	2293			15.5		5
17	**1252**	**APOA1**	**Apolipoprotein A-I**	**50.9**	**14**	
	1248			40.8		8
	1294			40.8		6
	1410			29.2		4
	1553			17.2		3
	1566			13.9		3
	1570			14.2		2
	2137			23.6		6
	2472			28.8		6
18	**1294**	**APOA4**	**Apolipoprotein A-IV**	**5.6**	**2**	
	2351			10.1		4
19	**1570**	**B0YIW2**	**Apolipoprotein C-III variant 1**	**23.1**	**2**	
20	**1074**	**APOE**	**Apolipoprotein E**	**7.6**	**2**	
21	**929**	**ACTBM**	**Beta-actin-like protein 3**	**7.7**	**3**	
22	**1222**	**CRP**	**C-reactive protein**	**13.8**	**3**	
23	**1752**	**CALR**	**Calreticulin**	**6.2**	**2**	
24	**2314**	**CASPE**	**Caspase-14**	**9.5**	**2**	
25	**2436**	**B4DPP6**	**Serum albumin**	**32.2**	**17**	
	748			12.5		5
	1323			9.7		4
	761			19.6		10
	2437			25.9		8
	753			18.4		8
	754			17.1		9
	756			22.3		10
26	**2183**	**B4DPP2**	**Vitamin D-binding protein**	**15.6**	**5**	
	2188			12.5		3
	2273			15.6		4
27	**1476**	**B2R4M6**	**S100A9**	**44.7**	**2**	
	1498			44.7		3
	2038			51.8		5
28	**948**	**CO3**	**Complement C3**	**10.9**	**16**	
	941			3.1		4
	992			5.1		7
	2003			5.9		6
29	**729**	**CO9**	**Complement C9**	**14.7**	**7**	
	751			10.6		4
30	**579**	**B0QZQ6**	**Complement factor B**	**17.9**	**7**	
	589			17.7		11
	595			13.9		7
	596			16.2		9
	2429			13.1		4
31	**1751**	**CYTA**	**Cystatin-A**	**24.5**	**2**	
32	**2314**	**DESP**	**Desmoplakin**	**2.4**	**5**	
33	**1859**	**LACRT**	**Extracellular glycoprotein lacritin**	**27.5**	**2**	

34	**975**	**FIBB**	**Fibrinogen beta chain**	**50.7**	**29**	
	849			16.7		5
	887			10.2		3
	968			28.5		9
	971			24.6		8
	972			41.5		19
	841			8.8		29
	1669			49.9		27
	1670			41.3		14
	2335			36.5		10
35	**1999**	**FIBG**	**Fibrinogen gamma chain**	**50.6**	**18**	
	574			32.5		11
	618			10.4		5
	628			21.6		7
	658			21		9
	665			30.5		13
	667			9.3		3
	877			37.7		17
	919			14.1		5
	1689			31.8		12
	1723			19.9		6
	1743			17.4		7
	1746			21		9
	1774			23		7
	1775			11.5		5
	1962			35.8		16
	572			32.5		11
	2001			35.1		14
	2254			33.6		14
	2430			21		7
36	**1907**	**GELS**	**Gelsolin**	**4.2**	**3**	
37	**1146**	**GSTO1**	**Glutathione S-transferaseomega 1**	**9.5**	**2**	
38	**972**	**HPT**	**Haptoglobin**	**38.2**	**13**	
	923			24.1		8
	929			29.6		11
	932			20.2		5
	958			29.6		12
	965			29.6		9
	968			27.3		9
	971			24.9		8
	911			20		5
	980			16		5
	986			27.1		7
	987			33.5		10
	993			26.8		6
	998			28.8		9
	1010			21.2		5
	1323			15.8		5
	1344			11.8		5
	1351			22.9		13
	1357			11.3		5
	1381			13.3		7
	1388			15.5		4
	1498			14.5		6
	1670			29.6		10
	2275			26.6		10
	2302			29.8		13
	2339			26.8		4
	2387			31.3		9
39	**1419**	**HPTR**	**Haptoglobin-related protein**	**6.6**	**5**	
	2121			6.3		2
40	**745**	**HSP7C**	**Heat shock cognate 71 kDa**	**21.5**	**15**	
41	**1570**	**Q9BX83**	**Hemoglobin alpha 1 globin chain**	**24**		

42	1388	HBB	Hemoglobin subunit beta	55.1	4	
	1351			40.8		3
	1417			49		5
	1482			40.8		4
	1489			49		5
	1579			32.7		3
43	2360	HEMO	Hemopexin	33.1	9	
	736			13.9		4
	739			9.7		3
	911			20		5
	2434			28.1		8
	2436			15.2		5
44	971	Q6PEJ8	HP protein	40.8	7	
45	738	IGHA1	Ig alpha-1 chain C region	19.8	5	
	790			17.8		5
	791			13.6		4
	793			13		4
	800			17.8		5
	802			17.8		5
	803			13		3
	809			13		4
	877			10.8		4
46	2162	IGKC	Ig kappa chain C region	34.9	2	
47	685	IGHM	Ig mu chain C region	13.7	5	
	689			13.7		4
	2403			10.4		4
48	1262	Q6P5S8	IGK@ protein	23.7	2	
49	2107	Q0KKI6	Immunoblobulin light chain	31.1	4	
50	1672	IGJ	Immunoglobulin J chain	48.2	3	
51	904	A2GL	Leucine-rich alpha-2-glycoprotein	4.9	2	
52	1674	ILEU	Leukocyte elastase inhibitor	37.5	10	
	940			26.6		7
53	1859	LCN1	Lipocalin-1	21	3	
	965			12.5		2
	745			12.5		2
54	2240	REG1A	Lithostathine-1-alpha	12.7	2	
55	1859	SG2A1	Mammaglobin-B	21.1	2	
56	1624	Q5J7W2	Migration-inducing gene 9 protein	26.1	2	
57	1488	MYL6	Myosin light polypeptide 6	43.7	5	
58	1411	ML12A	Myosin regulatory light chain 12A	29.8	6	
59	1344	NDKA	Nucleoside diphosphate kinase A	52.6	8	
60	1410	PRDX2	Peroxiredoxin-2	44.4	7	
	1294			19.2		3
61	1219	PRDX6	Peroxiredoxin-6	30.8	5	
62	779	PLSL	Plastin-2	15.5	7	
	877			3		2
63	1388	PACAP	Proapoptotic caspase adapter protein	13.2	2	
64	2447	B2R8F6	Proteasome subunit alpha type	25.7	6	
65	1561	S10AB	Protein S100-A11	27.6	3	
66	1602	S10A6	Protein S100-A6	16.7	2	
67	1559	S10A8	Protein S100-A8	38.7	3	
68	905	GDIB	Rab GDP dissociation inhibitor β	15.7	4	
69	1388	RET4	Retinol-binding protein 4	25.9	5	
	1294			15.4		2
70	1417	GDIR2	Rho GDP-dissociation inhibitor 2	23.4	4	
71	2403	TRFE	Serotransferrin	41.5	25	
	691			29.4		16
	2151			36.4		18
	2283			22.8		13
	2331			36.5		21
	685			20.5		8
72	2434	ALBU	Serum albumin	25.5	12	
73	1100	SAMP	Serum amyloid P- comp	23.8	5	

74	1474	TTHY	Transthyretin	63.3	5	
	1476			48.3		4
75	1498	TBCA	Tubulin-specific chaperone A	18.5	3	
76	2448	A8K1N0	Tyrosine 3-monooxygenase	22.9	4	
77	2314	UBIQ	Ubiquitin	44.7	2	
78	1411	B0YJC4	Vimentin variant 3	4.9	2	
79	2187	Q6GTG1	Vitamin D binding protein	39.5	28	
	1498			15.2		12
	2319			16.9		6
80	941	ZA2G	Zinc-alpha-2-glycoprotein	29.5	5	
	2003			13.2		2

Figure 17. Mastermap of bile in hilar CCA.

500 μg pooled bile from 4 patients was separated by 2D PAGE using 24 cm pH 4-7 IPG strips and 10% gels and protein detected by silver staining. The gel was analysed using SameSpots™. Spot numbers refer to identities reported in Table 7.

4.3.1 GeLC-MS-MS 'Shotgun' Proteomic approach to bile in Hilar CCA

4.3.1.1 Sample Processing and Protein Identification

In parallel ,protein (100ug) depleted of albumin/IgG and subjected to TCA precipitation from each bile sample was separated on 10% SDS PAGE using the Hoefer SE600X Chroma Deluxe Electrophoresis Unit and stained with InstantBlue Coomassie stain. Each lane was divided into 53 gel slices (Figure 17) and subjected to tryptic digestion and LC-MS/MS.
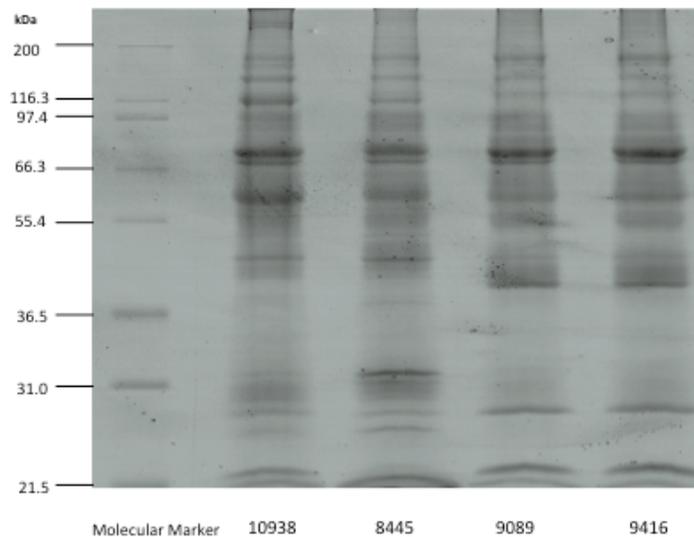
Figure 17. GeLC-MS/MS analysis of bile.

4 bile samples from patients with hilar CCA were processed by TCA precipitation and Albumin/IgG depletion. 100 µg of protein was loaded in each lane, analysed on a 10% SDS PAGE gel and visualized with InstantBlue Coomassie stain.

For each patient, the intermediate files produces by Mascot were combined, peptides with probability scores p>0.05 excluded and any redundancy removed. Proteins required at least one unique significant peptide to be considered as identified. The false discovery rate was determined as 1.8% by searching against a decoy database. For proteins identified by a single peptide, spectra were inspected manually to ensure quality control before inclusion. Full details of the proteins identified and including mass spectra for single peptide based identifications are provided as Supplementary Data Tables 1 & 2 respectively.

4.3.1.2 Results

Overall 3576, 2126, 2162 and 2298 peptides corresponding to 741, 509, 378, and 389 proteins were identified in the four bile samples from patients with hilar CCA of which 491, 350, 268, and 295 proteins respectively were identified by at least 2 significant peptides. Following the removal of keratins and immunoglobulins, redundancy was further reduced by taking the conservative approach of collapsing the data set down to the gene level (that is, removing the potential complexity arising from the presence of different or multiple forms of a particular protein in a particular bile sample). For protein entries with no gene, peptides were searched against MSDB (proteomics.leeds.ac.uk) to find this missing data; a small number of entries for which this was not possible were not considered in downstream analysis. Following this data reduction, products from a total of 813 unique genes were identified in the four patients, with 185 being present in 4, 83 in 3 and 152 in 2 out of the 4 samples and 393 being unique to a single patient. The degree of overlap is shown in Figure 18 and identities of all proteins are provided in Supplementary Table 3.
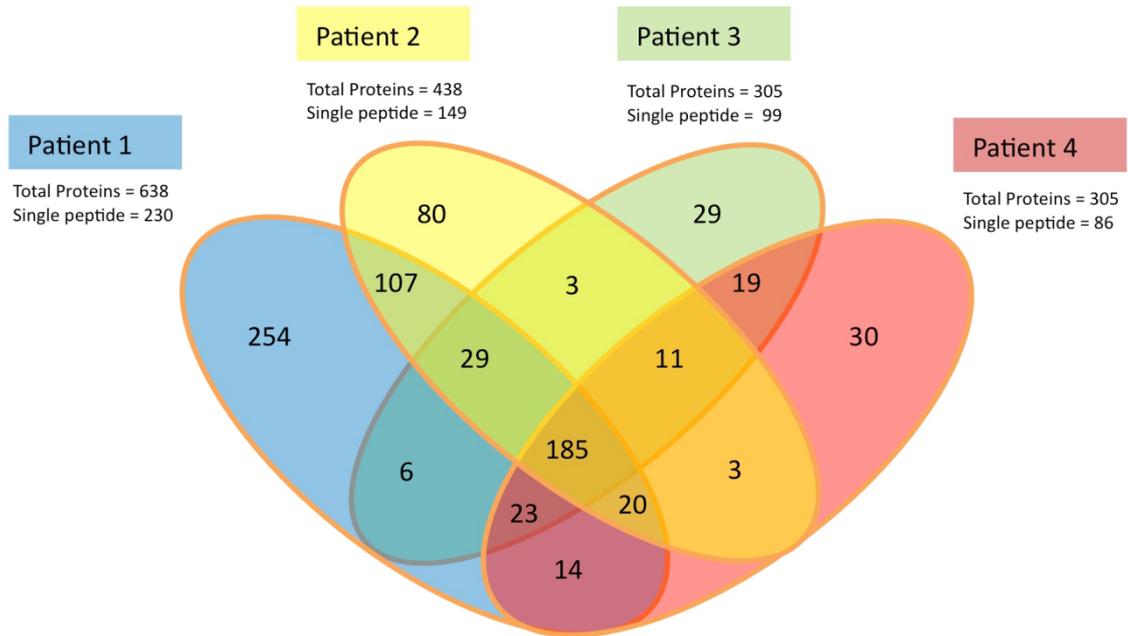
Figure 18: Venn diagram illustrating the number of proteins in each patient and the degree of overlap between individual patients. A total number of 813 unique proteins were identified.

Of the 185 proteins common to all patients, 21 specific proteins have been previously reported in the literature as being differentially expressed in CCA, which together with sub-cellular location and function, are highlighted in Table 10. The proteins identified in all 4 patients are listed in Supplementary Table 4.

Table 10 – Proteins associated with CCA and found in all four patients

Gene Name	Protein	Location	Functions	Up/ down in CCA
ANXA1	Annexin A1	extracellular	Receptor/ligand binding	up
ANXA2	Annexin A2	extracellular	Receptor/ligand binding	up
ANXA 4	Annexin A4	extracellular	Receptor/ligand binding	up
DMBT	Deleted in malignant brain tumours 1	plasma membrane	scavenger receptor activity	up
HSP90AA1	Heat shock protein HSP 90-alpha	cytoplasm	nucleotide/protein binding	up
MUC1	Mucin-1	plasma membrane	pathogen binding, cell signalling	up
LIPOCALIN-2	Neutrophil gelatinase-associated lipocalin	extracellular	ligand binding	up
PGK1	Phosphoglycerate kinase 1	cytoplasm	nucleotide binding, phosphorylation	up
PKM2	Pyruvate kinase M2	nucleus	nucleotide binding, apoptosis	up
VIM	vimentin	cytoplasm	cytoskeleton, apoptosis	up
ENO1	enolase	cytoplasm	glycolysis, growth, hypoxia tolerance	up
MMP9	MMP 9	extracellular	proteolysis extracellular matrix, leucocyte migration	up
GSTP1	Glutathione S-transferase P	cytoplasm	transferase	up
APOAII	Apolipoprotein A-II	secreted	metabolism	up
A1BG	alpha-1-B glycoprotein	extracellular	unknown function	down
AMBP	alpha-1-microglobulin	extracellular	transporter activity; calcium channel inhibitor activity	down
C1S	complement component 1 S	extracellular	serine-type endopeptidase activity	down
FGA	Fibrinogen alpha chain	extracellular	Receptor/protein binding	down
GC	group-specific component (vitamin D binding protein)	extracellular	Actin/vit D binding	down
HPX	HPX Hemopexin	extracellular	heme transporter activity	down
KNG1	Kininogen-1	extracellular	peptidase inhibitor activity/receptor binding	down

The gene products in the entire dataset were analysed using a combination of the Gene Ontology database version 7.0 (www.pantherdb.org) and Ingenuity Pathway Analysis. Cellular component analysis identified the majority of proteins as locating to cytoplasm (55%), extracellular (20%), membrane (10%) and nucleus (8%). The terms molecular function and biological processes were explored and in total 772 out of the 813 unique proteins were annotated. Proteins involved in catalytic activity, followed by protein binding and structural activity were the most common molecular functions (Figure 19). Specifically, hydrolase proteins made up more than 50% of the catalytic proteins and the serine-type peptidases and metelloproteinases account for 50% and 30% of them respectively. Calmodulin, cytosleleton, receptor and DNA/RNA binding proteins accounted for more 65% of the binding proteins.

Focussing on biological processes, the proteins in our dataset in order of frequency relate to metabolism, cellular processes and cellular communication. All other biologic processes are shown in Figure 19. Further interrogation of the gene ontology for metabolism identifies more than 90% of proteins relating to cellular and cell – matrix adhesion proteins. It is important to highlight that a number of proteins listed in Figure 19 occur in one or more categories.

Molecular Function

Function/ activity	Proteins in dataset	Proteins in all 4 patients
1. Ion channel	11 (1%)	1 (1%)
2. Transport	24 (3%)	9 (5%)
3. Translation	15 (2%)	0
4. Transcription	18 (2%)	0
5. Enzyme regulation	67 (8%)	28 (14%)
6. Catalytic	370 (45%)	70 (36%)
7. Motor	6 (1%)	2 (1%)
8. Receptor	31 (4%)	13 (7%)
9. Antioxidant	11 (1%)	6 (3%)
10. Structural molecular	86 (11%)	21 (11%)
11. Binding	245 (30%)	54 (28%)

Biologic Process

Process	Proteins in dataset	Proteins in all 4 patients
1. Cell comm	131 (16%)	37 (19%)
2. Cellular	222 (27%)	58 (30%)
3. Transport	121 (15%)	33 (17%)
4. Cell organisation	63 (8%)	18 (9%)
5. Apoptosis	31 (4%)	4 (2%)
6. Systems	36 (4%)	16 (8%)
7. Response to stimulus	112 (14%)	54 (28%)
8. Development	83 (10%)	22 (11%)
9. Metabolic	479 (58%)	103 (53%)
10. Cell cycle	47 (6%)	9 (5%)
11. Immune	163 (20%)	71 (38%)
12. Cell adhesion	46 (6%)	13 (6%)

Figure 19. Gene ontology classification.

Classification of molecular function (upper panel) and biologic process (lower panel) determined using gene ontology is shown for entire data set and proteins common to all four samples.

4.4 Discussion

A number of studies to date have started to characterise the bile proteome, generating a body of data that will underpin future studies using bile in comparative analyses to look for disease biomarkers. In this chapter the complementary approaches of 2D PAGE and shotgun MS were used to compile a comprehensive compendium of bile proteins specifically in CCA.

4.4.1 2D PAGE Bile Mastermap in Hilar CCA

To date no 2D PAGE protein master map of bile in CCA has been reported. We have produced a protein master map of high resolution of 80 unique proteins, the vast majority of which represented abundant serum proteins. Gene ontology classification identified the majority of proteins relating to molecular processes involving protein binding, enzyme regulation, catalytic and structural activity. The biological processes range equally across metabolic, immune system, transport, response to stimulus and cell communication.

An important advantage of 2D PAGE is the ability to visualise up to 2000 intact proteins and enable the resolution of isoforms and post-translational modifications of proteins that would be lost with shotgun strategies. In bile, as with other biological fluids, fewer proteins were visualised probably reflecting the presence of a small number of highly abundant proteins which dominate the profile. Although there were more than 600 hundred protein spots visualised in our gels less than half were amenable to automated robotic spot cutting. Furthermore several factors may explain the failure to obtain identifications for some spots cut from preparative gels. Spots may not contain sufficient protein to generate adequate mass spectra for identification and there may

have been failure of tryptic digestion. Furthermore keratin contamination during sampling handing and tryptic digestion can mask the identity of lower abundance proteins undergoing MS analysis. Finally the human proteome has yet to be fully described and some of the proteins may not be present in the current databases. The high throughput analysis of proteomes using this technique remains challenging as observed for the time consuming process of spot picking (even if automated) and subsequent spot digestion and analysis of each spot from a 2DE gel. Other limitations of this approach are the relative inability to detect lower abundance and proteins of extreme pI and hydrophobic proteins [210].

Nevertheless 2D PAGE approaches have been applied to the analysis of bile in disease. Chen et al compared 2D PAGE profiles between bile from malignant (CCA) and non-malignant disease and identified 16 differentially expressed protein spots in cancer [184]. However the identities of these protein spots were not reported. Zhou et al used a complementary approach of 2D PAGE and shotgun proteomics without depletion strategies to study bile obtained from gallbladders of patients undergoing surgery for gallstones [183]. Given bile protein is more concentrated in the gallbladder than from the bile duct it is of interest to note that the number of proteins identified by their study was 48 (from 106 spots excised) compared to 80 (from 323 spots) in this study. Both studies reported abundant serum proteins such as albumin, transferrin, immunoglobulins, and complement. These studies and results from our experiments support the feasibility and scope for development of this proteomic approach in bile biomarker discovery.

4.4.2 GeLC/MS/MS – 'Shotgun' Proteomic Analysis of Bile in Hilar CCA

Prior to this work, the first large scale effort to generate a bile catalogue was undertaken in 2004 in an unfractionated sample of bile taken at ERCP from one CCA patient and analysed by GeLC-M/MS, which identified 59 unique proteins [182]. A further 28 proteins were identified after prefractionation incorporating lectin affinity chromatography using Con A and WGA. Three proteins not previously related to CCA were focussed on: MAC-2 binding protein, lipocalin-2 and deleted in malignant brain tumours 1 protein.

An alternative approach using hexapeptide libraries to concentrate and reduce low and high abundance proteins respectively in pooled gallbladder bile by Guerrier et al, identified 222 gene products, of which 143 had not been previously reported [185]. In another study examining biliary stenosis caused by periampullary tumour, GeLC-MS/MS of bile obtained at ERCP identified 127 proteins (of which 34 were novel) based on more than one tryptic peptide; of these more than 80% were intracellular, most likely reflecting that this study profiled both the pellet and supernatant generated by centrifugation of bile [186].

The small number of studies to date attempting to extend the biliary proteome have incorporated the analysis of mainly single patients, utilised different sampling techniques and processing protocols. Furthermore investigations have sampled bile from with a variety of pathologies and explain some of the differences in the proteins identified. Overall using both 2D PAGE and shotgun approaches, there have been <300 proteins identified in the four major bile-profiling studies published prior to ours and these are summarised in Figure 20.
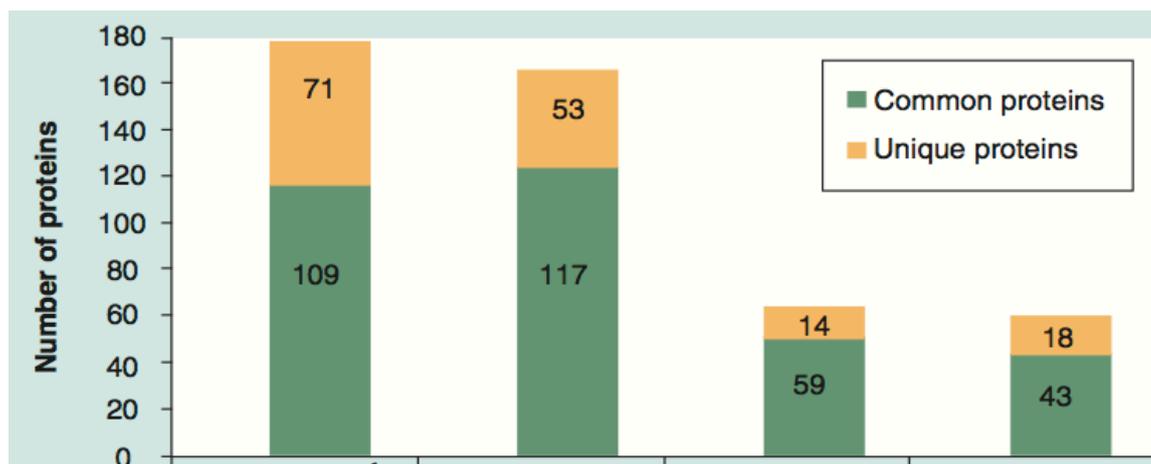
Author	Farina et al 2009	Guerrier et al 2007	Kristiansen et al 2004	Zhou et al 2005
Bile source	ERCP	Cholecystectomy	ERCP	Cholecystectomy
Disease	Pancreatic Ca	Normal bile	CCA	Gallstones
Prefractionation	Immunodepletion	No immunodepletion	No immunodepletion	No immunodepletion
Proteomic approach	SDS PAGE LC/MS/MS	SDS PAGE LC/MS/MS SELDI TOF/MS	SDS PAGE LC/MS/MS	2DE/2D LC.MS/MS
No of proteins identified with > 2 peptides	180	170	73	61

Figure 20. Summary of number of proteins reported in bile to date from proteomic analysis. In green are proteins found in common in all studies and in yellow proteins not identified in other studies

Overall 813 unique proteins were identified in this study after removing keratins immunoglobulins and albumin and reducing redundancy further by collapsing the data down to gene level. As expected there was considerable overlap between previous studies and the proteins identified in our dataset. 28 proteins from other studies were not identified witin the 813 proteins compiled in our catalogue. This can arise from the variation in the analytical methodology, sample collection/preparation, and the different disease categories.

Proteins identified were largely of cytoplasmic origin, involved in catalytic and structural molecular activity and involved with metabolic and cellular communication activity. A significant proportion of our proteins and those from other studies are known major abundant plasma proteins but include a number of known CCA associated proteins as summarised in Table 10 and confirm the utility of analysis of proximal fluids such as bile.

There was significant variability between the bile samples analysed here, probably reflecting individual and disease heterogeneity, the level of sample contamination as well as under-sampling, which is a characteristic of the shotgun strategy adopted. Nevertheless it represents the largest GeLC-MS/MS derived catalogue of biliary proteins identified in multiple patients with hilar CCA and extends the biliary proteome considerably with the largest previous report identifying 170 proteins (222 proteins if based upon single hit peptides)..

Both lipocalin 2 and deleted in malignant brain tumours 1 protein were identified in all four patients with CCA in our study. Elevated levels of lipocalin 2 in bile were shown to be able to distinguish between malignant from benign disease and when incorporated with the currently used serum marker CA19-9 improved the diagnostic accuracy achieved [144]. However translation to levels in lipocalin 2 in urine and serum samples failed to show significant differences between benign and malignant biliary disease.

4.4.3 Comparison of Proteins Identified by 2D PAGE and Shotgun Approaches

A total of 813 unique proteins were identified in a GeLC/M/MS approach and 81 proteins using a 2D PAGE resulting in a total of 824 unique proteins. There were 11 proteins (Acyl-CoA-binding protein, Alpha-synuclein, Cystatin A, Alpha-2-glycoprotein, Lithostathine-1-alpha, Migration-inducing gene-9, Myosin regulatory light chain 12A, Plastin-2, Proapoptotic caspase adapter protein, Retinol-binding protein 4, Tubulin-specific chaperone A) were identified in the master map and not included in the shotgun catalogue. Analysis using a combination of 2D PAGE and 'shotgun' approaches to analyse gallbladder bile without any depletion steps from an individual with cholethiasis, identified a total of 222 (mainly plasma) proteins – (48 and 218 from 2DE LC-MS/MS and 2D-LC-MS/MS, respectively) [182]. 44 of these proteins were common to both the techniques and 27 and 143 proteins in common with our study results.

4.4.4 Conclusions

In summary we present a comprehensive catalogue of bile proteins in four patients with hilar CCA and have considerably extended current knowledge of the malignant bile proteome. Prior to the current study, work carried in our laboratory using a similar parallel proteomic approach in a single patient with CCA identified 32 and 447 unique proteins. The shotgun approach used here shows clear advantages over gel-based top down techniques in speed, sensitivity and scope of analysis.

Profiling studies such as the one described here have formed the basis for examining the expression of a number of proteins in bile in relation to disease, as illustrated by the work taken forward on Mac2 binding protein, which was identified as a bile protein in a shotgun study of bile from a CCA patient and subsequently shown to have some potential as a biomarker [169]. In the period after our completion of out work two further studies have added to and extended the catalogue of biliary proteins albeit in non-cancerous disease or pancreatic carcinoma. Barbhuiya et al fractionated non-cancerous bile and used a multipronged proteomic platform approach (SDS PAGE, SCX, OFFGEL) followed by MS analysis using a more powerful mass spectrometer (LTQ-Orbitrap-Velos) than in our study. Overall 2552 proteins were identified - the largest number of proteins reported in human bile to date [211]. In malignant bile obtained by ERCP in a single a patient with pancreatic carcinoma 445 unique proteins with at least 2 significant peptides (812 proteins if single-hit proteins were included) were identified by SDS PAGE, in gel tryptic digestion and LC-MS/MS [212]. The current dataset together with other studies in the literature constitute an important prelude highlighting the potential promise of comparative quantitative proteomic studies in CCA biomarker discovery.

5.0 Label-Free Quantitative Proteomic Comparison of Bile in CCA and Benign Biliary Disease

5.1 Introduction

The ability to quantitate changes in protein expression that occur in disease is an important goal in proteomics. Studies have identified differences in protein profiles in bile from patients with malignant biliary disease compared to benign disease and a few proteins have gone on to demonstrate potential biomarker utility in initial validation studies. Examples include CEACAM-6 and MUC-1 (CA19-9) which were shown to be higher in the bile of CCA and pancreatic carcinoma patients compared to benign controls using an SDS PAGE and LC-MS/MS approach [186]. Kawase and colleagues identified 38 differentially expressed proteins in two paired cancer and normal bile duct tissues using a label free LC-MS/MS approach and validated findings in further tissue samples [150]. A number of proteins previously reported in CCA including MUC5AC [151], moesin [152] galectin 1 [153] and keratin 903 [154] were increased in CCA tissues. Western blotting and immunohistochemical validation of four novel proteins (actinin-1, actinin-4, protein DJ-1 and cathepsin B) in the original samples and four additional CCA cases showed overexpression compared with normal bile duct tissue.

To build on our previous shotgun-based characterisation of the bile proteome, a comparative analysis of bile from patients with benign and malignant disease was carried out. A label-free approach for quantitation was selected, as it does not require expensive labelling reagents and multistep labelling protocols, which can potentially raise issues with reproducibility and loss of target peptides. At the time of this study a profiling pipeline was being developed in the lab using an LTQ-Orbitrap Velos mass

spectrometer and MaxQuant-based quantitation. The application of such an approach to bile has been recently reported after completion of our study in a comparative analysis of pancreatic cancer and benign biliary disease in which samples from 8 patients were subjected to 1D SDS PAGE and subsequent MS analysis using an LTQ linear ion trap mass spectrometer [144]. Over 200 unique proteins were identified with 10 proteins varying significantly between study groups. The investigators took forward lipocalin-2 as a potential biomarker in differentiating malignant and biliary obstruction and confirmed its higher abundance in cancer. The application of label free sample preparation in combination with a LTQ-Orbitrap Velos mass spectrometer offers a technological advance in identifying a higher number of proteins with less potential of sample preparation bias.

5.1.1 Aims

(1) To use shotgun MS with label free quantitation to identify differences in abundance of biliary proteins between samples from 5 patients with CCA and 5 patients with benign biliary disease (common bile duct stones).

(2) To take forward selected proteins displaying increased abundance in bile from patients with malignant disease by immunoblotting in a larger patient cohort (13 CCA vs. 13 benign disease) to explore whether selected proteins may have clinical utility as biomarkers.

5.2 Sample processing and MS analysis for Label Free Quantitative Proteomic Comparison of Bile

Bile samples were obtained from patients undergoing diagnostic or therapeutic ERCP for evaluation/treatment of biliary obstruction either as a result of malignancy or common bile duct stones. Demographics and clinical details are provided in Table 11. Patients were age and sex matched. Bilirubin, AST and ALT values were all higher in malignant patients, but this was not statistically significant. Samples underwent in-solution tryptic digestion using FASP and MS analysis as described in Sections 2.11.2 and 2.11.3 respectively. Data was searched against an International Protein Index (IPI 3.68) human sequence database with MaxQuant 1.1.1.14 and Wilcoxon-Mann-Whitney used to compare intensities between the two groups.

Table 11. Demographic and clinical data for the patients used in the discovery screen.

Patient	Age	Sex	Disease	Bilirubin (mmol/l)	ALP (U/l)	ALT (U/l)
1 8386	63	M	CBD stones	67	213	41
2 8397	57	M	CBD stones	47	222	55
3 9074	81	F	CBD stones	91	249	61
4 9847	58	M	CBD stones	65	289	41
5 10844	36	F	CBD stones	81	301	53
6 8628	77	M	HIlar CCA	98	321	41
7 9075	54	F	Hilar CCA	101	465	88
8 9089	72	M	Hilar CCA	87	354	67
9 9417	65	F	Hilar CCA	31	145	126
10 10498	47	M	PSC/Hilar	121	406	59

5.3 Results

In total 1608 proteins were identified in the total data set of which 1276 were identified with one or more peptides and with valid quantitative data. Data for each protein in the study including intensity minimum, 1st quartile, median, mean, 3rd quartile, maximum and standard deviation in each group, followed by the p value and associated q value are provided in Supplementary Data Table 5. The overlap in proteins between the different samples in the benign and malignant disease groups is illustrated in Venn diagrams (Figure 21) and the number of unique proteins is summarised in Supplementary Data Table 6. In the benign and cancer samples there were 201 and 215 proteins present in all biological replicates.
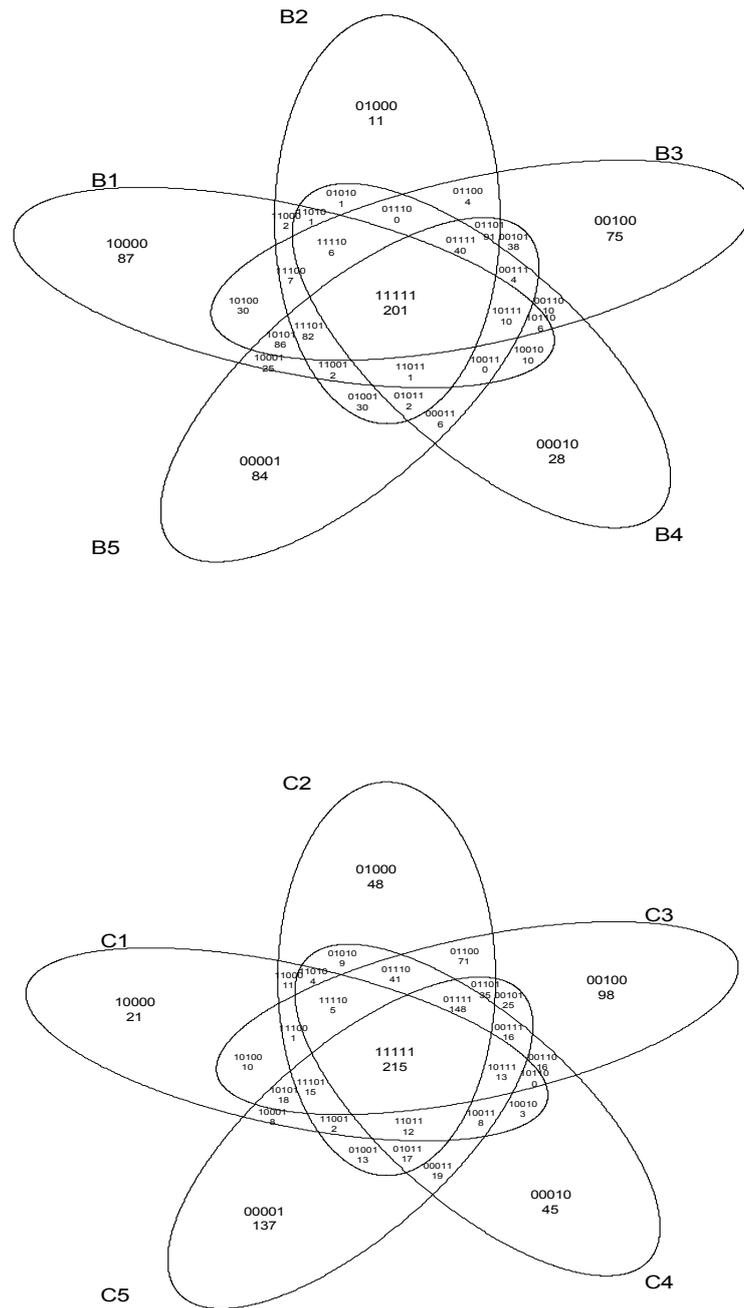
Figure 21. Venn diagram showing overlap in quantitated proteins in benign (above) and cancer (below) samples.

The five-figure indicator above each number is a presence/absence binary indicator for a protein being present in a sample. For example, the central region, which is the overlap between all regions, is shown as 11111, this indicates presence in all replicates.

The gene products in the entire data set were subjected to Gene Ontology Database classification (version 7.0 www.pantherdb.org) and of the 1276 proteins, 1201 were annotated and analysis of these is shown in Figure 22. The majority of proteins related to catalytic activity (38%), protein binding (30%) and structural molecular activity (10%) for molecular function.   Proteins involved in metabolism (27%), cellular processes (10%), cell communication (10%), immunity (10%) and transport (10%) formed the majority of biologic processes. Finally a wide range of classes of proteins existed with the most abundant relating to hydrolase activity (15%).

Molecular Function

- Antioxidant activity (1%)
- Binding (30%)
- Catalytic activity (38%)
- Enzyme regulator (7%)
- Ion channel (1%)
- Motor activity (1%)
- Receptor activity (6%)
- Structural molecule activity (10%)
- Transcription regulator activity (1%)
- Translation regulator activity (1%)
- Transporter regulator activity (4%)

Biological Process

- Apoptosis (2%)
- Cell adhesion (4%)
- Cell communication (10%)
- Cell cycle (3%)
- Cellular component organisation (9%)
- Cellular process (16%)
- Development process (6.5%)
- Generation of precursor metabolites (2%)
- Homoestasis (0.1%)
- Immune system (10%)
- Localisation (0.1%)
- Metabolic process (27%)
- Regulation of biological process (0.2%)
- Reproduction (0.7%)
- Response to stimuli (7%)
- System process (3%)
- Transport (10%)

Protein Class

- Calcium binding (3%)
- Cell adhesion molecule (3%)
- Cell junction protein (0.1%)
- Chaperone (3%)
- Cytoskeletal (7%)
- Defense/immunity (5%)
- Enzyme modulator (8%)
- Extracellular matrix (3%)
- Hydrolase (15%)
- Isomerase (2%)
- Kinase (1.6%)
- Ligase (1.2%)
- Lyase (2%)
- Membrane traffic (2%)
- Nucleic acid binding (7%)
- Oxidoreductase (7.5%)
- Phosphatase (1.5%)
- Protease (7%)
- Receptor (5%)
- Signalling molecule (5%)
- Storage protein (0.1%)
- Structural protein (1.2%)
- Transcription factor (1%)
- Transferase (3.6%)
- Transfer (5.3%)
- Transmembrane (1%)
- Transporter (4%)

Figure 22. Gene ontology classification of 1201 proteins from dataset.

In total there were 37 proteins with differential abundance between study groups (p<0.05) of which 13 had higher abundance in CCA and these are summarised in Table 12. Of these, 5 were selected for downstream validation using Western blotting of bile samples: MMP-9, Chitinase -1, Ly-GDI, Annexin A3 and Nicotinamide phosphoribosyltransferase (PBEF). The intensity value for each protein in each patient is shown in Figure 23 and their biological functions and previous findings in CCA are summarised in Table 13.

Table 12. Protein identity and fold change of proteins differentially expressed in bile from patients with malignant and benign disease.

| No	Protein ID	Protein Name	Fold Change	p value
1	*IPI00027509*	*MMP-9*	*+5.4*	*0.03*
2	*IPI00478217*	*Chitinase-1*	*+4.5*	*0.02*
3	*IPI00003817*	*Ly-GDI (RhoGDP dissociation inhibitor 2)*	*+4.25*	*0.03*
4	*IPI00024095*	*Annexin A3*	*+4*	*0.02*
5	*IPI00018873*	*Nicotinamide phosphoribosyltransferase (PBEF)*	*+3.5*	*0.03*
6	IPI00022389	C-reactive protein	+2.8	0.01
7	IPI00304612	60S Ribosomal protein	+2.6	0.02
8	IPI00021827	Defensin, alpha 3	+2.5	0.01
9	IPI00010154	Guanosine diphosphate dissociation inhibitor	+2.5	0.02
10	IPI00014625	Calcium activated chloride channel protein	+2.3	0.02
11	IPI00029039	Human proislet peptide	+2.4	0.04
12	IPI00010706	Glutathione synthase	+2.1	0.02
13	IPI00298994	Talin-1	+2.3	0.03
14	IPI00022431	Alpha-2-HS glycoprotein	-1.4	0.03
15	IPI00022426	Alpha-1 microglycoprotein	-1.4	0.03
16	IPI00026199	Extracellular glutathione peroxidase	-1.4	0.007
17	IPI00550991	Alpha-1- antichymotrypsin	-1.8	0.03
18	IPI00032179	Antithrombin-III	-1.8	0.015
19	IPI00291867	C3B/C4B inactivator	-1.9	0.015
20	IPI00553177	Alpha-1 protease inhibitor	-1.9	0.01
21	IPI00021842	Apolipoprotein E	-2.1	0.01
22	IPI00021841	Apolipoprotein A1	-2.3	0.007
23	IPI00744685	Biotinidase	-2.3	0.007
24	IPI00292950	Serpin peptidase inhibitor	-2.4	0.04
25	IPI00009896	Epoxide hydratase	-2.4	0.04
26	IPI00400826	Aging associated gene 4 protein	-2.4	0.03
27	IPI00843975	Cytovillin	-2.6	0.01
28	IPI00021854	Apolipoprotein A2	-2.6	0.007
29	IPI00925635	Insulin like growth factor	-3.0	0.04
30	IPI00009802	Chondroitin sulphate proteoglycan core protein	-3.0	0.04
31	IPI00171411	Golgi membrane protein 1	-3.1	0.045
32	IPI00465213	Annexin A13	-3.3	0.01
33	IPI00010399	Serum amyloid a protein	-3.6	0.03
34	IPI00328609	Kallilrein inhibitor	-3.6	0.007
35	IPI00003527	Ezrin-radixin-moesinbinding phosphoprotein	-3.6	0.01
36	IPI00021000	Bone sialoprotein 1	-4.7	0.03
37	IPI00006114	Cell proliferation inducing gene 35 protein	-4.7	0.04
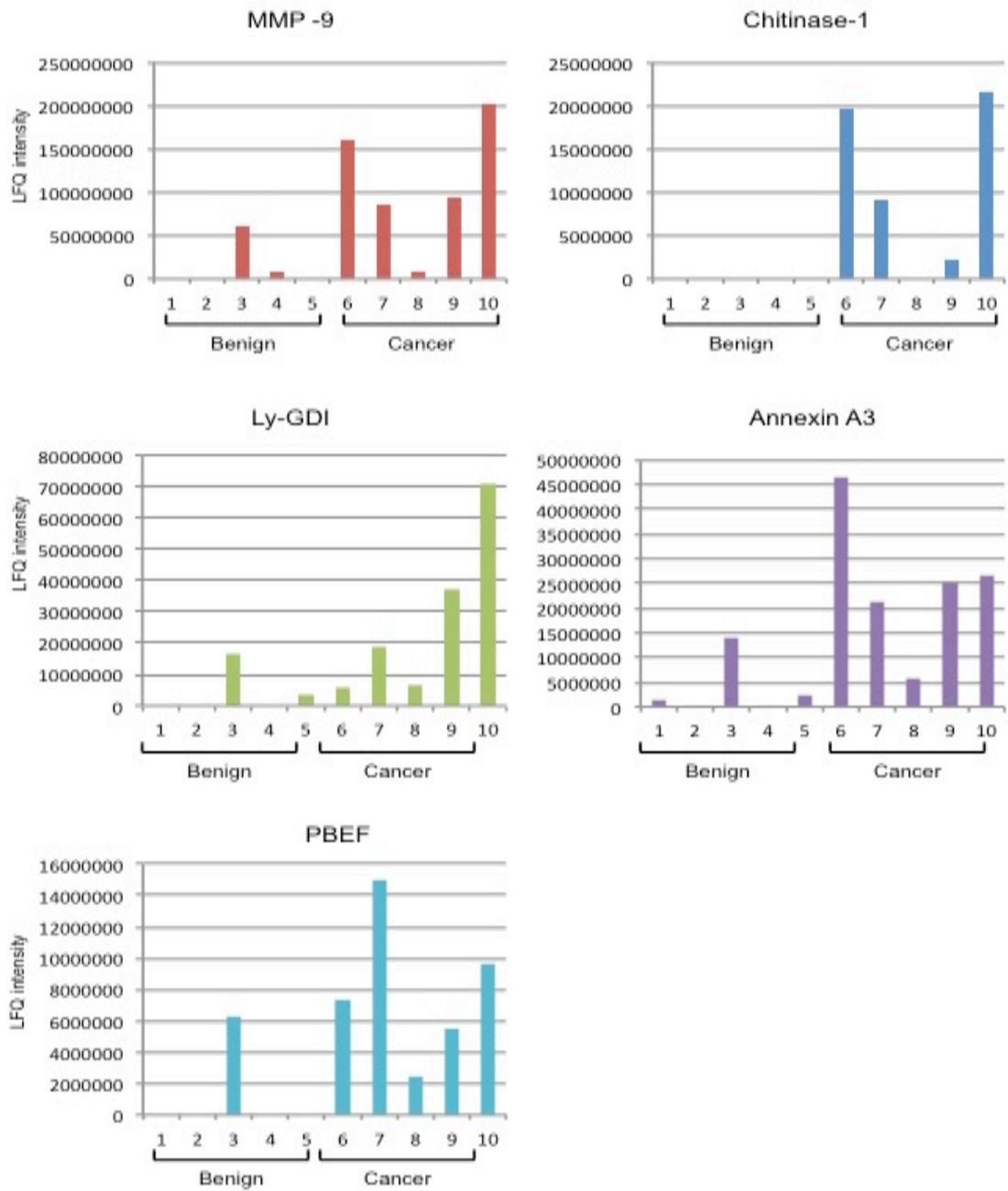
Figure 23. MS data for proteins selected for downstream analysis.

The relative abundance (LFQ intensity) of each protein in each patient is shown.

Table 13. Candidate biomarkers taken forward for downstream validation.

Protein	MW (kDa)	Function	Reported in disease	Reported in CCA
MMP-9	82	Belongs to family of metalloproteinases able to degrade all components of the extracellular matrix	Cancer – including HCC, pancreatic, colorectal Inflammatory – Osteoarthritis, Aneurysm	Yes (up)
Chitinase-1	52	Human CHIT-1 is considered to provide defense against chitin containing pathogens: crustaceans, insects, fungi, parasites	Chronic inflammatory: Gaucher disease, asthma, sarcoidosis, juvenile idiopathic arthritis, helicobacter pylori gastritis, coronary disease, periodontitis liver fibrosis, parasitic/fulipocalin-2 infections	No
LyGDI	42-44	Belong to the inhibitors of Rho family GDPases and in cancer functions in metastasis by anchoring Rho proteins to the cell membrane	Gastric, heamapeoic bladder cancer	No
Annexin A3	33-36	Belong to annexin family which have a range of functions including anticoagulation, anti-inflammatory, endocytosis and exocytosis, signal transduction, cell proliferation, cell differentiation and cell apoptosis	Ovarian, pancreatic, hepatic, colorectal cancers	No
PBEF	55	PBEF catalyzes the first rate-limiting step in converting nicotinamide to NAD+, essential for cellular metabolism, energy production, and DNA repair	Acute/Chronic inflammatory - acute lung injury, atherosclerosis, diabetes, rheumatoid arthritis, sepsis Cancer – colorectal, brain	No

5.3.1 Validation of Selected Differentially Expressed Proteins in Bile

Western blotting was used to compare the expression of annexin A3, chitinase -1, Ly-GDI, MMP-9 and PBEF between 13 benign and 13 malignant bile samples. In initial antibody work up, antibodies to chitinase-1 failed to identify a band of the correct size so this was not taken further. All other antibodies passed this validation and identifying a band of appropriate size and showed linearity when probing serial dilutions of bile. Patient demographics of the cohort used for validation are shown in Table 14. A sample from the original label-free experiment (9089), which contained all the proteins selected above, was included as a positive control. Median age between groups was 67 years. (IQR 55-74) and 70 years. (IQR 60-83), p=ns and the degree of biliary obstruction, as measured by bilirubin between cancer and benign patients was increased; 101 mmol/l (IQR 85-122) vs. 81 mmol/l (IQR 67-89) p=0.03.

Table 14. Demographics and clinical details for validation screen

Patient		Age	Sex	Disease	Bilirubin (mmol/l)	ALP (U/l)	ALT (U/l)
A	10319	61	F	CCA (Hilar)	121	564	101
B	11469	84	M	CCA (Intra)	90	411	143
C	9075	54	F	CCA (Hilar)	101	465	88
D	9037	67	M	CCA (Intra)	66	322	71
E	8619	57	M	CCA (Hilar)	132	439	111
F	8445	73	M	CCA (Hilar)	119	399	81
G	11162	82	F	Benign	67	201	79
H	9074	81	F	Benign	47	222	55
I	9059	62	F	Benign	88	345	61
J	9044	67	M	Benign	110	301	100
K	9741	87	F	Benign	86	275	46
L	8386	63	M	Benign	67	213	41
M	9066	70	M	Benign	90	387	52
N	9847	58	M	Benign	65	289	41
O	10844	36	F	Benign	81	301	53
P	11515	45	M	Benign	98	222	41
Q	11517	88	F	Benign	80	351	67
R	11162	82	F	Benign	77	401	55
S	10310	84	M	Benign	90	377	61
T	8628	77	M	CCA (Hilar)	98	321	41
U	10498	47	M	CCA (Intra)	91	401	87
V	8619	57	M	CCA (Hilar)	123	512	70
W	10092	75	M	CCA (Intra)	76	499	89
X	11451	73	F	CCA (Hilar)	118	1019	102
Y	11543	72	F	CCA (Hilar)	126	766	89
Z	9363	25	M	CCA(Hilar)/PSC	80	400	58

Annexin A3 expression was observed in 4/13 cases of CCA compared with 2/12 patients with benign disease (Figure 24). Annexin A3 expression was only found in hilar CCA samples, 50% of which had detectable levels. Interestingly, the two patients with benign biliary disease showing expression had bilirubin levels of 91 and 61 mmol/l, which represented some of the highest values in the benign group.
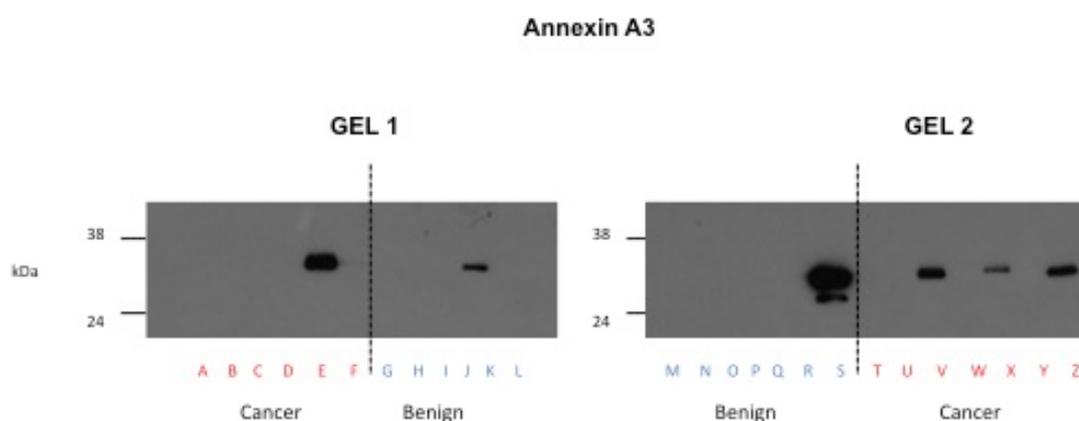

Figure 24. Expression of Annexin A3 in benign and cancer bile samples.

Biliary protein (5µg) was separated by SDS-PAGE and analysed by immunoblotting with antibodies to Annexin A3. Signals were normalised by Coomassie staining of parallel gels [213]. Samples labelled A-Z correspond to Table 14.

MMP-9 expression was shown to be at higher levels in bile from patients with malignant disease, being present in 11/13 cases compared with 4/13 with benign disease (Figure 24). The highest fold change in the initial screen was detected for MMP-9 (>5.4). All patients with hilar CCA demonstrated expression of MMP-9.
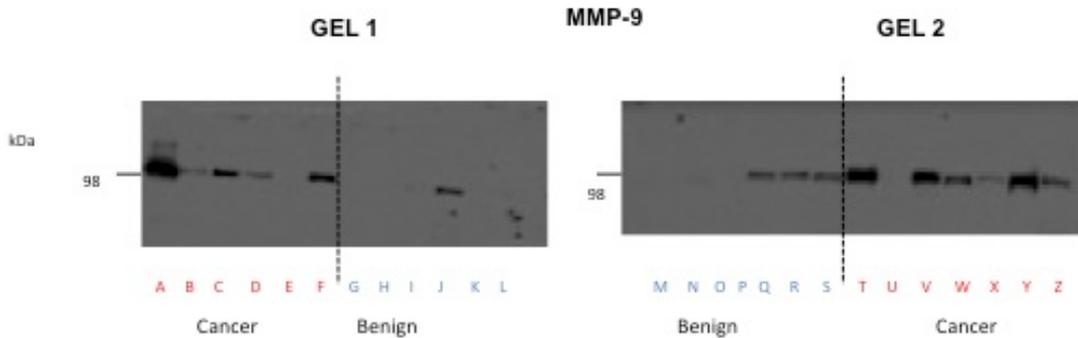

Figure 25. Expression of MMP-9 in benign and cancer bile samples.
Biliary protein (5μg) was separated by SDS-PAGE and analysed by immunoblotting with antibodies to MMP-9. Signals were normalised by Coomassie staining of parallel gels (213). Samples labelled A-Z correspond to Table 14.

LyGDI expression was observed in 6/13 of cases CCA compared with 4/13 with benign disease (Figure 26A). Its expression showed similar variation in both groups and did not differentiate based on type of cancer, degree of biliary obstruction or age. Similarly expression of PBEF was observed in all biological samples with no significant difference in the level of expression across groups (Figure 26B).
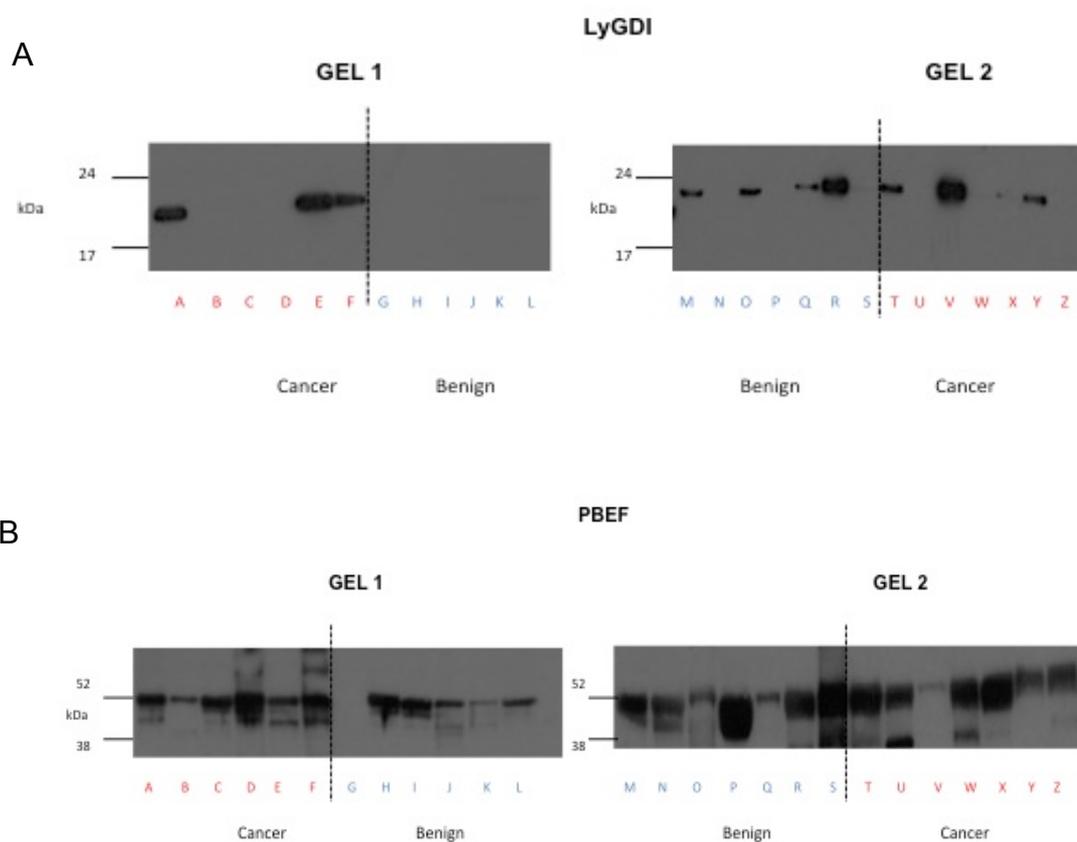

Figure 26A & 26B. Expression of LyGDI and PBEF in benign and cancer bile samples
Biliary protein (5µg) was separated by SDS-PAGE and analysed by immunoblotting with antibodies to LyGDI and PBEF. Signals were normalised by Coomassie staining of parallel gels. Samples labelled A-Z correspond to Table 14.

5.3.2 Preliminary Analysis of MMP-9 and lipocalin-2 and its Complex in CCA

Previous work in this laboratory showed higher expression of MMP-9 and lipocalin-2 in bile from a small cohort of patients with CCA (Bonney G, unpublished data). In human neutrophils, some of circulating MMP-9 exists in a covalent complex with lipocalin-2, which has a molecular weight of 125-130 kDa, and is reduction sensitive and able to bind to tissue inhibitor of metalloproteinase – 1 (TIMP-1) [214-215]. The complex has been shown to be involved in regulating the activity and stability of MMP-9 by protecting it from auto degradation in *in vitro* and *in vivo* studies [215].

In light of the observation of increased MMP-9 expression in our screen, the study was extended to investigate a potential role of the MMP-9 – lipocalin-2 complex in CCA. Initially the expression of lipocalin-2 was examined in bile from 7 patients with CCA and benign biliary disease then subsequent experiments went on to examine the MMP-9/lipocalin-2 complex, which can be visualised using non-reducing gels.

5.3.3 Expression of lipocalin-2 in CCA

Lipocalin-2 was detected at relatively low levels in bile from patients with benign biliary diseases (CBD stones) compared to CCA where there was higher abundance in 6/7 cases, consistent with our earlier results (Bonney, unpublished data).
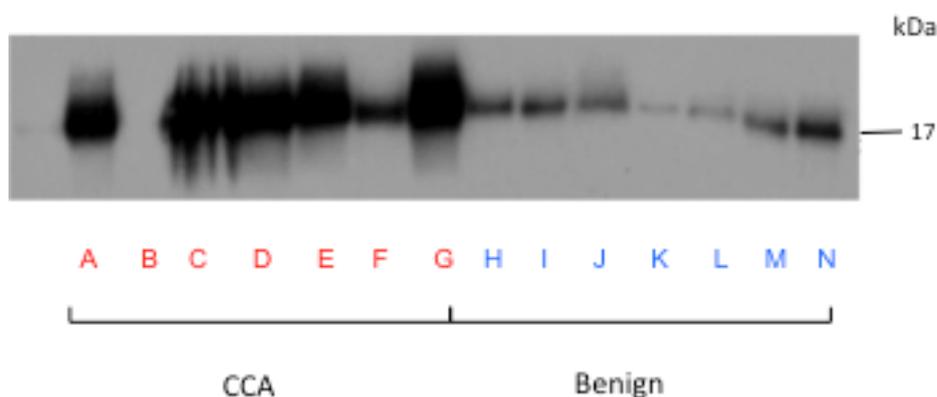

Figure 27. Expression of Lipocalin-2 in bile from patients with CCA cancer and benign disease

Expression of Lipocalin-2 was examined in bile from patients with CCA and benign biliary disease. Protein (5µg) was separated by SDS PAGE and analysed by immunoblotting with antibodies to Lipocalin-2. Signals were normalised relative to parallel Coomassie stained gels. Samples A-N corresponds to patients in Table 14.

5.3.4 Expression of MMP-9 - lipocalin-2 Complex in Bile in CCA

Using immunoblotting under non-reducing conditions (i.e. without β-mercaptoethanol) the MMP-9-lipocalin-2 complex was examined in bile from patients with CCA, which confirmed the presence of a 125 kDa band recognised by antibodies to MMP-9 and lipocalin-2 which co-migrated with recombinant MMP-9-lipocalin-2 complex. This band was not demonstrated in bile samples run on gels run under reducing conditions (Figure 28 A & B).
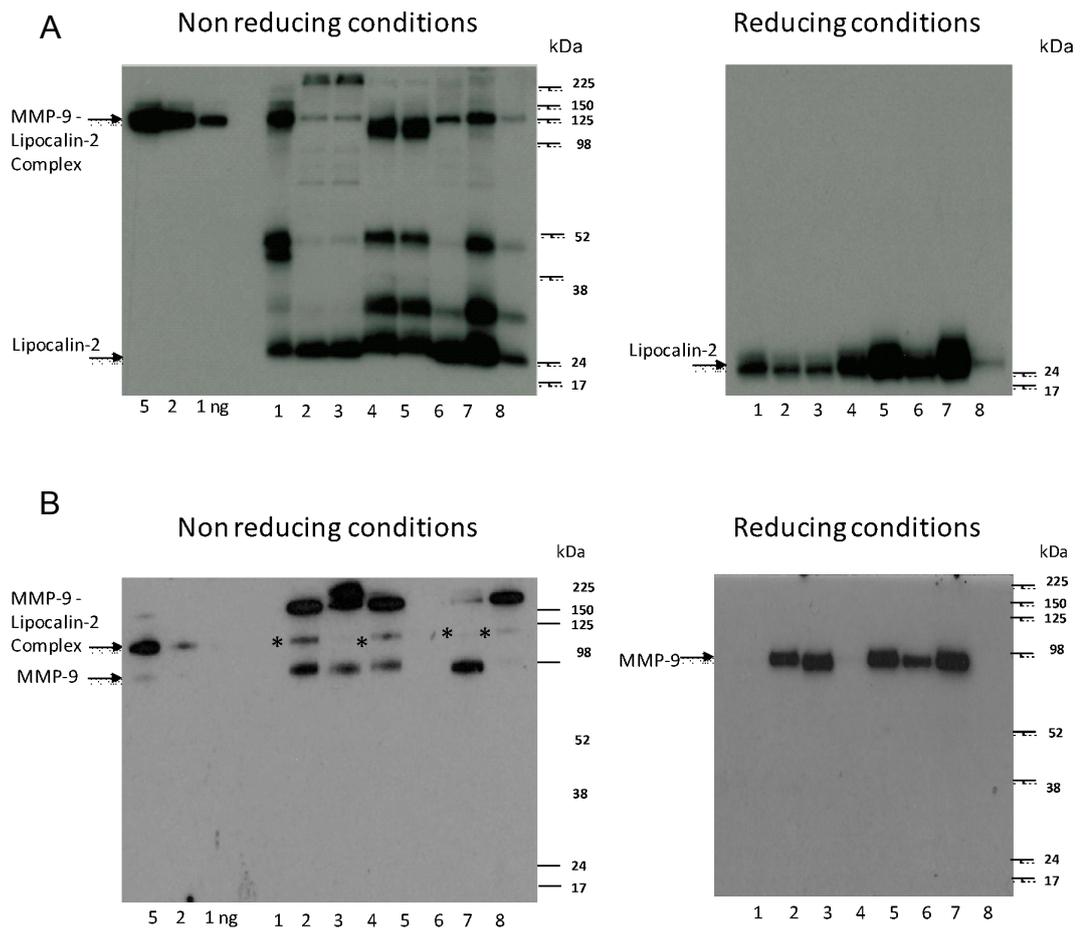
Figure 28 A & B. Immunoblotting of MMP-9, Lipocalin-2 and the MMP-9- Lipocalin-2 complex.

In total 8 patients with CCA were probed with antibodies to lipocalin-2 and MMP-9 respectively. Biliary protein (5µg) (lanes 1-8), and serial dilution of MMP-9-lipocalin-2 complex (lanes 5, 2, 1 ng) were separated by SDS PAGE in the absence or presence of β-mercaptoethanoland (i.e. reducing and non reducing conditions) and analysed by immunoblotting with antibodies to lipocalin-2 and MMP-9. Signals were normalised by densitometry analysis of parallel Coomassie stained gels. Results confirmed the presence of the complex in non-reducing conditions using antibodies to both MMP-9 and lipocalin-2.

In a subsequent experiment the MMP-9-lipocalin-2 complex was compared in bile samples of patients with CCA and benign biliary disease (Figure 29). Results showed the presence of the complex in greater number /intensity in cancer compared to benign bile samples.
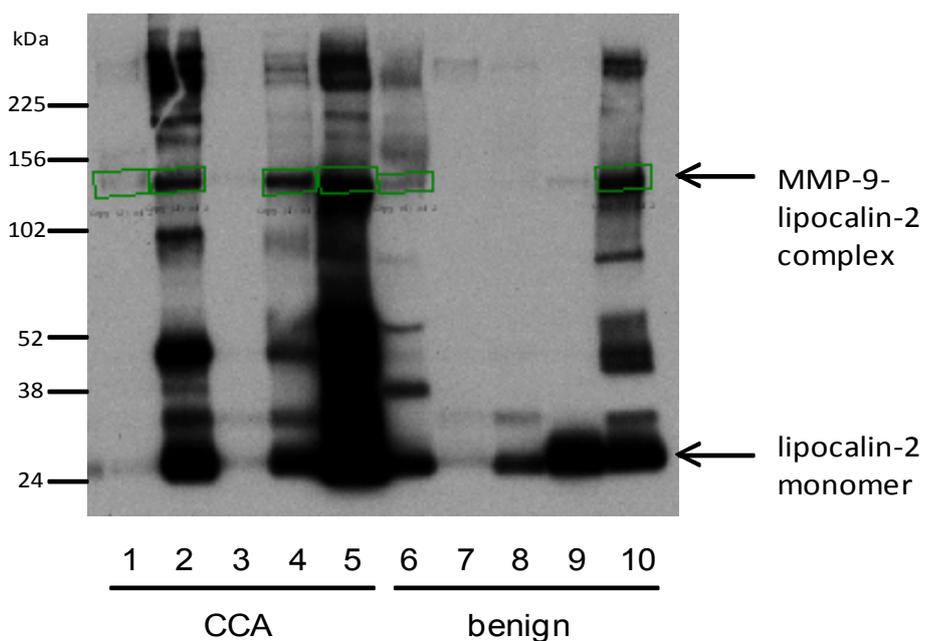

Figure 29. Immunoblotting of MMP-9- lipocalin-2 complex in bile CCA vs. benign biliary disease. Biliary protein (5µg) and serial dilution of MMP-9-lipocalin-2 complex in sample 1 (9089) were separated by SDS PAGE in non reducing conditions and analysed by immunoblotting with antibodies to lipocalin-2. Signals were normalised by densitometry analysis of parallel Coomassie stained gels.

5.4 Discussion

At the time of this study and to our knowledge, to date, this is the first shotgun study with label free-based comparative analysis carried out to identify differences in protein abundance between patients with CCA and benign biliary disease (common bile duct stones). The entire dataset yielded 1276 unique proteins and represents one of the largest bile protein compendiums generated at the time, extending our previous shotgun catalogue of bile proteins in hilar CCA by more than 400.

The majority of proteins related to catalytic activity (38%), protein binding (30%) and structural molecular activity (10%) for molecular function. Proteins involved in metabolism (27%), cellular processes (10%), cell communication (10%), immunity (10%) and transport (10%) formed the majority of biologic processes. Finally a wide range of classes of proteins existed with the most abundant relating to hydrolase activity such as metalloproteinases (15%).

The comparative screen comparing 5 patients in each of two groups (CCA and benign biliary disease) identified 37 proteins with significant differences in fold change with 13 proteins being more than 2 fold increased in malignant samples. From this list the top 5 proteins with the most abundant fold changes were selected for further validation: a nnexin A3, chitinase-1, Ly-GDI, MMP-9 and PBEF immunoblotting showed a clear pattern of overexpression of MMP-9 in bile from patients with CCA compared to benign disease. Of note all the patients in the screen with hilar CCA (n=8) demonstrated staining for MMP-9 compared to 3/5 with intrahepatic CCA. Annexin A3, PBEF and LyGDI did not demonstrate an ability to differentiate benign and malignant disease.

In light of previous experiments showing elevated levels of lipocalin-2 in CCA, other studies suggesting a protective role for lipocalin-2 against MMP-9 auto degradation, promoting angiogenesis and tumour growth in cancer and the finding that an MMP-9-lipocalin-2 complex could differentiate pancreatic malignancy from benign disease in bile prompted further investigation (Bonney G et al unpublished data), [144], [216].

In this study 7 patients with CCA were compared to 7 with benign biliary disease and increased lipocalin-2 expression was observed in in 6/7 patients with cancer. At the time of this study no studies had investigated the expression of the MMP-9/lipocalin-2 complex in bile or bile duct tissue in CCA. We detected a 125 kDa band in bile from patients with CCA by probing with MMP-9 and lipocalin-2 antibodies which co-migrated with MMP-9 – lipocalin-2 complex under non-reducing conditions. Taking the results of the label free comparative experiment in identifying MMP-9, elevated levels of lipocalin-2 in previous laboratory studies and here, and the presence of a complex in malignant bile, there is potential scope for further biomarker investigation. Indeed in a preliminary comparison of 5 cancer and 5 benign disease samples we did observe strong expression of the complex in cancer.

5.4.1 Matrix Metalloproteinase – 9 (MMP-9)

MMPs are a family of zinc and calcium dependent endopeptidases with the combined ability to degrade all components of the extracellular matrix. MMPs have been implicated in cell migration during cancer invasion, tumour growth, angiogenesis and metastasis in several gastrointestinal cancers including colorectal, esophageal and hepatopancreaticobiliary [217-221]. Furthermore their ability to degrade cellular and extracellular matrices has seen them play a role in tissue damage caused by

inflammation [222-224]. The detailed role of the 23 types and 6 groups of MMPs in disease has been reviewed elsewhere [225-226].

MMP-9 is synthesised as a (92kDa) proenzyme form which is activated into a (82kDa) form that digests components of the basement membrane including gelatin (hence also known as gelatinase B), laminin and type IV collagen. MMP-9 is produced by myofibroblasts, monocytes, keratinocytes, tissue macrophages, and polymorphonuclear leucocytes and its expression/secretion induced and controlled by cytokines, chemokines and peptidoglycans. MMP-9 is usually only expressed at low levels in normal healthy, resting tissues and elevated in environments requiring tissue repair/remodelling or in inflammation and disease [225], [227]. Secretion of MMP-9 is accompanied by a TIMP-1 which is the major endogenous regulator of its activity, ensuring a balance between matrix synthesis and degradation [226].

Fibrosis is one hallmark of CCA caused by recurrent infection and inflammation within the bile ducts and is a dynamic process where progression and regression can be present alternatively during its evolution resulting in complex remodelling of the extracellular matrix. This dysfunction in the extracellular matrix alters normal cell function, leading to cell proliferation and myofibroblast-induced inflammation and angiogenesis to facilitate cancer development [228]. Furthermore the secretion of MMPs by myofibroblasts can degrade the basement membrane and lead to tumour invasion and metastasis.

The specific overexpression of MMP-9 in CCA was initially reported in 1996 in which Terada et al analysed the expression of MMPs and TIMP in 10 normal livers, 11 surgically resected intrahepatic CCA and 6 surgically resected HCC. In normal livers, MMPs and TIMPs were infrequently and faintly expressed in bile ducts, but were not

expressed in hepatocytes. In the 11 CCAs, MMP-1, MMP-2, MMP3, MMP-9, TIMP-1, and TIMP-2 were expressed in tumour cells and/or tumour stroma in 11 (100%), 5 (45%), 8 (73%), 3 (27%), 9 (82%), and 9 (82%), respectively [230]. Their expression was stronger in CCA with severe invasion than in specimens with mild invasion. In contrast, MMPs and TIMP were not expressed in any cases of HCC. Their results showed that intrahepatic bile duct cells may neoexpress or overexpress MMPs and TIMPs after malignant transformation but hepatocytes do not, and suggest that in CCA, MMPs and TIMP play an important role in cell invasion by degrading extracellular matrix proteins.

More recently studies in hamsters infected with Opisthorchis viverrini and injected with nitrosamine compounds were observed to be associated with increased peribilary fibrosis and CCA development [231]. In this model, MMP-9 levels were increased and correlated with levels of myofibroblast activity, fibrosis and cholangiocarcinogenesis. The authors proposed that MMP-9 stimulates myofibroblast-mediated fibrosis leading to tumour formation, and then cancer-associated myofibroblasts promote tumour progression by secretion of further MMP-9.

Overproduction of NO by iNOS and ROS results in activation of MMPs and degradation of the basement membrane [232]. In the same animal model above investigators confirmed co-localisation of MMP-9 and iNOS in inflamed and malignant tissue and provided further evidence to link iNOS expression and NO production mediated activation of MMP-9 leading to CCA development via matrix degradation. Furthermore MMP-9 expression was associated with expression of Rac1 and together can promote ROS mediated DNA damage and invasive potential [231]. Other important inflammatory mediators of MMP-9 expression include TNF-α and fascin, which confer increased migration ability of CCA cells.

Given the close link between inflammation and cancer progression, the role of MMPs in immune response regulation is being studied. One mechanism of interest outlines the effect of TNFα on NF-kβ signaling in tumour cells creating a "feed-forward loop" stimulating chemokines, and MMPs (including MMP-9) in tumour and surrounding stromal cells [233]. Chemical Inhibitors of iNOS and MMP-9 have been shown to reduce incidence of CCA in animal models and attenuate angiogenesis and tumour invasion in prostate cancer respectively [234-237]. Currently no specific MMP-9 inhibitors have been assessed in patients with CCA.

As a biomarker of diagnostic and prognostic value, MMP-9 has been studied in breast, colorectal, gastric, bladder and pancreatic cancers but is yet to be translated into the clinical [238-242]. Expression of MMP-9 in these studies correlated with stage, degree of invasion and identified groups at highest risk of recurrence, which could be targeted with adjuvant therapies. Only a few studies have reported detection of MMP-9 in serum or tissue in CCA. Li et al analysed the expression of MMP-9 in tissue using reverse transcription – PCR and demonstrated higher levels in CCA compared to normal biliary tissue and showed that within CCAs, high levels conferred a more aggressive phenotype [243]. Itatsu et al examined the expression of various MMPs in surgically resected specimens of CCA using immunohistochemistry and found that nearly 50% of specimens expressed MMP-9 and another study confirmed its presence in intra and extrahepatic CCA [221],[244].

In serum, studies comparing MMP-7 and 9, combined with the established markers CEA and CA19-9 as biomarkers to differentiate benign and malignant biliary disease only MMP-7 demonstrated sufficient sensitivity and specificity [245]. However this study did not address pre-analytical factors that may affect MMP-9 namely that citrate plasma is suggested as the sample of choice to measure circulating MMP-9 and not

serum collected with or without clot activator [246]. No subsequent studies have investigated MMP-9 expression in blood in CCA. However, in the context of liver fibrosis (a predisposing factor for malignant conversion in CCA) MMP-9 plays an important role in matrix remodeling and is considered to represent an accurate biochemical marker of fibrotic activity and novel ELISA based assays have been developed to measure MMP-9 mediated changes in the extracellular matrix [247], [248]. In summary, while evidence implicating MMP-9 as a biomarker in the diagnosis and prognosis in other gastrointestinal cancers is growing, in CCA current focus remains on its biology in relation to matrix remodeling and tumour invasion in cholangiocarcinogenesis.

5.4.2 Lipocalin-2

Lipocalin-2 (also know as neutrophil gelatinase-associated lipocalin (NGAL), siderocalin, uterocalin and oncogene 24p3) is a 25-kDa protein belonging to the lipocalin superfamily. This includes 20 secreted lipoproteins, which are similar in molecular structure, that serve to bind and transport specific hydrophobic ligands (prostaglandins, retinoids, arachidonic acid, hormones) [249].

Lipocalin-2 binds to iron particles and transports them intracellularly via the 24p3R membrane receptor. It is known as a "stress protein" that can be induced by various cells including neutrophils and epithelial cells in order to activate iron dependant enzymatic defence systems, inhibit bacterial growth and respond to oxidative stress [250]. In humans, high levels of lipocalin-2 are found in epithelial tissues susceptible to infection including respiratory and gastrointestinal tracts, sites of inflammation (particularly in response to oxidative stress induced by ROS) and cancer [251-252]. The greatest clinical interest in this molecule currently relates to it being an early biomarker

of acute kidney injury, graft rejection post transplantation and severity of heart failure [253-255]. However, evidence continues to grow implicating a role in cell growth, development and differentiation from as early as the embryonic phase and lipocalin-2 has subsequently received interest in tumorigenesis and tumor progression of various human cancers including breast, colorectal, liver and pancreatic [144],[256-259].

Lipocalin-2 is expressed in CCA and knock down of the gene by siRNA has been shown to significantly suppress invasiveness of CCA cells *in vitro* [260]. Proteomic based studies of the secretome of CCA cell lines have shown lipocalin-2 in conditioned media and subsequent small-scale validation confirmed increased lipocalin-2 in paired tumour versus normal tissue in 12 patients [261]. However whether lipocalin-2 regulates CCA proliferation *in vivo* remains to be determined.

In bile, a label free proteomic analysis comparing malignant (n=22) versus benign (n=16) bile showed lipocalin-2 levels were significantly elevated in the malignant group [144]. Investigators then assessed biomarker utility by ELISA in serum and urine. No significant differences in serum and urine lipocalin-2 levels were found between benign and malignant disease but combining biliary lipocalin-2 and serum CA 19-9 improved diagnostic accuracy for distinguishing benign from malignant biliary obstruction (sensitivity 85%, specificity 82%, positive predictive value 79%, and negative predictive value 87%). In a clinical serum study, Leelawat et al compared lipocalin-2 in combination with CA19-9 in 50 patients with CCA and benign biliary disease and demonstrated increased CA19-9 and lipocalin-2 in CCA, a correlation between cancer stage and lipocalin-2 levels and the ability of both molecules in combination to provide the most accurate differentiation between malignant and benign biliary disease (sensitivity 90%, specificity 66%) [263].

Consistent with other studies in the literature, in our western blot anlaysis of lipocalin-2 under non reducing conditions (Figure 28 A) we observed that in addition to the monomeric 25kDa form of the protein, additional bands were seen including the lipocalin-2-MMP-9 complex. This has implications for studies attempting to measure lipocalin-2 levels in biological samples and in comparing such studies as there may be differences in the detection of the free monomer or complexed forms of the molecule. Both the studies described above used the same research use only ELISA kits available from R&D Systems (NGAL; Abingdon, UK) for lipocalin -2 measurement but do not discuss the form of lipocalin that was detected [144], [263]. Indeed variations in the relative amounts of lipocalin-2 forms have been reported depending on the assay and antibody configurations used in research studies [264], [265]. Kift et al compared the analytical performance of five commercially available assays for measuring lipocalin-2 in urine and highlighted that variability exists in the performance of several available kits and has to be taken into account when interpreting results [266]. Furthermore, the investigators proposed that as the relative proportions of the different forms of lipocalin -2 change over time, for example following surgery, more meaningful information could be achieved if assays measuring specific forms were used.

In conclusion, while there is increasing literature reporting the role of lipocalin-2 as a multifaceted modulator in various cancers, its study in CCA is limited. Proteomic investigations have shown it to be present at higher level in malignant bile and immunohistochemical staining confirmed lipocalin-2 overexpression in tumour compared to normal tissue whilst there are conflicting reports regarding its elevation in serum. The influence of lipocalin-2 on carcinogenesis is yet to be fully defined, but involves pathways related to apoptosis/survival and migration/invasion and blockade by monoclonal antibodies has been shown to halt development of metastases in animal models. Further studies investigating the mechanisms underlying its role may

raise the potential of use of lipocalin-2 as both a diagnostic/prognostic biomarker and a therapeutic target. From the results of this study the finding of both MMP-9 and lipocalin-2 expression in CCA raises further interest in the link between both molecules and their complex and this is discussed below.

5.4.3 MMP-9-Lipocalin-2 Complex

As shown in human neutrophils, some of MMP-9 exists in a covalent complex with lipocalin-2, has a molecular weight of 125-130 kDa, and is reduction sensitive and able to bind to TIMP-1 [215]. The complex has been shown to be involved in regulating the activity and stability of MMP-9 by protecting it from auto degradation in *in vitro* and *in vivo* studies. Zhang et al have shown that the activity of the MMP-9-lipocalin-2 complex correlated with the depth of oesophageal squamous cell tumour invasion and postulated that lipocalin-2 promoted metastasis by complex formation resulting in preservation of MMP-9 activity [259]. In breast and gastric cancer this was reported to facilitate angiogenesis/tumour growth and result in poor prognosis [216], [260]. Studies in serum and urine of patients with cerebral, breast and oesophageal cancer have assessed the utility of the complex as a diagnostic biomarker with some success but have yet to be validated in large scale studies [216], [267-269].

Lipocalin-2 knockdown by siRNA in CCA cells suppressed tumour invasion by reducing complex formation and its effect on MMP-9 activity [261]. To relate their *in vitro* findings to the *in vivo* situation, the study was extended to immunohistochemistry of tumour specimens which demonstrated strong lipocalin-2 expression in 75% (18/24) of cases. However, in contrast to the study by Zhang et al in oesophageal cancer, the level of complex expression did not significantly correlate with tumour differentiation, lymph node involvement and metastatic status [267]. A potential co-founding factor in this finding may relate to the significant differences in the biological features of the tumours and the limited number of specimens. Nevertheless the authors provided new insights linking lipocalin-2 and direct promotion of cancer invasiveness by regulating cell motility and by stabilizing MMP-9 via complex formation, which may ultimately lead to potential therapeutic targets. Indeed in our preliminary immunoblotting analysis comparing CCA bile and benign bile we did observe strong presence of the complex in malignant samples and supports further investigation.

5.4.4 Limitations of current study and future directions

In all studies attempting to quantify protein differences in biological fluids including bile from patients in different clinical groups, a major criticism remains the small number of samples (typically <10) used in the analysis and the fact that findings are not validated in larger patient cohorts [178]. Furthermore the selection of proteins overexpressed in cancer for validation set with a fold change of 2 or more can be questioned as potentially excluding other proteins of interest. However a 2-fold change was selected as an appropriate cut-off as changes smaller than this would be below the range of technical variability expected for the label free MS approach. Similarly in the label-free comparative proteomic approach used here, a significant limitation in the study design was the restriction of including only 10 samples in the original screen (each with 3

injections). Reasons accounting for this included availability of mass spectrometry time and the life span of the LC columns used for LC-MS/MS. Consequently in the label - free screen, 5 patients with CCA were compared with 5 patients with benign disease. Examining the distribution of every pair -wise ratio of protein intensities in the dataset, the 95th percentile corresponded to a fold change of 6 as illustrated in scatter plots of protein abundance (Figure 30). This indicates that the screen had reduced power to identify differences associated with disease against a background of biological variation. This may in part explain why some molecules selected from the original screen (Annexin A3, PBEF, Ly-GDI) did not validate by Western blotting (that is, they were false positives). Furthermore studies proposing the underlying biological mechanisms driving carcinogenesis between intrahepatic and hilar CCA being quite different, including both groups in the larger validation screen increases the variability and potentially masks relevant low abundance molecules specific to each form of cancer. Therefore in retrospect the impact of biological variability could have been reduced by the inclusion of only hilar or intrahepatic CCA for proteomic driven biomarker discovery experiments.

Evidence from both in house laboratory experiments preceding this study and in the literature linking elevated lipocalin-2 in pancreaticobiliary malignancy adds further weight to our observation of increased expression of lipocalin-2 in a set of 7 patients with CCA by immunoblotting. A larger screen incorporated more than significant samples in each arm and validation by immunohistochemistry has been undertaken to confirm this in investigations following this study (Nair A et al personal communication, 2013). The finding of the MMP-9-lipocalin complex in bile from patients with CCA is important as it has yet to be reported elsewhere and the underlying biology of its actions implicating a possible role in cholangiocarcinogenesis. In our small comparison of 5 patients with CCA and benign biliary disease we did observe stronger expression

of MMP-9 – lipocalin-2 complex but would require further detailed comparative analysis between malignant and benign biliary disease. Validated overexpression of complex could then be extended to ELISA based assessment in serum of patients.

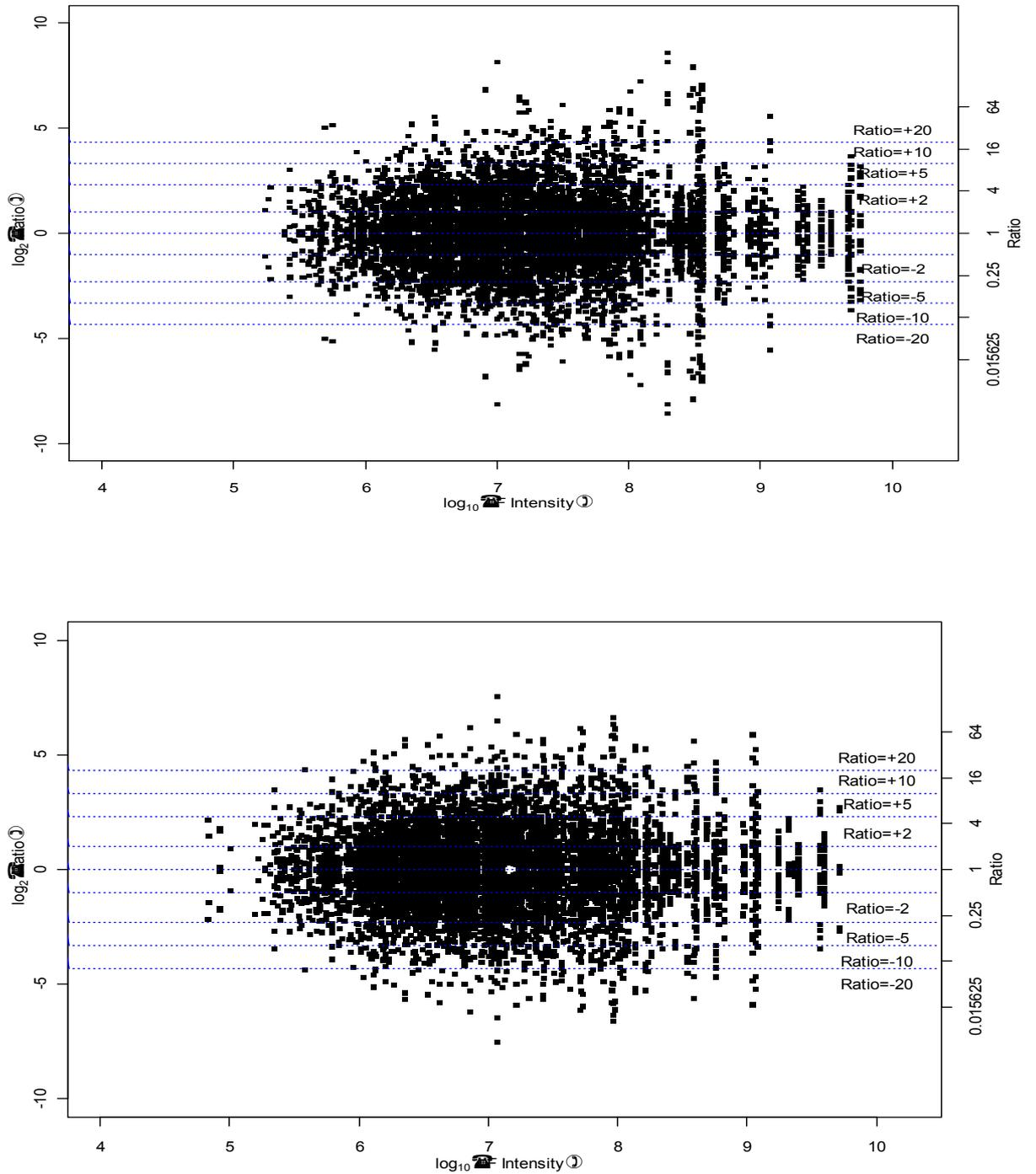**Figure 30. Overall reproducibility of protein abundance in benign and CCA bile.**

Scatter plots of abundance (log10LFQ intensity) versus log2 (ratio) for every possible pair-wise comparison are plotted for bile samples from patients with benign (upper panel) and malignant (lower panel) disease. In the dataset, the 95th percentile corresponded to a fold change of 6 as illustrated in scatter plots of protein abundance

5.4.5 Conclusions

In this chapter we have undertaken the first reported label-free comparative analysis of bile protein profiles between patients with CCA and benign biliary disease caused by common bile duct stones. As a result we have catalogued more than 1200 proteins in bile, extending our previous shotgun catalogue in hilar CCA and at the time of study this represented the largest available compendium. There were 37 proteins displaying differential expression between malignant and benign disease bile and 4 specific proteins (Annexin A3, Ly-GDI, MMP-9 and PBEF) were taken forward in an immunoblotting based validation screen. Only MMP-9 was shown to more abundant in CCA. Interestingly lipocalin-2 - a molecule not identified by the initial screen but of interest from previous in-house studies and its ability to form a complex with MMP-9 – was shown to be more abundant in the bile of CCA patients compared to those with benign biliary disease. Interest in the role of the protective effect of lipocalin-2 on MMP-9 activity via complex formation resulted in one of the first reports of the 125kDa MMP-9-lipocalin-2 complex in CCA. Developments from the results presented here should focus attention on experimental design that can incorporate greater sample numbers and with more homogenous disease (i.e. hilar or intrahepatic) to address the challenge of biological variability. MMP-9 and lipocalin-2 and more recently their complex continue to receive interest in disease associated with inflammation and matrix remodelling such as CCA and warrant further investigation.

6.0 Concluding Remarks

CCA is a devastating disease associated with late presentation and poor survival. Current diagnostic markers such as CA19-9 lack sufficient sensitivity and specificity to identify early disease and distinguish benign from malignant strictures in PSC. A biomarker or panel of biomarkers is urgently required for earlier diagnosis and facilitate better survival outcomes following potentially curative surgical treatment. Proteomics is a discipline that holds promise to identify such molecules and its systematic application to bile in CCA remains in its infancy.

The study herein builds on work carried out previously in the laboratory establishing protocols for the bile sample collection, processing and application to proteomic platforms. Specifically an attempt has been made to (1), investigate the differences in protein profiles between serum and bile using 2D PAGE, (2) apply albumin/IgG depletion strategies to bile before proteomic analysis, (3) compile a compendium of proteins found in the bile of patients with CCA and (4) compare the proteomic profiles between patients with CCA and benign biliary disease (cholethiasis) and identify potential biomarker proteins for preliminary validation.

2D PAGE and DIGE is a powerful approach to separate a large number of proteins simultaneously with high throughput. In our experiment we have produced 2D gels with good resolution and reproducibility, reflecting optimal sample preparation and revealed more than 200 protein spots with difference in levels between bile and serum. The separation of bile and serum protein profiles supports the hypothesis that bile represents a sufficiently unique proximal bio fluid to warrant further analysis for disease-specific biomarker discovery. To date no other investigations using DIGE labelling has been applied to proteomic investigation of bile in CCA.

The bile proteome is similar to serum in having a large dynamic range, which limits proteome coverage. Depletion of abundant proteins is one strategy that can facilitate profiling of lower abundance proteins. Depletion of albumin and immunoglobulin from bile was reproducible and a 2D DIGE comparison of un-depleted and depleted samples showed that there were limited changes in the global protein profile. In PCA of the entire dataset little separation existed between samples from different patients subjected to albumin/depletion with patient variability giving much greater separation between gels. Thus depletion effectively removed major abundant proteins but did not confer significant changes to the overall proteome profile. However with any depletion strategy it is acknowledged that there is the issue of balance between increasing proteome coverage at the expense of co-depletion of potential proteins of interest. This protocol was subsequently used in compiling a compendium of proteins in bile of patients with CCA and applied to the comparative analysis of proteins between those with CCA and benign biliary disease.

Protein biomarker discovery in biliary disease is facilitated by the availability of comprehensive descriptions of the malignant bile proteome. The complementary approaches of 2D PAGE and GeLC MS/MS were utilised to compose the first reported bile protein mastermap and the largest (at the time) catalogue of proteins in malignant bile from four patients with hilar CCA. A total of 80 proteins using 2D PAGE and 813 proteins using a GeLC MS/MS approach were identified resulting in a total of 824 unique proteins. Since then a multiplatform proteomic approach analysed non-cancerous bile samples and identified 2552 proteins to establish the largest catalogue in human bile to date [271] and in another study using 2D PAGE, in-gel digestion and LC-MS/MS approach to malignant bile from pancreatic cancer revealed 445 unique proteins [212]. Both studies have cited and acknowledged the work presented here.

Few studies were noted to have identified differences in protein profiles in bile from patients with malignant biliary disease compared to benign disease and few proteins have gone on to demonstrate potential biomarker utility. We have performed the first comparative analysis of protein profiles in bile with CCA and benign biliary disease using a label free proteomic approach. All together more than 1200 unique proteins were identified from all bile samples and 37 proteins with differential abundance between cancer and benign disease. Of these proteins 13 had higher abundance in CCA and ultimately 4 proteins: MMP-9, Ly-GDI, Annexin A3 and PBEF were taken forward in validation by immunoblotting. Only MMP-9 was overexpressed in bile from patients with CCA compared to benign biliary disease. This 82 kDa protein belongs to the family of zinc and calcium dependent endopeptidases with the combined ability to degrade all components of the extracellular matrix and reported to have an important role in call invasion, metastasis and tumour progression in several cancers including CCA.

In light of previous work in our laboratory demonstrating lipocalin-2 overexpression in bile samples from patients with CCA (Bonney G, unpublished work) and the reported relationship of lipocalin-2 in forming a complex with MMP-9 and protecting it from auto degradation in other gastrointestinal cancers prompted further investigation in CCA. We have confirmed our previous findings of overexpression of lipocalin-2 in CCA and highlighted the presence of the complex in bile from patients with CCA. In humans, high levels of lipocalin-2 (25kDa) are found in epithelial tissues susceptible to infection, sites of inflammation (particularly in response to oxidative stress induced by ROS) and cancer [272]. Our data offers some support to published data linking lipocalin-2 promoting the invasiveness of the CCA cells by forming a complex with MMP-9, stabilizing its activity and rendering the cancer cells to be more invasive [261].

Despite advances in proteomics there remains a lack of successfully identified biomarkers of clinical utility in oncology and specifically in CCA. Proteomic platforms by their very nature require the integration and collaboration of both technology and specialist personnel respectively to facilitate robust, clinically relevant data. All of this requires considerable utilisation of economic and logistical resource and inclusion of high numbers of samples in study designs has to be balanced with this in mind. Bile sample collection remains invasive and as a fluid for proteomic analysis has innate challenges requiring robust processing protocols. Nevertheless continued efforts are required to focus expansion of samples with accurate clinical and staging detail to allow optimal selection for proteomic study design.

Finally, future directions from the work presented here include the assessment of MMP-9, lipocalin-2 and its complex in tissue by immunohistochemistry and ELISA of plasma/serum samples. Indeed immunohistochemistry-based studies have now been initiated in our laboratories and early results have already demonstrated overexpression of MMP-9 and lipocalin-2 in CCA-tumour tissue compared to surrounding non-involved tissue (Nair A, 2013, unpublished work).

7.0 References

1. Khan SA, Thomas HC, Davidson BR, Taylor-Robinson SD. Cholangiocarcinoma. Lancet. 2005;366:1303-1314.

2. Blechacz BR, Gores GJ. Cholangiocarcinoma. Clin Liver Dis. 2008;12:131-50, ix.

3. Kondo S, Takada T, Miyazaki M et al. Guidelines for the management of biliary tract and ampullary carcinomas: surgical treatment. J Hepatobiliary Pancreat Surg. 2008;15:41-54.

4. Chung YE, Kim MJ, Park YN, Lee YH, Choi JY. Staging of extrahepatic cholangiocarcinoma. Eur Radiol. 2008;18:2182-2195.

5. Klatskin G. Adenocarcinoma of the hepatic duct at its bifurcation within the porta hepatis. An unusual timor with distinctive clinical and pathological features. Am J Med. 1965;38:241-256.

6. Nakeeb A, Pitt HA, Sohn TA et al. Cholangiocarcinoma. A spectrum of intrahepatic, perihilar, and distal tumors. Ann Surg. 1996;224:463-73; discussion 473-5.

7. Bismuth H, Nakache R, Diamond T. Management strategies in resection for hilar cholangiocarcinoma. Ann Surg. 1992;215:31-38.

8. Lim JH, Park CK. Pathology of cholangiocarcinoma. Abdom Imaging. 2004;29:540-547.

9. Jarnagin WR, Shoup M. Surgical management of cholangiocarcinoma. Semin Liver Dis. 2004;24:189-199.

10. Ishak KG APP, Sobin LH. Histological typing of tumours of the liver. WHO International Histological Classification of Tumours. 1994.11.   Patel T. Worldwide trends in mortality from biliary tract malignancies. BMC Cancer. 2002;2:10.

12. Shaib Y, El-Serag HB. The epidemiology of cholangiocarcinoma. Semin Liver Dis. 2004;24:115-125.

13. Taylor-Robinson SD, Toledano MB, Arora S et al. Increase in mortality rates from intrahepatic cholangiocarcinoma in England and Wales 1968-1998. Gut. 2001;48:816-820.

14. Khan SA, Toledano MB, Taylor-Robinson SD. Epidemiology, risk factors, and pathogenesis of cholangiocarcinoma. HPB (Oxford). 2008;10:77-82.

15. Khan SA, Taylor-Robinson SD, Toledano MB, Beck A, Elliott P, Thomas HC. Changing international trends in mortality rates for liver, biliary and pancreatic tumours. J Hepatol. 2002;37:806-813.

16. Hammill CW, Wong LL. Intrahepatic cholangiocarcinoma: a malignancy of increasing importance. J Am Coll Surg. 2008;207:594-603.

17. McGlynn KA, Tarone RE, El-Serag HB. A comparison of trends in the incidence of hepatocellular carcinoma and intrahepatic cholangiocarcinoma in the United States. Cancer Epidemiol Biomarkers Prev. 2006;15:1198-1203.

18. Welzel TM, McGlynn KA, Hsing AW, O'Brien TR, Pfeiffer RM. Impact of classification of hilar cholangiocarcinomas (Klatskin tumors) on the incidence of intra- and extrahepatic cholangiocarcinoma in the United States. J Natl Cancer Inst. 2006;98:873-875.

19. Gatto M, Alvaro D. New insights on cholangiocarcinoma. World J Gastrointest Oncol. 2010;2:136-145.

20. Tanaka M, Tanaka H, Tsukuma H, Ioka A, Oshima A, Nakahara T. Risk factors for intrahepatic cholangiocarcinoma: a possible role of hepatitis B virus. J Viral Hepat. 2010;17:742-748.

21. El-Serag HB, Engels EA, Landgren O et al. Risk of hepatobiliary and pancreatic cancers after hepatitis C virus infection: A population-based study of U.S. veterans. Hepatology. 2009;49:116-123.

22. Aljiffry M, Walsh MJ, Molinari M. Advances in diagnosis, treatment and palliation of cholangiocarcinoma: 1990-2009. World J Gastroenterol. 2009;15:4240-4262.

23. Kumar M. Molecular carcinogenesis of hepatocellular carcinoma and intrahepatic cholangiocarcinoma: one step closer to personalised medicine? Cell and Bioscience. 2011;1:1-13.

24. Lazaridis KN, Gores GJ. Primary sclerosing cholangitis and cholangiocarcinoma. Semin Liver Dis. 2006;26:42-51.

25. Mendes F, Lindor KD. Primary sclerosing cholangitis: overview and update. Nat Rev Gastroenterol Hepatol. 2010;7:611-619.

26. Morris-Stiff G, Bhati C, Olliff S et al. Cholangiocarcinoma complicating primary sclerosing cholangitis: a 24-year experience. Dig Surg. 2008;25:126-132.

27. Nashan B, Schlitt HJ, Tusch G et al. Biliary malignancies in primary sclerosing cholangitis: timing for liver transplantation. Hepatology. 1996;23:1105-1111.

28. Burak K, Angulo P, Pasha TM, Egan K, Petz J, Lindor KD. Incidence and risk factors for cholangiocarcinoma in primary sclerosing cholangitis. Am J Gastroenterol. 2004;99:523-526.

29. Jesudian AB, Jacobson IM. Screening and diagnosis of cholangiocarcinoma in patients with primary sclerosing cholangitis. Rev Gastroenterol Disord. 2009;9:E41-7.

30. Abbas G, Lindor KD. Cholangiocarcinoma in primary sclerosing cholangitis. J Gastrointest Cancer. 2009;40:19-25.

31. Wise C, Pilanthananond M, Perry BF, Alpini G, McNeal M, Glaser SS. Mechanisms of biliary carcinogenesis and growth. World J Gastroenterol. 2008;14:2986-2989.

32. Hassid VJ, Orlando FA, Awad ZT et al. Genetic and molecular abnormalities in cholangiocarcinogenesis. Anticancer Res. 2009;29:1151-1156.

33. Xia X, Francis H, Glaser S, Alpini G, LeSage G. Bile acid interactions with cholangiocytes. World J Gastroenterol. 2006;12:3553-3563.

34. Tao LY, He XD, Qu Q et al. Risk factors for intrahepatic and extrahepatic cholangiocarcinoma: a case-control study in China. Liver Int. 2010;30:215-221.

35. Wu TT, Levy M, Correa AM, Rosen CB, Abraham SC. Biliary intraepithelial neoplasia in patients without chronic biliary disease: analysis of liver explants with alcoholic cirrhosis, hepatitis C infection, and noncirrhotic liver diseases. Cancer. 2009;115:4564-4575.

36. de Martel C, Plummer M, Franceschi S. Cholangiocarcinoma: descriptive epidemiology and risk factors. Gastroenterol Clin Biol. 2010;34:173-180.

37. Welzel TM, Graubard BI, El-Serag HB et al. Risk factors for intrahepatic and extrahepatic cholangiocarcinoma in the United States: a population-based case-control study. Clin Gastroenterol Hepatol. 2007;5:1221-1228.

38. Chang SC, Rashid A, Gao YT et al. Polymorphism of genes related to insulin sensitivity and the risk of biliary tract cancer and biliary stone: a population-based case-control study in Shanghai, China. Carcinogenesis. 2008;29:944-948.

39. Funabiki T, Matsubara T, Miyakawa S, Ishihara S. Pancreaticobiliary maljunction and carcinogenesis to biliary and pancreatic malignancy. Langenbecks Arch Surg. 2009;394:159-169.

40. Yamada S, Hosoda S, Tateno H, Kido C, Takahashi S. Survey of thorotrast-associated liver cancers in Japan. J Natl Cancer Inst. 1983;70:31-35.

41. Zhu AX, Lauwers GY, Tanabe KK. Cholangiocarcinoma in association with Thorotrast exposure. J Hepatobiliary Pancreat Surg. 2004;11:430-433.

42. Nyberg U, Nilsson B, Travis LB, Holm LE, Hall P. Cancer incidence among Swedish patients exposed to radioactive thorotrast: a forty-year follow-up survey. Radiat Res. 2002;157:419-425.

43. Greene FL PDL, Fleming ID. AJCC (American Joint Committee on Cancer) cancer staging manual. 2002

44. Jarnagin WR, Fong Y, DeMatteo RP et al. Staging, resectability, and outcome in 225 patients with hilar cholangiocarcinoma. Ann Surg. 2001;234:507-17; discussion 517-9.

45. Kim HJ. [TNM staging of hilar cholangiocarcinoma]. Korean J Gastroenterol. 2005;46:20-27.

46. Otto G, Romaneehsen B, Hoppe-Lotichius M, Bittinger F. Hilar cholangiocarcinoma: resectability and radicality after routine diagnostic imaging. J Hepatobiliary Pancreat Surg. 2004;11:310-318.

47. Bold RJ, Goodnight JEJ. Hilar cholangiocarcinoma: surgical and endoscopic approaches. Surg Clin North Am. 2004;84:525-542.

48. Alvaro D, Macarri G, Mancino MG, 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 Oct 2;147(7):451-9.

49. Wiseman H, Halliwell B. Damage to DNA by reactive oxygen and nitrogen species: role in inflammatory disease and progression to cancer. Biochem J. 1996;313:17-29.

50. Hofseth LJ, Hussain SP, Wogan GN, Harris CC. Nitric oxide in cancer and chemoprevention. Free Radic Biol Med. 2003;34:955-968.

51. Prawan A, Buranrat B, Kukongviriyapan U, Sripa B, Kukongviriyapan V. Inflammatory cytokines suppress NAD(P)H:quinone oxidoreductase-1 and induce oxidative stress in cholangiocarcinoma cells. J Cancer Res Clin Oncol. 2009;135:515-522.

52. Mon NN, Kokuryo T, Hamaguchi M. Inflammation and tumor progression: a lesson from TNF-alpha-dependent FAK signaling in cholangiocarcinoma. Methods Mol Biol. 2009;512:279-293.

53. Pinlaor S, Hiraku Y, Ma N et al. Mechanism of NO-mediated oxidative and nitrative DNA damage in hamsters infected with Opisthorchis viverrini: a model of inflammation-mediated carcinogenesis. Nitric Oxide. 2004;11:175-183.

54. Pinlaor S, Sripa B, Ma N et al. Nitrative and oxidative DNA damage in intrahepatic cholangiocarcinoma patients in relation to tumor invasion. World J Gastroenterol. 2005;11:4644-4649.

55. Harris AL. Hypoxia--a key regulatory factor in tumour growth. Nat Rev Cancer. 2002;2:38-47.

56. Wang D, Dubois RN. The role of COX-2 in intestinal inflammation and colorectal cancer. Oncogene. 2010;29:781-788.

57. Yoon JH, Higuchi H, Werneburg NW, Kaufmann SH, Gores GJ. Bile acids induce cyclooxygenase-2 expression via the epidermal growth factor receptor in a human cholangiocarcinoma cell line. Gastroenterology. 2002;122:985-993.

58. Hayashi N, Yamamoto H, Hiraoka N et al. Differential expression of cyclooxygenase-2 (COX-2) in human bile duct epithelial cells and bile duct neoplasm. Hepatology. 2001;34:638-650.

59. Nzeako UC, Guicciardi ME, Yoon JH, Bronk SF, Gores GJ. COX-2 inhibits Fas-mediated apoptosis in cholangiocarcinoma cells. Hepatology. 2002;35:552-559.

60. Han C, Leng J, Demetris AJ, Wu T. Cyclooxygenase-2 promotes human cholangiocarcinoma growth: evidence for cyclooxygenase-2-independent mechanism in celecoxib-mediated induction of p21waf1/cip1 and p27kip1 and cell cycle arrest. Cancer Res. 2004;64:1369-1376.

61. Goydos JS, Brumfield AM, Frezza E, Booth A, Lotze MT, Carty SE. Marked elevation of serum interleukin-6 in patients with cholangiocarcinoma: validation of utility as a clinical marker. Ann Surg. 1998;227:398-404.

62. Sugawara H, Yasoshima M, Katayanagi K et al. Relationship between interleukin-6 and proliferation and differentiation in cholangiocarcinoma. Histopathology. 1998;33:145-153.

63. Tadlock L, Patel T. Involvement of p38 mitogen-activated protein kinase signaling in transformed growth of a cholangiocarcinoma cell line. Hepatology. 2001;33:43-51.

64. Isomoto H, Kobayashi S, Werneburg NW et al. Interleukin 6 upregulates myeloid cell leukemia-1 expression through a STAT3 pathway in cholangiocarcinoma cells. Hepatology. 2005;42:1329-1338.

65. Wu T, Han C, Lunz JGr, Michalopoulos G, Shelhamer JH, Demetris AJ. Involvement of 85-kd cytosolic phospholipase A(2) and cyclooxygenase-2 in the proliferation of human cholangiocarcinoma cells. Hepatology. 2002;36:363-373.

66. Tanimura Y, Kokuryo T, Tsunoda N et al. Tumor necrosis factor alpha promotes invasiveness of cholangiocarcinoma cells via its receptor, TNFR2. Cancer Lett. 2005;219:205-213.

67. Ohira S, Sasaki M, Harada K et al. Possible regulation of migration of intrahepatic cholangiocarcinoma cells by interaction of CXCR4 expressed in carcinoma cells with tumor necrosis factor-alpha and stromal-derived factor-1 released in stroma. Am J Pathol. 2006;168:1155-1168.

68. Komori J, Marusawa H, Machimoto T et al. Activation-induced cytidine deaminase links bile duct inflammation to human cholangiocarcinoma. Hepatology. 2008;47:888-896.

69. Berthiaume EP, Wands J. The molecular pathogenesis of cholangiocarcinoma. Semin Liver Dis. 2004;24:127-137.

70. Glaser SS, Gaudio E, Alpini G. Vascular factors, angiogenesis and biliary tract disease. Curr Opin Gastroenterol. 2010;26:246-250.

71. Gaudio E, Barbaro B, Alvaro D et al. Administration of r-VEGF-A prevents hepatic artery ligation-induced bile duct damage in bile duct ligated rats. Am J Physiol Gastrointest Liver Physiol. 2006;291:G307-17.

72. Gaudio E, Barbaro B, Alvaro D et al. Vascular endothelial growth factor stimulates rat cholangiocyte proliferation via an autocrine mechanism. Gastroenterology. 2006;130:1270-1282.

73. Mancino A, Mancino MG, Glaser SS et al. Estrogens stimulate the proliferation of human cholangiocarcinoma by inducing the expression and secretion of vascular endothelial growth factor. Dig Liver Dis. 2009;41:156-163.

74. Tang D, Nagano H, Yamamoto H et al. Angiogenesis in cholangiocellular carcinoma: expression of vascular endothelial growth factor, angiopoietin-1/2, thrombospondin-1 and clinicopathological significance. Oncol Rep. 2006;15:525-532.

75. French JJ, Midwinter MJ, Bennett MK, Manas DM, Charnley RM. A matrix metalloproteinase inhibitor to treat unresectable cholangiocarcinoma. HPB (Oxford). 2005;7:289-291.

76. Shirabe K, Shimada M, Kajiyama K et al. Expression of matrix metalloproteinase-9 in surgically resected intrahepatic cholangiocarcinoma. Surgery. 1999;126:842-846.

77. Maeda T, Taguchi K, Aishima S et al. Clinicopathological correlates of aspartyl (asparaginyl) beta-hydroxylase over-expression in cholangiocarcinoma. Cancer Detect Prev. 2004;28:313-318.

78. Shen FZ, Zhang BY, Feng YJ et al. Current research in perineural invasion of cholangiocarcinoma. J Exp Clin Cancer Res. 2010;29:24.

79. Gigliozzi A, Alpini G, Baroni GS et al. Nerve growth factor modulates the proliferative capacity of the intrahepatic biliary epithelium in experimental cholestasis. Gastroenterology. 2004;127:1198-1209.

80. Kim HJ, Kim JS, Kang CD et al. [Expression of epidermal growth factor receptor, ErbB2 and matrix metalloproteinase-9 in hepatolithiasis and cholangiocarcinoma]. Korean J Gastroenterol. 2005;45:52-59.

81. Nakanishi Y, Zen Y, Kondo S, Itoh T, Itatsu K, Nakanuma Y. Expression of cell cycle-related molecules in biliary premalignant lesions: biliary intraepithelial neoplasia and biliary intraductal papillary neoplasm. Hum Pathol. 2008;39:1153-1161.

82. Khan SA, Davidson BR, Goldin R et al. Guidelines for the diagnosis and treatment of cholangiocarcinoma: consensus document. Gut. 2002;51 Suppl 6:VI1-9.

83. Hidalgo E, Asthana S, Nishio H et al. Surgery for hilar cholangiocarcinoma: the Leeds experience. Eur J Surg Oncol. 2008;34:787-794.

84. Ito F, Cho CS, Rikkers LF, Weber SM. Hilar cholangiocarcinoma: current management. Ann Surg. 2009;250:210-218.

85. Morise Z, Sugioka A, Tokoro T et al. Surgery and chemotherapy for intrahepatic cholangiocarcinoma. World J Hepatol. 2010;2:58-64.

86. Veillette G, Castillo CF. Distal biliary malignancy. Surg Clin North Am. 2008;88:1429-47, xi.

87. Rea DJ, Rosen CB, Nagorney DM, Heimbach JK, Gores GJ. Transplantation for cholangiocarcinoma: when and for whom? Surg Oncol Clin N Am. 2009;18:325-37, ix.

88. Kaya M, de Groen PC, Angulo P et al. Treatment of cholangiocarcinoma complicating primary sclerosing cholangitis: the Mayo Clinic experience. Am J Gastroenterol. 2001;96:1164-1169.

89. Wiedmann M, Caca K, Berr F et al. Neoadjuvant photodynamic therapy as a new approach to treating hilar cholangiocarcinoma: a phase II pilot study. Cancer. 2003;97:2783-2790.

90. Charatcharoenwitthaya 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.

91. Conway JD, Mishra G. The role of endoscopic ultrasound in biliary strictures. Curr Gastroenterol Rep. 2008;10:157-162.

92. Hekimoglu K, Ustundag Y, Dusak A et al. MRCP vs. ERCP in the evaluation of biliary pathologies: review of current literature. J Dig Dis. 2008;9:162-169.

93. Slattery JM, Sahani DV. What is the current state-of-the-art imaging for detection and staging of cholangiocarcinoma? Oncologist. 2006;11:913-922.

94. Singh P, Patel T. Advances in the diagnosis, evaluation and management of cholangiocarcinoma. Curr Opin Gastroenterol. 2006;22:294-299.

95. Park MS, Kim TK, Kim KW et al. Differentiation of extrahepatic bile duct cholangiocarcinoma from benign stricture: findings at MRCP versus ERCP. Radiology. 2004;233:234-240.

96. Desa LA, Akosa AB, Lazzara S, Domizio P, Krausz T, Benjamin IS. Cytodiagnosis in the management of extrahepatic biliary stricture. Gut. 1991;32:1188-1191.

97. Temino Lopez-Jurado R, Cacho Acosta G, Arguelles Pintos M, Rodriguez Caravaca G, Lledo Navarro JL, Fernandez Rodriguez C. Diagnostic yield of brush cytology for biliary stenosis during ERCP. Rev Esp Enferm Dig. 2009;101:385-9, 390-4.

98. Gerhards MF, Vos P, van Gulik TM, Rauws EA, Bosma A, Gouma DJ. Incidence of benign lesions in patients resected for suspicious hilar obstruction. Br J Surg. 2001;88:48-51.

99. Malhi H, Gores GJ. Cholangiocarcinoma: modern advances in understanding a deadly old disease. J Hepatol. 2006;45:856-867.

100. Alvaro D. Serum and bile biomarkers for cholangiocarcinoma. Curr Opin Gastroenterol. 2009;25:279-284.

101. Briggs CD, Neal CP, Mann CD, Steward WP, Manson MM, Berry DP. Prognostic molecular markers in cholangiocarcinoma: a systematic review. Eur J Cancer. 2009;45:33-47.

102. Nehls O, Gregor M, Klump B. Serum and bile markers for cholangiocarcinoma. Semin Liver Dis. 2004;24:139-154.

103. Koprowski H, Steplewski Z, Mitchell K, Herlyn M, Herlyn D, Fuhrer P. Colorectal carcinoma antigens detected by hybridoma antibodies. Somatic Cell Genet. 1979;5:957-971.

104. Ong SL, Sachdeva A, Garcea G et al. Elevation of carbohydrate antigen 19.9 in benign hepatobiliary conditions and its correlation with serum bilirubin concentration. Dig Dis Sci. 2008;53:3213-3217.

105. 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.

106. Marrelli D, Caruso S, Pedrazzani C et al. CA19-9 serum levels in obstructive jaundice: clinical value in benign and malignant conditions. Am J Surg. 2009;198:333-339.

107. Locker GY, Hamilton S, Harris J et al. ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. J Clin Oncol. 2006;24:5313-5327.

108. Patel AH, Harnois DM, Klee GG, LaRusso NF, Gores GJ. The utility of CA 19-9 in the diagnoses of cholangiocarcinoma in patients without primary sclerosing cholangitis. Am J Gastroenterol. 2000;95:204-207.

109. Walker SL, McCormick PA. Diagnosing cholangiocarcinoma in primary sclerosing cholangitis: an "evidence based radiology" review. Abdom Imaging. 2008;33:14-17.

110. Albiin N, Smith IC, Arnelo U et al. Detection of cholangiocarcinoma with magnetic resonance spectroscopy of bile in patients with and without primary sclerosing cholangitis. Acta Radiol. 2008;49:855-862.

111. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69:89-95.

112. Hori S, Blanchet JS, McLoughlin J. From prostate-specific antigen (PSA) to precursor PSA (proPSA) isoforms: a review of the emerging role of proPSAs in the detection and management of early prostate cancer. BJU Int. 2012

113. Lievre A, Laurent-Puig P. Genetics: Predictive value of KRAS mutations in chemoresistant CRC. Nat Rev Clin Oncol. 2009;6:306-307.

114. Anderson NL, Anderson NG. Proteome and proteomics: new technologies, new concepts, and new words. Electrophoresis. 1998;19:1853-1861.

115. Dai Z, Zhou J, Qiu SJ, Liu YK, Fan J. Lectin-based glycoproteomics to explore and analyze hepatocellular carcinoma-related glycoprotein markers. Electrophoresis. 2009;30:2957-2966.

116. Chong PK, Lee H, Kong JW, Loh MC, Wong CH, Lim YP. Phosphoproteomics, oncogenic signaling and cancer research. Proteomics. 2008;8:4370-4382.

117. Hu S, Loo JA, Wong DT. Human body fluid proteome analysis. Proteomics. 2006;6:6326-6353.

118. Fung ET. A recipe for proteomics diagnostic test development: the OVA1 test, from biomarker discovery to FDA clearance. Clin Chem. 2010;56:327-329.

119. Banks RE, Stanley AJ, Cairns DA et al. Influences of blood sample processing on low-molecular-weight proteome identified by surface-enhanced laser desorption/ionization mass spectrometry. Clin Chem. 2005;51:1637-1649.

120. Banks RE. Preanalytical influences in clinical proteomic studies: raising awareness of fundamental issues in sample banking. Clin Chem. 2008;54:6-7.

121. Schrohl AS, Wurtz S, Kohn E et al. Banking of biological fluids for studies of disease-associated protein biomarkers. Mol Cell Proteomics. 2008;7:2061-2066.

122. Jackson DH, Banks RE. Banking of clinical samples for proteomic biomarker studies: a consideration of logistical issues with a focus on pre-analytical variation. Proteomics Clin Appl. 2010;4:250-270.

123. Apweiler R, Aslanidis C, Deufel T et al. Approaching clinical proteomics: current state and future fields of application in fluid proteomics. Clin Chem Lab Med. 2009;47:724-744.

124. Dos Santos A, Court M, Thiers V et al. Identification of cellular targets in human intrahepatic cholangiocarcinoma using laser microdissection and accurate mass and time tag proteomics. Mol Cell Proteomics. 2010;9:1991-2004.

125. Ding SJ, Li Y, Tan YX et al. From proteomic analysis to clinical significance: overexpression of cytokeratin 19 correlates with hepatocellular carcinoma metastasis. Mol Cell Proteomics. 2004;3:73-81.

126. Svasti J, Srisomsap C, Subhasitanont P et al. Proteomic profiling of cholangiocarcinoma cell line treated with pomiferin from Derris malaccensis. Proteomics. 2005;5:4504-4509.

127. Liu X, Zhang M, Go VL, Hu S. Membrane proteomic analysis of pancreatic cancer cells. J Biomed Sci. 2010;17:74.

128. Farina A, Dumonceau JM, Lescuyer P. Proteomic analysis of human bile and potential applications for cancer diagnosis. Expert Rev Proteomics. 2009;6:285-301.

129. Tu C, Rudnick PA, Martinez MY et al. Depletion of abundant plasma proteins and limitations of plasma proteomics. J Proteome Res. 2010;9:4982-4991.

130. Gorg A, Weiss W, Dunn MJ. Current two-dimensional electrophoresis technology for proteomics. Proteomics. 2004;4:3665-3685.

131. Tonge R, Shaw J, Middleton B et al. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. Proteomics. 2001;1:377-396.

132. Alban A, David SO, Bjorkesten L et al. A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. Proteomics. 2003;3:36-44.

133. Vasudev NS, Ferguson RE, Cairns DA, Stanley AJ, Selby PJ, Banks RE. Serum biomarker discovery in renal cancer using 2-DE and prefractionation by immunodepletion and isoelectric focusing; increasing coverage or more of the same? Proteomics. 2008;8:5074-5085.

134. Rajcevic U, Niclou SP, Jimenez CR. Proteomics strategies for target identification and biomarker discovery in cancer. Front Biosci. 2009;14:3292-3303.

135. Rabilloud T, Chevallet M, Luche S, Lelong C. Fully denaturing two-dimensional electrophoresis of membrane proteins: a critical update. Proteomics. 2008;8:3965-3973.

136. Veenstra TD, Prieto DA, Conrads TP. Proteomic patterns for early cancer detection. Drug Discov Today. 2004;9:889-897.

137. Walch A, Rauser S, Deininger SO, Hofler H. MALDI imaging mass spectrometry for direct tissue analysis: a new frontier for molecular histology. Histochem Cell Biol. 2008;130:421-434.

138. Chen X, Sun L, Yu Y, Xue Y, Yang P. Amino acid-coded tagging approaches in quantitative proteomics. Expert Rev Proteomics. 2007;4:25-37.

139. Miyagi M, Rao KC. Proteolytic 18O-labeling strategies for quantitative proteomics. Mass Spectrom Rev. 2007;26:121-136.

140. Boehm AM, Putz S, Altenhofer D, Sickmann A, Falk M. Precise protein quantification based on peptide quantification using iTRAQ. BMC Bioinformatics. 2007;8:214.

141. Neilson KA, Ali NA, Muralidharan S et al. Less label, more free: approaches in label-free quantitative mass spectrometry. Proteomics. 2011;11:535-553.

142. Xue H, Lu B, Zhang J et al. Identification of serum biomarkers for colorectal cancer metastasis using a differential secretome approach. J Proteome Res. 2010;9:545-555.

143. Sitek B, Waldera-Lupa DM, Poschmann G, Meyer HE, Stuhler K. Application of label-free proteomics for differential analysis of lung carcinoma cell line A549. Methods Mol Biol. 2012;893:241-248.

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

145. Zhu W, Smith JW, Huang CM. Mass spectrometry-based label-free quantitative proteomics. J Biomed Biotechnol. 2010;2010:840518.

146. America AH, Cordewener JH. Comparative LC-MS: a landscape of peaks and valleys. Proteomics. 2008;8:731-749.

147. Lengqvist J, Andrade J, Yang Y, Alvelius G, Lewensohn R, Lehtio J. Robustness and accuracy of high speed LC-MS separations for global peptide quantitation and biomarker discovery. J Chromatogr B Analyt Technol Biomed Life Sci. 2009;877:1306-1316.

148. Huang SK, Darfler MM, Nicholl MB et al. LC/MS-based quantitative proteomic analysis of paraffin-embedded archival melanomas reveals potential proteomic biomarkers associated with metastasis. PLoS ONE. 2009;4:e4430.

149. Zybailov B, Coleman MK, Florens L, Washburn MP. Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. Anal Chem. 2005;77:6218-6224.

150. Kawase H, Fujii K, Miyamoto M et al. Differential LC-MS-based proteomics of surgical human cholangiocarcinoma tissues. J Proteome Res. 2009;8:4092-4103.

151. Matull WR, Andreola F, Loh A et al. MUC4 and MUC5AC are highly specific tumour-associated mucins in biliary tract cancer. Br J Cancer. 2008;98:1675-1681.

152. Kristiansen TZ, Harsha HC, Gronborg M, Maitra A, Pandey A. Differential membrane proteomics using 18O-labeling to identify biomarkers for cholangiocarcinoma. J Proteome Res. 2008;7:4670-4677.

153. Shimonishi T, Miyazaki K, Kono N et al. Expression of endogenous galectin-1 and galectin-3 in intrahepatic cholangiocarcinoma. Hum Pathol. 2001;32:302-310.

154. Aishima S, Asayama Y, Taguchi K et al. The utility of keratin 903 as a new prognostic marker in mass-forming-type intrahepatic cholangiocarcinoma. Mod Pathol. 2002;15:1181-1190.

155. Yamazaki D, Kurisu S, Takenawa T. Regulation of cancer cell motility through actin reorganization. Cancer Sci. 2005;96:379-386.

156. Craig DH, Haimovich B, Basson MD. Alpha-actinin-1 phosphorylation modulates pressure-induced colon cancer cell adhesion through regulation of focal adhesion kinase-Src interaction. Am J Physiol Cell Physiol. 2007;293:C1862-74.

157. Kikuchi S, Honda K, Tsuda H et al. Expression and gene amplification of actinin-4 in invasive ductal carcinoma of the pancreas. Clin Cancer Res. 2008;14:5348-5356.

158. Vasseur S, Afzal S, Tardivel-Lacombe J, Park DS, Iovanna JL, Mak TW. DJ-1/PARK7 is an important mediator of hypoxia-induced cellular responses. Proc Natl Acad Sci U S A. 2009;106:1111-1116.

159. Podgorski I, Sloane BF. Cathepsin B and its role(s) in cancer progression. Biochem Soc Symp. 2003;263-276.

160. Dufek V, Petrtyl J, Klener P. [Cathepsin B-like substance in the monitoring of chemically induced primary cholangiocarcinoma in swine]. Vnitr Lek. 1996;42:551-554.

161. Iftikhar R, Kladney RD, Havlioglu N et al. Disease- and cell-specific expression of GP73 in human liver disease. Am J Gastroenterol. 2004;99:1087-1095.

162. Varambally S, Laxman B, Mehra R et al. Golgi protein GOLM1 is a tissue and urine biomarker of prostate cancer. Neoplasia. 2008;10:1285-1294.

163. Wang M, Long RE, Comunale MA et al. Novel fucosylated biomarkers for the early detection of hepatocellular carcinoma. Cancer Epidemiol Biomarkers Prev. 2009;18:1914-1921.

164. Piljic A, Schultz C. Annexin A4 self-association modulates general membrane protein mobility in living cells. Mol Biol Cell. 2006;17:3318-3328.

165. Zimmermann U, Balabanov S, Giebel J et al. Increased expression and altered location of annexin IV in renal clear cell carcinoma: a possible role in tumour dissemination. Cancer Lett. 2004;209:111-118.

166. Duncan R, Carpenter B, Main LC, Telfer C, Murray GI. Characterisation and protein expression profiling of annexins in colorectal cancer. Br J Cancer. 2008;98:426-433.

167. Welsch T, Endlich K, Giese T, Buchler MW, Schmidt J. Eps8 is increased in pancreatic cancer and required for dynamic actin-based cell protrusions and intercellular cytoskeletal organization. Cancer Lett. 2007;255:205-218.

168. Tietz P, de Groen PC, Anderson NL et al. Cholangiocyte-specific rat liver proteins identified by establishment of a two-dimensional gel protein database. Electrophoresis. 1998;19:3207-3212.

169. Koopmann J, Thuluvath PJ, Zahurak ML et al. Mac-2-binding protein is a diagnostic marker for biliary tract carcinoma. Cancer. 2004;101:1609-1615.

170. Srisomsap C, Sawangareetrakul P, Subhasitanont P et al. Proteomic analysis of cholangiocarcinoma cell line. Proteomics. 2004;4:1135-1144.

171. Schneider M, Hansen JL, Sheikh SP. S100A4: a common mediator of epithelial-mesenchymal transition, fibrosis and regeneration in diseases? J Mol Med (Berl). 2008;86:507-522.

172. Liu L, Wang J, Liu B et al. Serum levels of variants of transthyretin down-regulation in cholangiocarcinoma. J Cell Biochem. 2008;104:745-755.

173. Wang X, Dai S, Zhang Z et al. Characterization of apolipoprotein A-I as a potential biomarker for cholangiocarcinoma. Eur J Cancer Care (Engl). 2009

174. Sirisinha S, Tengchaisri T, Boonpucknavig S, Prempracha N, Ratanarapee S, A. P. Establishment and characterization of a cholangiocarcinoma cell line from a Thai patient with intrahepatic bile duct cancer. Asian Pac J Allergy Immunol. 1991;9:153-157.

175. Yonglitthipagon P, Pairojkul C, Chamgramol Y, Mulvenna J, Sripa B. Up-regulation of annexin A2 in cholangiocarcinoma caused by Opisthorchis viverrini and its implication as a prognostic marker. Int J Parasitol. 2010;40:1203-1212.

176. Pak JH, Moon JH, Hwang SJ, Cho SH, Seo SB, Kim TS. Proteomic analysis of differentially expressed proteins in human cholangiocarcinoma cells treated with Clonorchis sinensis excretory-secretory products. J Cell Biochem. 2009;108:1376-1388.

177. Shimonishi T, Miyazaki K, Nakanuma Y. Cytokeratin profile relates to histological subtypes and intrahepatic location of intrahepatic cholangiocarcinoma and primary sites of metastatic adenocarcinoma of liver. Histopathology. 2000;37:55-63.

178. 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.

179. He C, Jngst D. Electrophoretic analysis of biliary proteins: application of high resolution two-dimensional polyacrylamide gel electrophoresis with immobilized pH gradient in the first dimension. Electrophoresis. 1996;17:617-619.

180. 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.

181. Stark M, Jornvall H, Johansson J. Isolation and characterization of hydrophobic polypeptides in human bile. Eur J Biochem. 1999;266:209-214.

182. Kristiansen TZ, Bunkenborg J, Gronborg M et al. A proteomic analysis of human bile. Mol Cell Proteomics. 2004;3:715-728.

183. Zhou H, Chen B, Li RX 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.

184. Chen B, Dong JQ, Chen YJ et al. Two-dimensional electrophoresis for comparative proteomic analysis of human bile. Hepatobiliary Pancreat Dis Int. 2007;6:402-406.

185. Guerrier L, Claverol S, Finzi L et al. Contribution of solid-phase hexapeptide ligand libraries to the repertoire of human bile proteins. J Chromatogr A. 2007;1176:192-205.

186. 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.

187. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26:1367-1372.

188. Hubner NC, Bird AW, Cox J et al. Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. J Cell Biol. 2010;189:739-754.

189. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A. 2001;98:5116-5121.

190. Kristiansen TZ MA, Pandey A. Proteomics of human bile. In: Proteomics of Human Bodily Fluids. In: Proteomics of Human Bodily Fluids. 2007;399-414.

191. Albers CJ, Huizenga JR, Krom RA, Vonk RJ, Gips CH. Composition of human hepatic bile. Ann Clin Biochem. 1985;22:129-132.

192. Gallinger S, Harvey PR, Petrunka CN, Ilson RG, Strasberg SM. Biliary proteins and the nucleation defect in cholesterol cholelithiasis. Gastroenterology. 1987;92:867-875.

193. Mullock BM, Shaw LJ, Fitzharris B et al. Sources of proteins in human bile. Gut. 1985;26:500-509.

194. Keulemans YC, Mok KS, de Wit LT, Gouma DJ, Groen AK. Hepatic bile versus gallbladder bile: a comparison of protein and lipid concentration and composition in cholesterol gallstone patients. Hepatology. 1998;28:11-16.

195. Reuben A. Biliary proteins. Hepatology. 1984;4:46S-50S.

196. Rong Y, Jin D, Hou C et al. Proteomics analysis of serum protein profiling in pancreatic cancer patients by DIGE: up-regulation of mannose-binding lectin 2 and myosin light chain kinase 2. BMC Gastroenterol. 2010;10:68.

197. Heywood WE, Madgett TE, Wang D et al. 2D DIGE analysis of maternal plasma for potential biomarkers of Down Syndrome. Proteome Sci. 2011;9:56.

198. Fernandez-Costa C, Calamia V, Fernandez-Puente P, Capelo-Martinez JL, Ruiz-Romero C, Blanco FJ. Sequential depletion of human serum for the search of osteoarthritis biomarkers. Proteome Sci. 2012;10:55.

199. Anderson NL, Polanski M, Pieper R et al. The human plasma proteome: a nonredundant list developed by combination of four separate sources. Mol Cell Proteomics. 2004;3:311-326.

200. Ahmed N, Barker G, Oliva K et al. An approach to remove albumin for the proteomic analysis of low abundance biomarkers in human serum. Proteomics. 2003;3:1980-1987.

201. Bjorhall K, Miliotis T, Davidsson P. Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. Proteomics. 2005;5:307-317.

202. Echan LA, Tang HY, Ali-Khan N, Lee K, Speicher DW. Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. Proteomics. 2005;5:3292-3303.

203. Huang HL, Stasyk T, Morandell S et al. Enrichment of low-abundant serum proteins by albumin/immunoglobulin G immunoaffinity depletion under partly denaturing conditions. Electrophoresis. 2005;26:2843-2849.

204. Smith MP, Wood SL, Zougman A et al. A systematic analysis of the effects of increasing degrees of serum immunodepletion in terms of depth of coverage and other key aspects in top-down and bottom-up proteomic analyses. Proteomics. 2011;11:2222-2235.

205. Qian WJ, Kaleta DT, Petritis BO et al. Enhanced detection of low abundance human plasma proteins using a tandem IgY12-SuperMix immunoaffinity separation strategy. Mol Cell Proteomics. 2008;7:1963-1973.

206. Millioni R, Tolin S, Puricelli L et al. High abundance proteins depletion vs low abundance proteins enrichment: comparison of methods to reduce the plasma proteome complexity. PLoS ONE. 2011;6:e19603.

207. D'Amici GM, Timperio AM, Rinalducci S, Zolla L. Combinatorial Peptide ligand libraries to discover liver disease biomarkers in plasma samples. Methods Mol Biol. 2012;909:311-319.

208. Moritz RL, Clippingdale AB, Kapp EA et al. Application of 2-D free-flow electrophoresis/RP-HPLC for proteomic analysis of human plasma depleted of multi high-abundance proteins. Proteomics. 2005;5:3402-3413.

209. Granger J, Siddiqui J, Copeland S, Remick D. Albumin depletion of human plasma also removes low abundance proteins including the cytokines. Proteomics. 2005;5:4713-4718.

210. Issaq HJ. The role of separation science in proteomics research. Electrophoresis. 2001;22:3629-3638.

211. Barbhuiya MA, Sahasrabuddhe NA, Pinto SM et al. Comprehensive proteomic analysis of human bile. Proteomics. 2011;11:4443-4453.

212. Farina A, Dumonceau JM, Delhaye M et al. A step further in the analysis of human bile proteome. J Proteome Res. 2011;10:2047-2063.

213. Ferguson RE, Carroll HP, Harris A, Maher ER, Selby PJ, Banks RE. Housekeeping proteins: a preliminary study illustrating some limitations as useful references in protein expression studies. Proteomics. 2005;5:566-571.

214. Kolkenbrock H, Hecker-Kia A, Orgel D, Kinawi A, Ulbrich N. Progelatinase B forms from human neutrophils. complex formation of monomer/lipocalin with TIMP-1. Biol Chem. 1996;377:529-533.

215. Malla N, Sjoli S, Winberg JO, Hadler-Olsen E, Uhlin-Hansen L. Biological and pathobiological functions of gelatinase dimers and complexes. Connect Tissue Res. 2008;49:180-184.

216. Kubben FJ, Sier CF, Hawinkels LJ et al. Clinical evidence for a protective role of lipocalin-2 against MMP-9 autodegradation and the impact for gastric cancer. Eur J Cancer. 2007;43:1869-1876.

217. Herszenyi L, Hritz I, Lakatos G, Varga MZ, Tulassay Z. The behavior of matrix metalloproteinases and their inhibitors in colorectal cancer. Int J Mol Sci. 2012;13:13240-13263.

218. Groblewska M, Siewko M, Mroczko B, Szmitkowski M. The role of matrix metalloproteinases (MMPs) and their inhibitors (TIMPs) in the development of esophageal cancer. Folia Histochem Cytobiol. 2012;50:12-19.

219. Giannopoulos G, Pavlakis K, Parasi A et al. The expression of matrix metalloproteinases-2 and -9 and their tissue inhibitor 2 in pancreatic ductal and ampullary carcinoma and their relation to angiogenesis and clinicopathological parameters. Anticancer Res. 2008;28:1875-1881.

220. Okazaki I, Inagaki Y. Novel strategies for hepatocellular carcinoma based on MMPs science. Anticancer Agents Med Chem. 2012;12:753-763.

221. Kirimlioglu H, Turkmen I, Bassullu N, Dirican A, Karadag N, Kirimlioglu V. The expression of matrix metalloproteinases in intrahepatic cholangiocarcinoma, hilar (Klatskin tumor), middle and distal extrahepatic cholangiocarcinoma, gallbladder cancer, and ampullary carcinoma: role of matrix metalloproteinases in tumor progression and prognosis. Turk J Gastroenterol. 2009;20:41-47.

222. Jang JW, Lee JK, Kim SH. Activation of matrix metalloproteinases-9 after photothrombotic spinal cord injury model in rats. J Korean Neurosurg Soc. 2011;50:288-292.

223. Ferroni P, Basili S, Martini F et al. Serum metalloproteinase 9 levels in patients with coronary artery disease: a novel marker of inflammation. J Investig Med. 2003;51:295-300.

224. Shuman Moss LA, Jensen-Taubman S, Stetler-Stevenson WG. Matrix metalloproteinases: changing roles in tumor progression and metastasis. Am J Pathol. 2012;181:1895-1899.

225. Snoek-van Beurden PA, Von den Hoff JW. Zymographic techniques for the analysis of matrix metalloproteinases and their inhibitors. BioTechniques. 2005;38:73-83.

226. Nagase H, Karamanos N. Metalloproteinases in health and disease: challenges and the future prospects. FEBS J. 2011;278:1.

227. Helmersson-Karlqvist J, Akerfeldt T, Gunningberg L, Swenne CL, Larsson A. Serum MMP-9 and TIMP-1 concentrations and MMP-9 activity during surgery-induced inflammation in humans. Clin Chem Lab Med. 2012;50:1115-1119.

228. Laothong U, Pinlaor P, Hiraku Y et al. Protective effect of melatonin against Opisthorchis viverrini-induced oxidative and nitrosative DNA damage and liver injury in hamsters. J Pineal Res. 2010;49:271-282.

229. Itatsu K, Sasaki M, Yamaguchi J et al. Cyclooxygenase-2 is involved in the up-regulation of matrix metalloproteinase-9 in cholangiocarcinoma induced by tumor necrosis factor-alpha. Am J Pathol. 2009;174:829-841.

230. Terada T, Okada Y, Nakanuma Y. Expression of immunoreactive matrix metalloproteinases and tissue inhibitors of matrix metalloproteinases in human normal livers and primary liver tumors. Hepatology. 1996;23:1341-1344.

231. Prakobwong S, Yongvanit P, Hiraku Y et al. Involvement of MMP-9 in peribiliary fibrosis and cholangiocarcinogenesis via Rac1-dependent DNA damage in a hamster model. Int J Cancer. 2010;127:2576-2587.

232. Ridnour LA, Windhausen AN, Isenberg JS et al. Nitric oxide regulates matrix metalloproteinase-9 activity by guanylyl-cyclase-dependent and -independent pathways. Proc Natl Acad Sci U S A. 2007;104:16898-16903.

233. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. Cell. 2010;140:883-899.

234. Lukivskaya O, Patsenker E, Lis R, Buko VU. Inhibition of inducible nitric oxide synthase activity prevents liver recovery in rat thioacetamide-induced fibrosis reversal. Eur J Clin Invest. 2008;38:317-325.

235. Aram G, Potter JJ, Liu X, Torbenson MS, Mezey E. Lack of inducible nitric oxide synthase leads to increased hepatic apoptosis and decreased fibrosis in mice after chronic carbon tetrachloride administration. Hepatology. 2008;47:2051-2058.

236. Mishima T, Tajima Y, Kuroki T et al. Chemopreventative effect of an inducible nitric oxide synthase inhibitor, ONO-1714, on inflammation-associated biliary carcinogenesis in hamsters. Carcinogenesis. 2009;30:1763-1767.

237. London CA, Sekhon HS, Arora V, Stein DA, Iversen PL, Devi GR. A novel antisense inhibitor of MMP-9 attenuates angiogenesis, human prostate cancer cell invasion and tumorigenicity. Cancer Gene Ther. 2003;10:823-832.

238. Somiari SB, Somiari RI, Heckman CM et al. Circulating MMP2 and MMP9 in breast cancer -- potential role in classification of patients into low risk, high risk, benign disease and breast cancer categories. Int J Cancer. 2006;119:1403-1411.

239. Cho YB, Lee WY, Song SY, Shin HJ, Yun SH, Chun HK. Matrix metalloproteinase-9 activity is associated with poor prognosis in T3-T4 node-negative colorectal cancer. Hum Pathol. 2007;38:1603-1610.

240. Zhang QW, Liu L, Chen R et al. Matrix metalloproteinase-9 as a prognostic factor in gastric cancer: a meta-analysis. Asian Pac J Cancer Prev. 2012;13:2903-2908.

241. Reis ST, Leite KR, Piovesan LF et al. Increased expression of MMP-9 and IL-8 are correlated with poor prognosis of Bladder Cancer. BMC Urol. 2012;12:18.

242. Mroczko B, Lukaszewicz-Zajac M, Wereszczynska-Siemiatkowska U et al. Clinical significance of the measurements of serum matrix metalloproteinase-9 and its inhibitor (tissue inhibitor of metalloproteinase-1) in patients with pancreatic cancer: metalloproteinase-9 as an independent prognostic factor. Pancreas. 2009;38:613-618.

243. Li Y, Zhang Y, Zheng Q. Expression of RECK gene and MMP-9 in hilar cholangiocarcinoma and its clinical significance. J Huazhong Univ Sci Technolog Med Sci. 2005;25:552-554.

244. Itatsu K, Zen Y, Yamaguchi J et al. Expression of matrix metalloproteinase 7 is an unfavorable postoperative prognostic factor in cholangiocarcinoma of the perihilar, hilar, and extrahepatic bile ducts. Hum Pathol. 2008;39:710-719.

245. Leelawat K, Sakchinabut S, Narong S, Wannaprasert J. Detection of serum MMP-7 and MMP-9 in cholangiocarcinoma patients: evaluation of diagnostic accuracy. BMC Gastroenterol. 2009;9:30.

246. Jung K. Is serum matrix metalloproteinase 9 a useful biomarker in detection of colorectal cancer? Considering pre-analytical interference that may influence diagnostic accuracy. Br J Cancer. 2008;99:553-4; author reply 555.

247. Veidal SS, Vassiliadis E, Barascuk N et al. Matrix metalloproteinase-9-mediated type III collagen degradation as a novel serological biochemical marker for liver fibrogenesis. Liver Int. 2010;30:1293-1304.

248. Barascuk N, Veidal SS, Larsen L et al. A novel assay for extracellular matrix remodeling associated with liver fibrosis: An enzyme-linked immunosorbent assay (ELISA) for a MMP-9 proteolytically revealed neo-epitope of type III collagen. Clin Biochem. 2010;43:899-904.

249. Bolignano D, Donato V, Lacquaniti A et al. Neutrophil gelatinase-associated lipocalin (NGAL) in human neoplasias: a new protein enters the scene. Cancer Lett. 2010;288:10-16.

250. Bao G, Clifton M, Hoette TM et al. Iron traffics in circulation bound to a siderocalin (Ngal)-catechol complex. Nat Chem Biol. 2010;6:602-609.

251. Alpizar-Alpizar W, Laerum OD, Illemann M et al. Neutrophil gelatinase-associated lipocalin (NGAL/Lcn2) is upregulated in gastric mucosa infected with Helicobacter pylori. Virchows Arch. 2009;455:225-233.

252. Chakraborty S, Kaur S, Guha S, Batra SK. The multifaceted roles of neutrophil gelatinase associated lipocalin (NGAL) in inflammation and cancer. Biochim Biophys Acta. 2012;1826:129-169.

253. Makris K, Rizos D, Kafkas N, Haliassos A. Neurophil gelatinase-associated lipocalin as a new biomarker in laboratory medicine. Clin Chem Lab Med. 2012;50:1519-1532.

254. Ronco C, Cruz D, Noland BW. Neutrophil gelatinase-associated lipocalin curve and neutrophil gelatinase-associated lipocalin extended-range assay: a new biomarker approach in the early diagnosis of acute kidney injury and cardio-renal syndrome. Semin Nephrol. 2012;32:121-128.

255. Maisel AS, Mueller C, Fitzgerald R et al. Prognostic utility of plasma neutrophil gelatinase-associated lipocalin in patients with acute heart failure: the NGAL EvaLuation Along with B-type NaTriuretic Peptide in acutely decompensated heart failure (GALLANT) trial. Eur J Heart Fail. 2011;13:846-851.

256. Zhang PX, Zhang FR, Xie JJ et al. Expression of NGAL and NGALR in human embryonic, fetal and normal adult tissues. Mol Med Report. 2012;6:716-722.

257. Bauer M, Eickhoff JC, Gould MN, Mundhenke C, Maass N, Friedl A. Neutrophil gelatinase-associated lipocalin (NGAL) is a predictor of poor prognosis in human primary breast cancer. Breast Cancer Res Treat. 2008;108:389-397.

258. Zhang XF, Zhang Y, Zhang XH et al. Clinical significance of Neutrophil gelatinase-associated lipocalin(NGAL) expression in primary rectal cancer. BMC Cancer. 2009;9:134.

259. Zhang Y, Fan Y, Mei Z. NGAL and NGALR overexpression in human hepatocellular carcinoma toward a molecular prognostic classification. Cancer Epidemiol. 2012;36:e294-9.

261. Nuntagowat C, Leelawat K, Tohtong R. NGAL knockdown by siRNA in human cholangiocarcinoma cells suppressed invasion by reducing NGAL/MMP-9 complex formation. Clin Exp Metastasis. 2010;27:295-305.

262. Srisomsap C, Sawangareetrakul P, Subhasitanont P et al. Proteomic studies of cholangiocarcinoma and hepatocellular carcinoma cell secretomes. J Biomed Biotechnol. 2010;2010:437143.

263. Leelawat K, Narong S, Wannaprasert J, Leelawat S. Serum NGAL to Clinically Distinguish Cholangiocarcinoma from Benign Biliary Tract Diseases. Int J Hepatol. 2011;2011:873548.

264. Cai L, Borowiec J, Xu S, et al. Assays of urine levels of HNL/NGAL in patients undergoing cardiac surgery and the impact of antibody configuration on theri clinical performance. Clin Chim Acta. 2009;403:121-5

265. Kjeldsen L, Koch C, Arnljots K, Borregaard N. Charecterisation of two ELISAs for NGAL , a newly described lipocalin in human neutrophils. J Immunol Methods. 1996:198:155-164

266. Kift R, Messenger MP, Wind T et al. A comparison of the analytical performance of five commercially available assays for neutrophil gelatinase-associated lipocalin using urine. Ann Clin Biochem. 2013;50:236-244

267. Zhang H, Xu L, Xiao D et al. Upregulation of neutrophil gelatinase-associated lipocalin in oesophageal squamous cell carcinoma: significant correlation with cell differentiation and tumour invasion. J Clin Pathol. 2007;60:555-561.

268. Fernandez CA, Yan L, Louis G, Yang J, Kutok JL, Moses MA. The matrix metalloproteinase-9/neutrophil gelatinase-associated lipocalin complex plays a role in breast tumor growth and is present in the urine of breast cancer patients. Clin Cancer Res. 2005;11:5390-5395.

269. Smith ER, Zurakowski D, Saad A, Scott RM, Moses MA. Urinary biomarkers predict brain tumor presence and response to therapy. Clin Cancer Res. 2008;14:2378-2386.

270. Yan L, Borregaard N, Kjeldsen L, Moses MA. The high molecular weight urinary matrix metalloproteinase (MMP) activity is a complex of gelatinase B/MMP-9 and neutrophil gelatinase-associated lipocalin (NGAL). Modulation of MMP-9 activity by NGAL. J Biol Chem. 2001;276:37258-37265.

271. Barbhuiya MA, Sahasrabuddhe NA, Pinto SM et al. Comprehensive proteomic analysis of human bile. Proteomics. 2011;11:4443-4453.

272. Moniaux N, Chakraborty S, Yalniz M et al. Early diagnosis of pancreatic cancer: neutrophil gelatinase-associated lipocalin as a marker of pancreatic intraepithelial neoplasia. Br J Cancer. 2008;98:1540-1547.

Appendix 1 – Contact details for the suppliers of the chemicals, consumables and equipment

Agilent Technologies UK Ltd
610 Wharfedale Road
IQ Winnersh
Wokingham
Berkshire
RG41 5TP

Applied Biosystems
120 Birchwood Boulevard
Warrington
Cheshire
WA3 7QH

BD Biosciences
The Danby Building
Edmund Halley Road
Oxford Science Park
Oxford
OX4 4DQ

Biotage AB
Box 8
SE-751 03 Uppsala
Sweden

BMG Labtech Ltd
Unit 5 Merlin Centre
Gatehouse Close
Aylesbury
Bucks
HP19 8DP

Dako
Cambridge House
St Thomas Place
Ely
Cambridgeshire
CB7 4EX

GE Healthcare
Amersham Place
Little Chalfont
Bucks
HP7 9NA

ICN Pharmaceuticals Ltd
Cedarwood Chineham Business Park
Crockford Lane
Chineham
Basingstoke
Hants
RG24 8WG

Invitrogen

3 Fountain Drive

Inchinnan Business Park

Paisley

PA4 9RF

Janke & Kunkel

Str-10

D-79219

Staufen

Germany

Merck

Frankfurter Straße 250

64293 Darmstadt

Germany

MRC-Holland

Willem Schoutenstraat 6

1057 DN Amsterdam

The Netherlands

Neomarkers

47790 Westinghouse Drive

Fremont

California 94539

United States

Nonlinear Dynamics Limited

Keel House

Garth Heads

Newcastle upon Tyne

NE1 2JE

Pierce Biotechnology Ltd

Century House

High Street

Tattenhall

Cheshire

CH3 9RJ

Promega UK

Delta House

Southampton Science Park

Southampton

SO16 7NS

Qiagen

House Fleming Way

Crawley

West Sussex

RH10 9NQ

Roche Diagnostics

Charles Avenue

Burgess Hill

West Sussex

RH15 9RY

Sigma-Aldrich Company Ltd

Fancy Rd

Poole

Dorset

BH12 4QH

Soft Genetics

100 Oakwood Ave

Suite 350

State College

PA 16803

USA

Thermo Shandon

171 Industry Drive

Pittsburgh

Pennsylvania

15275

USA

VWR International LTD

Merck House

Seldown Lane

Poole

Dorset

BH15 1TD

Appendix 2 - Buffer Recipes

<u>Laemmli Buffer</u>

62.5mM Tris-HCl pH 6.8

10% (v/v) glycerol

5% (v/v) β-mercaptoethanol

2% (w/v) SDS

A trace of bromophenol blue

<u>DIGE lysis buffer</u>

7 M urea

2 M thiourea

4% (w/v) CHAPS)

<u>SDS-PAGE running buffer</u>

25 mM Tris pH 8.3

192 mM glycine

0.1% (w/v) SDS

<u>Towbin's transfer buffer</u>

25 mM Tris

192 mM glycine

10% (v/v) methanol

pH 8.3

DIGE buffer

7 M urea

2 M thiourea

4% w/v CHAPS

2% w/v DTT

1.6% v/v pharmalytes pH 3-10

Reswell buffer

7 M urea

2 M thiourea

4% w/v CHAPS

0.46% w/v (30 mM) DTT

0.2% v/v pharmalytes pH 3-10

A trace of bromophenol blue

Equilibration buffer

0.05 M Tris-HCl pH 6.8

6 M urea

30% (v/v) glycerol

2% (v/v) SDS

1% (w/v) DTT

A trace of bromophenol blue

Appendix 3 Sample Details

Study Number	Disease	Date of sample	Age	Sex
8389	Benign	04/09/2006	74	F
8392	Benign	05/09/2006	84	F
9059	Benign	06/09/2007	62	F
9074	Benign	20/09/2007	21	F
9741	Benign	28/10/2008	87	F
9756	Benign	31/10/2008	81	F
10844	Benign	15/06/2009	36	F
11162	Benign	18/09/2009	82	F
11517	Benign	08/12/2009	88	F
11599	Benign	07/01/2010	85	F
11600	Benign	07/01/2010	75	F
8386	Benign	30/03/2006	63	M
8395	Benign	07/09/2006	63	M
8397	Benign	07/09/2006	57	M
9044	Benign	23/08/2007	67	M
9066	Benign	12/09/2007	70	M
9380	Benign	09/06/2008	63	M
9382	Benign	09/06/2008	59	M
9847	Benign	17/11/2008	18	M
10310	Benign	12/02/2009	84	M
10312	Benign	12/02/2009	45	M
11515	Benign	08/12/2009	45	M
11587	Benign	04/01/2010	82	M
11588	Benign	04/01/2010	90	M
11590	Benign	04/01/2010	69	M
8396	CCA	07/09/2006	67	F
8484	CCA	01/11/2006	77	F
8620	CCA	05/12/2006	74	F
9075	CCA	20/09/2007	54	F
9402	CCA	10/06/2008	85	F
9416	CCA	17/06/2008	72	F
9417	CCA	18/06/2008	65	F
9432	CCA	01/07/2008	56	F
9494	CCA	11/08/2008	59	F
10319	CCA	13/01/2009	61	F
10325	CCA	17/02/2009	66	F
10938	CCA	14/07/2009	64	F
11166	CCA	11/09/2009	55	F
8431	CCA	14/09/2006	56	M
8445	CCA	28/04/2006	73	M
8570	CCA	14/12/2006	77	M
8600	CCA	23/12/2006	66	M
8619	CCA	26/01/2007	57	M
8628	CCA	26/01/2007	77	M
8662	CCA	07/02/2007	65	M
9037	CCA	22/08/2007	67	M
9089	CCA	10/10/2007	72	M
9369	CCA	21/05/2008	52	M
9851	CCA	17/11/2008	75	M

10092	CCA	17/12/2008	75	M
10408	CCA	03/03/2009	76	M
10569	CCA	06/04/2009	71	M
10726	CCA	18/05/2009	76	M
10873	CCA	23/06/2009	72	M
10874	CCA	23/06/2009	76	M
11174	CCA	16/09/2009	61	M
11469	CCA	26/11/2009	84	M
11201	CCA	25/09/2009	62	M
11451	CCA	18/11/2009	73	F
11543	CCA	15/12/2009	72	F
11455	CCA	20/11/2009	46	M
11589	CCA	04/01/2010	84	M
11027	LIVER DONOR	06/08/2009	35	F
11128	LIVER DONOR	27/08/2009	53	F
8849	LIVER DONOR	18/05/2007	54	F
8940	LIVER DONOR	25/07/2007	57	F
9039	LIVER DONOR	23/08/2007	34	F
9078	LIVER DONOR	25/02/2007	42	M
9097	LIVER DONOR	16/11/2007	48	M
9241	LIVER DONOR	22/02/2007	45	M
9082	PSC	29/09/2007	43	F
9258	PSC	19/03/2008	56	F
9260	PSC	26/03/2008	45	F
9262	PSC	26/03/2008	64	F
9376	PSC	04/06/2008	49	F
11373	PSC	28/10/2009	20	F
11513	PSC	07/12/2009	73	F
9246	PSC	12/03/2008	39	M
9247	PSC	12/03/2008	48	M
9251	PSC	14/03/2008	51	M
9261	PSC	26/03/2008	54	M
9263	PSC	26/03/2008	54	M
9363	PSC	13/05/2008	25	M
9367	PSC	21/05/2008	66	M
9368	PSC	21/05/2008	42	M
9473	PSC	22/07/2008	40	M
9890	PSC	27/11/2008	24	M
11188	PSC	21/09/2009	47	M
11330	PSC	16/10/2009	66	M
11374	PSC	28/10/2009	71	M
11512	PSC	07/12/2009	61	M
11516	PSC	08/12/2009	60	M
9164	PSC	23/01/2008	50	M
10498	PSC/CCA	17/03/2009	47	M
10601	PSC/CCA	20/04/2009	26	M

DATASET BRIEF

Shotgun proteomics of human bile in hilar cholangiocarcinoma

Shahid G. Farid[1,2], Rachel A. Craven[2], Jianhe Peng[2], Glenn K. Bonney[1,2], David N. Perkins[2], Peter J. Selby[2], K. Rajendra Prasad[1] and Rosamonde E. Banks[2]

[1] Department of Hepatopancreaticobiliary Surgery, St James's University Hospital, Beckett Street, Leeds, UK
[2] Clinical and Biomedical Proteomics Group, Cancer Research UK Centre, Leeds Institute of Molecular Medicine, St James's University Hospital, Leeds, UK

The need to find biomarkers for hepatobiliary diseases including cholangiocarcinoma (CCA) has led to an interest in using bile as a proximal fluid in biomarker discovery experiments, although there are inherent challenges both in its acquisition and analysis. The study described here greatly extends previous studies that have started to characterise the bile proteome. Bile from four patients with hilar CCA was depleted of albumin and immunoglobulin G and analysed by GeLC-MS/MS. The number of proteins identified per bile sample was between 378 and 741. Overall, the products of 813 unique genes were identified, considerably extending current knowledge of the malignant bile proteome. Of these, 268 were present in at least 3 out of 4 patients. This data set represents the largest catalogue of bile proteins to date and together with other studies in the literature constitutes an important prelude to the potential promise of expression proteomics and subsequent validation studies in CCA biomarker discovery.

Cholangiocarcinomas (CCAs) are neoplasms arising from the bile ducts and are classified according to the involvement of the intrahepatic and/or extrahepatic (hilar or distal) portions. Overall, CCA accounts for 3% of all gastrointestinal malignancies and hilar CCA accounts for more than 60% of cases [1]. At present, surgical resection or transplantation in selected cases remains the mainstay for treatment, but only a third of patients are amenable to this potentially curative intervention [2]. Unfortunately, survival outcomes have not changed significantly in the last 30 years and even after R0 resection, that is, with a microscopically clear resection margin, 5-year survival is around 20–40% in most series [3]. Factors responsible for the poor outcome

associated with CCA include the limitations of diagnostic modalities, with currently available biomarkers lacking sensitivity and specificity to facilitate early detection of disease. Development of better biomarkers would facilitate earlier surgical intervention thereby improving prognosis.

Proteomics provides a powerful approach for the identification of biomarkers with potential clinical utility. Studies of serum and plasma can be used for the discovery of such molecules, but this approach is significantly limited by the dynamic range of protein expression, which exceeds >10 orders of magnitude, with proteins originating from tumour cells being of relatively low abundance. One approach to overcome this problem, at least in part, is to use more proximal fluids that would be likely to represent an enriched source of tumour-derived proteins, thereby improving their chance of detection. In the case of CCA, bile would be predicted to be a good source of biomarkers shed or secreted by malignant biliary epithelium [4]. However, a number of factors make the study of bile in biomarker discovery studies a challenge. First, there are difficulties with collection of suitably large sample banks from patients with CCA and

Correspondence: Professor Rosamonde E. Banks, Clinical and Biomedical Proteomics Group, Cancer Research UK Centre, Leeds Institute of Molecular Medicine, St James's University Hospital, Leeds, LS9 7TF, UK
E-mail: r.banks@leeds.ac.uk
Fax: +44-113-242-9886

Abbreviation: CCA, cholangiocarcinoma

appropriate controls such as patients with benign disease or the predisposing condition primary sclerosing cholangitis (PSC). Collection of bile is invasive, being possible during endoscopic retrograde cholangiopancreaticography (ERCP) or open surgery. Second, bile is a complex fluid with protein comprising a relatively minor component compared with other constituents such as bile salts and lipids, which can interfere with protein separation and analysis.

A number of studies have started to characterise the bile proteome, generating a body of data that will underpin future studies using bile in comparative analyses to look for disease biomarkers [5]. In one of the first large-scale efforts to generate an inventory of bile proteins, unfractionated bile from a patient with CCA was analysed by GeLC-MS/MS and 59 unique proteins were identified with a further 28 being found by incorporating pre-fractionation by lectin affinity chromatography using Con A and WGA [6]. More recent studies have extended this catalogue of human bile proteins. Analysis of bile from a cholesterol stone patient by 2-D PAGE and 2-D-LC-MS/MS identified a total of 222 proteins, 98 of which were identified based on more than one tryptic peptide [7]; strikingly, 114 of these proteins were predicted to have a signal peptide. Using hexapeptide ligand libraries to concentrate less abundant protein species, 222 gene products were identified, 143 of which had not been previously reported in bile [8]. Finally, in a study examining biliary stenosis caused by pancreatic adenocarcinoma, GeLC-MS/MS of bile identified 127 proteins (of which 34 were novel) based on more than one tryptic peptide; of these, >80% were intracellular, most likely reflecting that this study profiled both the pellet and supernatant generated by centrifugation of bile [9]. The small number of studies to date which often involve single patients, combined with different sampling techniques and processing protocols together with the variety of pathological conditions studied, explains in part the differences between the proteins identified; however, taken together, approaching 300 proteins have been identified in these four major studies published to date [5]. The study described here markedly extends these data sets, focusing on characterisation of the bile proteome in four patients with CCA.

Bile was collected from patients undergoing diagnostic/therapeutic ERCP intervention with ethics approval (REC 06/Q1206/136) and informed consent. After placement of the catheter, 5–20 mL Omnipaque dye (GE Healthcare, Amersham, UK) was inserted to confirm its position within the bile duct. An equal volume of fluid to dye injected was discarded to minimise contamination, before bile was collected into a sterile syringe and transferred to a Falcon tube. Bile was transported to the processing laboratory on ice within 15 min of collection, centrifuged (13 000 × g for 15 min at 4°C) aliquoted and stored at −80°C until analysis. To potentially improve the detection of lower abundance proteins, two major bile constituents, albumin and immunoglobulin G, were largely removed using the Albumin/IgG Removal Kit (GE Healthcare) following the manufacturer's

instructions. Desalting and concentration of bile samples was carried out by trichloroacetic acid (TCA) precipitation. In brief, an equal volume of 20% w/v TCA was added to bile (typically a volume equivalent to 150 µg protein), the mixture was allowed to precipitate on ice for 30 min before centrifugation at 18 000 × g for 10 min at 4°C. The pellet was washed with ice-cold acetone, allowed to dry and resuspended in DIGE lysis buffer (7 M urea, 2 M thiourea, 4% w/v CHAPS). Protein concentration was determined by densitometric analysis of InstantBlue Coomassie® stained gels (Expedeon, Cambridge, UK) of proteins separated by SDS-PAGE using a standard curve formed using a serial dilution of albumin-depleted serum (ADS) of known concentration. Standard protein assays were not adopted as they gave inconsistent results or did not show a linear trend upon dilution of the bile samples being measured.

Protein (100 µg) from each sample was separated on 10% SDS-PAGE gels using the Hoefer SE600X Chroma Deluxe Electrophoresis Unit. Gels were stained with InstantBlue Coomassie® and each lane was divided into 53 gel slices, which were reduced with DTT, alkylated with iodoacetamide and digested with trypsin. Extracted peptides were analysed using an Agilent 1100 Series nano-LC System (Agilent Technologies, South Queensferry, UK) coupled online with a QSTAR XL quadrupole TOF hybrid mass spectrometer (Applied Biosystems) as previously described [10]. An MS scan from 400 to 1800 m/z was performed for 1 s; the three most abundant doubly and triply charged ions (m/z 400–1000) with intensities over 40 counts were selected for MS/MS analysis, which was acquired from 80 to 1800 m/z for 1 s in the Enhance All mode and precursors were then excluded for 200 s. The MS/MS data was processed by Analyst (version 2.0, Applied Biosystems) and searched using a local Mascot search engine (version 2.3, Matrix Science, London, UK) with the following parameters – database: IPI human (89 652 sequences, version 3.74); enzyme: trypsin; fixed modification: carbamindomethyl (C); variable modification: oxidation (M), deamidated (NQ); peptide mass tolerance: ±0.15 Da; fragment mass tolerance: ±0.1 Da; maximum missed cleavages: 1; instrument type: ESI-QUAD-TOF. The false-positive discovery rate was determined as 1.81% by searching against a decoy database. The Mascot Dat files have been submitted to the PRIDE database (PRIDE Converter 19587657; accession number 15872) and are available at proteomics.leeds.ac.uk/supplementary_data/bile. A summary of the dat files for each patient is provided as Supporting Information Table 1.

For each patient, the intermediate files produced by Mascot were combined, peptides with probability scores with $p > 0.05$ were excluded and redundancy then removed. Proteins required at least one unique significant peptide to be considered identified. Details of the proteins and peptides identified are supplied as Supporting Information Tables 2.1–2.4. For proteins identified with a single peptide, spectra were inspected manually and only those passing this quality control were
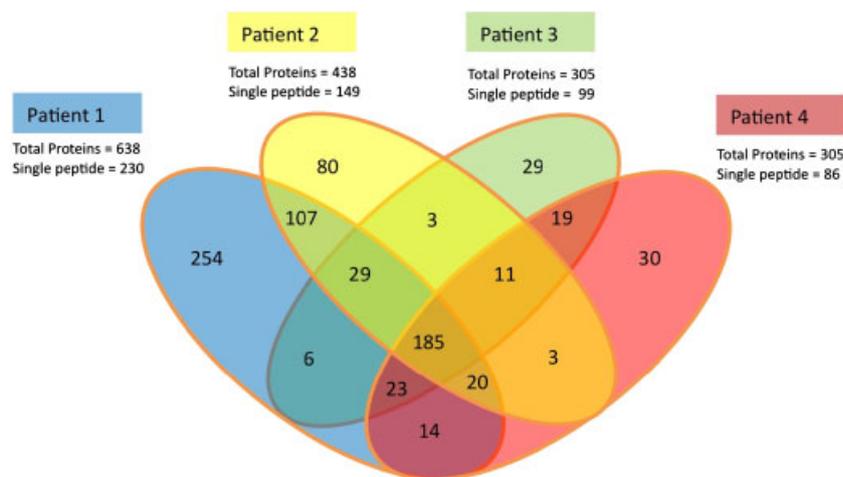
Figure 1. Venn diagram illustrating the number of proteins in each patient and degree of overlap between individual patients. A total number of 813 unique gene products were identified.
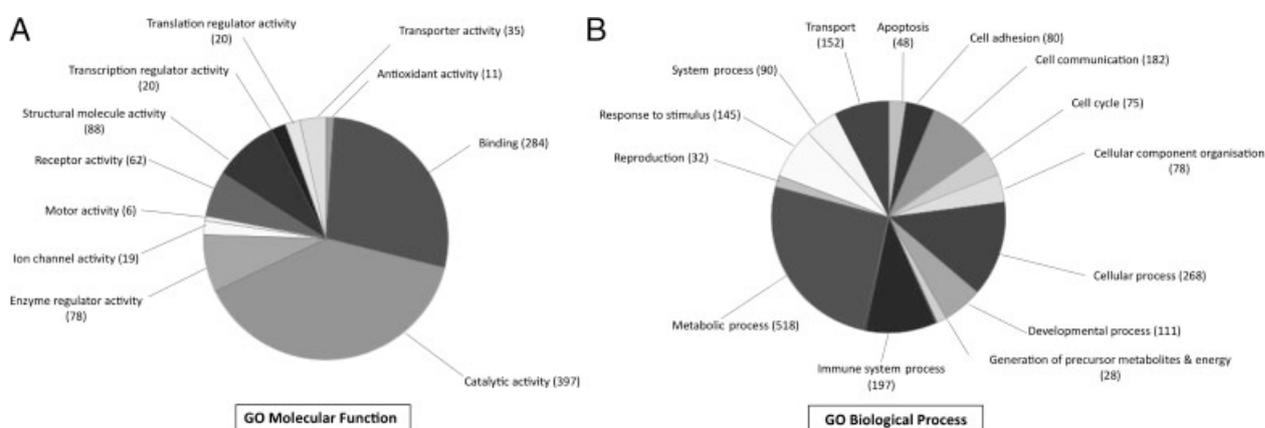

Figure 2. Diagram showing the assignment of gene ontology terms to the proteins identified in bile: (A) molecular function and (B) biological process.

included; these are available at proteomics.leeds.ac.uk/data/supplementary_data/bile. Overall 3576, 2126, 2162 and 2298 peptides corresponding to 741, 509, 378 and 389 proteins were identified in the four bile samples from patients with CCA, with 491, 350, 268 and 295 proteins being identified by at least two significant peptides, respectively. Following removal of keratins and immunoglobulins, redundancy was further reduced by taking the conservative approach of collapsing the data set down to the gene level (that is, removing the potential complexity arising from the presence of different or multiple forms of a particular protein in a particular bile sample). For protein entries with no gene, peptides were searched against MSDB (proteomics.leeds.ac.uk) to find this missing data; a small number of entries for which this was not possible were not considered in downstream analysis. Following this data reduction, products from a total of 813 unique genes were identified in the four patients, with 185 being present in 4, 83 in 3 and 152 in 2 out of the 4 samples and 393 being unique to a single patient. A summary of this data is available as

Supporting Information Table 3). The degree of overlap between samples is shown in Fig. 1.

These gene products were analysed using the Gene Ontology database version 7.0 (www.pantherdb.org/) and Ingenuity Pathway Analysis®. Cellular component analysis identified the majority of proteins to be cytoplasmic in origin (55%), with extracellular space, membrane and nucleus accounting for 20, 10 and 8%, respectively. The terms molecular function and biological processes were explored and in total 772 genes were annotated. Proteins involved in catalytic activity were the most common when looking at molecular function followed by protein binding and structural activity (Fig. 2A). In biological processes, proteins associated with metabolism, cellular and immune processes and cell communication were dominant (Fig. 2B).

As expected, there was considerable overlap between previous studies and the proteins identified in our data set; however, the catalogue reported here extends the biliary proteome considerably with the largest published study

Table 1. Proteins present in current bile dataset which have been reported previously to be up or down regulated in CCA tissue or fluids (from IPA)

Gene name	Protein	Location	Functions	Up/down expression in CCA	No. of patients in data set
A1BG	α-1-B glycoprotein	Extracellular	Unknown function	Down	P1, P2, P3, P4
ADH1B	Alcohol dehydrogenase 1B	Cytoplasm	Zinc ion binding	Down	P2
ADH1C	Alcohol dehydrogenase 1C	Cytoplasm	Oxidoreductase activity; metal ion binding	Down	P1, P2
AMBP	α-1-microglobulin	Extracellular	Transporter activity; calcium channel inhibitor activity	Down	P1, P2, P3, P4
ANXA1	Annexin A1	Extracellular	Receptor/ligand binding	Up	P1, P2, P3, P4
ANXA2	Annexin A2	Extracellular	Receptor/ligand binding	Up	P1, P2, P3, P4
C1S	Complement component 1 S	Extracellular	Serine-type endopeptidase activity	Down	P1, P2, P3, P4
CD14	CD14 antigen, lipopolysaccharide receptor	Extracellular	Lipopolysaccharide binding; opsonin receptor activity	Down	P1, P4
DMBT	Deleted in malignant brain tumours 1	Plasma membrane	Scavenger receptor activity	Up	P1, P2, P3, P4
FGA	Fibrinogen α chain	Extracellular	Receptor/protein binding	Down	P1, P2, P3, P4
GC	Group-specific component (vitamin D binding protein)	Extracellular	Actin/vit D binding	Down	P1, P2, P3, P4
GNB2L1	Guanine nucleotide binding protein (G protein)	Cytoplasm	Receptor/protein binding	Up	P1
HPX	Hemopexin	Extracellular	Heme transporter activity	Down	P1, P2, P3, P4
HSP90AA1	Heat shock protein 90-α	Cytoplasm	Nucleotide/protein binding	Up	P1, P2, P3, P4
HSP90AB1	Heat shock protein 90-β	Cytoplasm	Nucleotide/protein binding	Up	P1, P2, P4
MUC1	Mucin-1	Plasma membrane	Pathogen-binding, cell signalling	Up	P1, P2, P3, P4
KNG1	Kininogen-1	Extracellular	Peptidase inhibitor activity/receptor binding	Down	P1, P2, P3, P4
NGAL	Neutrophil gelatinase-associated lipocalin	Extracellular	Ligand binding	Up	P1, P2, P3, P4
PGK1	Phosphoglycerate kinase 1	Cytoplasm	Nucleotide binding, phosphorylation	Up	P1, P2, P3, P4
PKM2	Pyruvate kinase M2	Nucleus	Nucleotide binding, apoptosis	Up	P1, P2, P3, P4
RPLP0	Ribosomal phosphoprotein P0	Cytoplasm	RNA binding	Up	P1
SERPINA6	Serpin peptidase inhibitor, clade A member 6	Extracellular	Serine-type endopeptidase inhibitor activity	Down	P1, P2, P3, P4
SERPINC1	Serpin peptidase inhibitor, clade C (antithrombin), member 1	Extracellular	Protease binding	Down	P1, P2, P3, P4
VIM	Vimentin	Cytoplasm	Structural constituent of cytoskeleton, protein kinase binding	Up	P1, P2, P3, P4

reporting 222 proteins [7, 8]. As has been noted previously, a significant proportion of these proteins are similar to major abundant proteins found in plasma [11] but include known CCA-associated proteins including MUC 1 [12], neutrophil gelatinase-associated lipocalin (NGAL) [13] and vimentin [14, 15] confirming the utility of analysis of proximal fluids such as bile. A preliminary Ingenuity Pathway Analysis of proteins reported in our data set has identified a number of proteins previously reported to be differentially expressed in studies of tissue, cell lines, serum and bile in CCA (Table 1). There was significant variability between the bile samples analysed here in terms of the number and identity of proteins, probably reflecting individual and disease heterogeneity, the level of sample contamination as well as undersampling, which is a characteristic of the shotgun strategy adopted.

In summary, we present a comprehensive catalogue of bile proteins in four patients with hilar CCA and have considerably extended current knowledge of the malignant bile proteome. Profiling studies such as the one described here have formed the basis for examining the expression of a number of proteins in bile in relation to disease, as illustrated by the work taken forward on Mac2-binding protein, which was identified as a bile protein in a shotgun study of bile from a CCA patient [6] and subsequently shown to have some potential as a biomarker [16]. The current data set together with other studies in the literature constitute an important prelude highlighting the potential promise of comparative quantitative proteomic studies in CCA biomarker discovery.

The authors have declared no conflict of interest.

References

[1] Khan, S. A., Thomas, H. C., Davidson, B. R., Taylor-Robinson, S. D., Cholangiocarcinoma. *Lancet* 2005, *366*, 1303–1314.

[2] Hidalgo, E., Asthana, S., Nishio, H., Wyatt, J. et al., Surgery for hilar cholangiocarcinoma: the Leeds experience. *Eur. J. Surg. Oncol.* 2008, *34*, 787–794.

[3] Ito, F., Cho, C. S., Rikkers, L. F., Weber, S. M., Hilar cholangiocarcinoma: current management. *Ann. Surg.* 2009, *250*, 210–218.

[4] Bonney, G. K., Craven, R. A., Prasad, R., Melcher, A. F. et al., Circulating markers of biliary malignancy: opportunities in proteomics? *Lancet Oncol.* 2008, *9*, 149–158.

[5] Farina, A., Dumonceau, J. M., Lescuyer, P., Proteomic analysis of human bile and potential applications for cancer diagnosis. *Expert Rev. Proteomics* 2009, *6*, 285–301.

[6] Kristiansen, T. Z., Bunkenborg, J., Gronborg, M., Molina, H. et al., A proteomic analysis of human bile. *Mol. Cell. Proteomics* 2004, *3*, 715–728.

[7] Zhou, H., Chen, B., Li, R. X., Sheng, Q. H. 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.

[8] Guerrier, L., Claverol, S., Finzi, L., Paye, F. et al., Contribution of solid-phase hexapeptide ligand libraries to the repertoire of human bile proteins. *J. Chromatogr. A* 2007, *1176*, 192–205.

[9] Farina, A., Dumonceau, J. M., Frossard, J. L., Hadengue, A. et al., Proteomic analysis of human bile from malignant biliary stenosis induced by pancreatic cancer. *J. Proteome Res.* 2009, *8*, 159–169.

[10] Aggelis, V., Craven, R. A., Peng, J., Harnden, P. et al., Proteomic identification of differentially expressed plasma membrane proteins in renal cell carcinoma by stable isotope labeling of a von Hippel-Lindau transfectant cell line model. *Proteomics* 2009, *9*. 2118–2130.

[11] Schenk, S., Schoenhals, G. J., de Souza, G., Mann, M., A high confidence manually validated human blood plasma protein reference set. *BMC Med. Genomics* 2008, *15*, 1–41.

[12] Park, S. Y., Roh, S. J., Kim, Y. N., Kim, S. Z. et al., Expression of MUC1, MUC2, MUC5AC and MUC6 in cholangiocarcinoma: prognostic impact. *Oncol. Rep.* 2009, *22*, 649–657.

[13] Srisomsap, C., Sawangareetrakul, P., Subhasitanont, P., Chokchaichamnankit, D. et al., Proteomic studies of cholangiocarcinoma and hepatocellular carcinoma cell secretomes. *J. Biomed. Biotechnol.* 2010, *2010*, 437143.

[14] Korita, P. V., Wakai, T., Ajioka, Y., Inoue, M. et al., Aberrant expression of vimentin correlates with dedifferentiation and poor prognosis in patients with intrahepatic cholangiocarcinoma. *Anticancer Res.* 2010, *30*, 2279–2285.

[15] Dos Santos, A., Court, M., Thiers, V., Sar, S. et al., Identification of cellular targets in human intrahepatic cholangiocarcinoma using laser microdissection and accurate mass and time tag proteomics. *Mol. Cell. Proteomics* 2010, *9*, 1991–2004.

[16] Koopmann, J., Thuluvath, P. J., Zahurak, M. L., Kristiansen, T. Z. et al., Mac-2-binding protein is a diagnostic marker for biliary tract carcinoma. *Cancer* 2004, *101*, 1609–1615.