A MULTI-MODALITY APPROACH FOR ENHANCING THE DIAGNOSIS OF CHOLANGIOCARCINOMA

CHRISTOPHER ANTONY WADSWORTH

HEPATOLOGY & GASTROENTEROLOGY SECTION
DEPARTMENT OF MEDICINE
IMPERIAL COLLEGE LONDON

September 2012

For SJL

Abstract

Background

Cholangiocarcinoma (CC) is a malignancy of the bile ducts and mortality is high as patients present too late for curative surgery. In most cases of CC the aetiology is unknown, whilst diagnosis and staging are challenging. The hepatobiliary system excretes carcinogenic toxins and genetic mutations in biliary transporters lead to dysfunction and cholestasis, potentially contributing to cholangiocarcinogenesis. Polymorphisms in the NKG2D receptor have previously been associated with CC in primary sclerosing cholangitis (PSC). Such a role has not been investigated in sporadic CC. CC is difficult to diagnose, particularly in those with PSC. The transition from benign to malignant biliary disease is likely to be reflected in changes to the plasma proteome. However, current plasma biomarkers do not reliably distinguish benign from malignant biliary strictures. Elevation of neutrophil gelatinase-associated lipocalin (NGAL) has been demonstrated in the bile of patients with CC but has not been investigated as a plasma protein biomarker. Staging of CC is inaccurate, with only a minority of operated patients cured. Higher resolution MRI would improve diagnosis and staging. The work presented in this thesis represents a multimodality approach to enhance the diagnosis of CC:

Genetic studies

Genetic variation in major biliary transporter proteins, and the NKG2D receptor, were investigated. Single nucleotide polymorphisms (SNPs) in candidate genes were selected using HapMap. DNA from 173 CC patients and 265 healthy controls was genotyped. SNPs in *ABCB11*, *MDR3* and *ATP8B1* were nominally associated with altered susceptibility to CC*,* suggesting a potential role in cholangiocarcinogenesis. The previous association of NKG2D variation with CC in PSC was not replicated in sporadic CC, suggesting a possible difference in pathogenesis.

Protein studies

Plasma from subjects with CC, benign disease, and from healthy controls was studied. Two proteomic techniques, liquid chromatography-tandem mass spectrometry (LC-MS/MS) and surfaced enhanced laser desorption ionization time-of-flight MS (SELDI-TOF MS), were utilised. Differentially expressed proteins were identified where possible. LC-MS/MS fully identified six proteins that were differentially expressed in CC compared to gall stone disease patients. SELDI-TOF MS identified seven *m/z* peaks that showed significant utility in discriminating CC from PSC controls. An ELISA approach was used to study plasma NGAL levels in CC. Although differentially expressed between CC and healthy control groups, the utility of NGAL in discriminating CC from PSC was limited.

Imaging studies

An endoscope-mounted MR coil and intraductal MR detector coil were developed. Quantitative resolution and signal-to-noise-ratio (SNR) testing, and qualitative tissue discrimination appraisal, were undertaken. Sub-0.7mm resolution and excellent SNRs have been demonstrated. High-resolution has been demonstrated in imaged tissue. Imaging with the new devices compares favourably with endoscopic ultrasound imaging.

Acknowledgements

I thank my supervisors, Dr Shahid Khan, Dr Robert Edwards and Professor Simon Taylor-Robinson for taking me on as their student, for sharing their wisdom and for their patience in guiding a sensible strategy through the widely differing aspects of my research programme. I am grateful to Dr Peter Dixon and Professor Catherine Williamson, Institute of Reproductive and Developmental Biology, Imperial College London, for advice on the methodology of my genetic studies and to Prof John Whittaker, Hospital for Tropical Diseases, University College London, for his advice on statistical analyses of genotype data. I am indebted to Dr Verena Horneffer-van der Sluis for her enthusiasm, meticulous advice and assistance with laboratory proteomic techniques and I thank Ms Magda Gierula for her technical assistance with the LC-MS/MS work. I thank Professor Richard Syms and Professor Ian Young FRS, Faculty of Engineering, Imperial College London, for the inspiring insight into a parallel world of academic electronic engineering, and their advice on my MR detector coil studies, and to Dr Munir Ahmed for his assistance with technical aspects. I am grateful to the entire Imaging team at Imperial, including Dr Marc Rea, Ms Julie Fitzpatrick and Mr Warren Caspersz for affording access to the MRI scanner at St Mary's Hospital and for assisting with my MR studies there; and to Dr Steve Pereira, Dr Mike Chapman, Dr Harpreet Wasan, Mr Duncan Spalding, Dr Siobhan Ralphs, Dr Mark McPhail, Sr Mary Crossey and, especially, to Dr Kirsten Boberg (of NoPSC, Oslo, Norway) for their assistance in collecting samples for my studies. I thank Professor Mark Thursz and Dr Janice Main from the Hepatology and Gastroenterology Section, Division of Diabetes Endocrinology and Metabolism, Department of Medicine, Imperial College London for their general advice on the strategy of my research period.

I gratefully acknowledge the charitable bodies that have part funded my work; the Imperial Healthcare Charity, the British Liver Trust, AMMF - The Cholangiocarcinoma Charity, the ESPRC and The Wellcome Trust.

Statement of originality

I have personally developed the study protocols, obtained ethics approval, collected samples, processed samples and undertaken the genetic and proteomic analyses described in this document.

KASPar genotyping is performed on an automated system and was outsourced to KBioscience. I prepared and plated all samples. KBioscience supplied raw genotyping data and I have undertaken all analyses myself.

The MRI microcoils are novel, and have been designed, developed and constructed by Professor Syms and his Electronic Engineering team. I have been intimately involved in their design specification. I have been directly involved in the conception, design, execution and reporting of all the imaging experiments described in this document.

The work described in this report is my own. Where collaborations have been entered into, these are clearly identified.

Abbreviations

α1AT	alpha-1 antitrypsin
ABCB11	ATP-binding cassette, sub-family B member 11
ABCB4	ATP-binding cassette, sub-family B member 4
ABCC2	ATP-binding cassette sub-family C member 2
AFP	Alpha-fetoprotein
AKI	Acute kidney injury
AMMF	Alan Morement Memorial Fund
ATP	Adenosine triphosphate
ATP8B1	Probable phospholipid-transporting ATPase IC
BRIC	Benign recurrent intrahepatic cholestasis
BSEP	Bile salt exporter pump
CA 19-9	Carbohydrate antigen 19-9
CATT	Cochran Armitage trend testing
CC	Cholangiocarcinoma
CEA	Carcinoembryonic antigen
CEUS	Contrast enhanced ultrasound
CI	Confidence interval
C-P	Child-Pugh
CRF	Case Record Form
CRP	C-reactive protein
CS	*Clonorchis sinensis*
CT	Computed tomography
DIA	Digital image analysis
EDTA	Ethylenediaminetetraacetic acid
EHCC	Extrahepatic cholangiocarcinoma
ERCP	Endoscopic retrograde cholangiopancreatography
ELISA	Enzyme-linked immunosorbent assay
EPSRC	Engineering and Physical Sciences Research Council
EUS	Endoscopic ultrasound
FAM	A fluorescent dye marker used to label molecules
FDR	False discovery rate
FIC1	Familial intrahepatic cholestasis 1 (protein)
FISH	Fluorescent in situ hybridization
FNA	Fine needle aspiration (of cells)
FOV	Field of view
FR-FSE	Fast Relaxation Fast Spin Echo
FSE	Fast spin echo
FXR	Farnesoid X receptor
GeLC-MS/MS	Gel fractionation, liquid chromatography and MS/MS
GSD	Gallstone disease
HC	Healthy control (or healthy control cohort)
HCC	Hepatocellular carcinoma
HLA	Human leukocyte antigen

HNPCC	Hereditary non-polyposis colorectal cancer/Lynch syndrome
HPD	Hematoporphyrin derivative
HR	Hazard ratio
HWE	Hardy Weinberg equilibrium
ICL	Imperial College London
ICH	Imperial College Healthcare NHS Trust
ICP	Intrahepatic cholestasis of pregnancy
IDUS	Intra-ductal ultrasound
IHCC	Intrahepatic cholangiocarcinoma
IQ	Interquartile
KASPar	KBioscience allele-specific PCR SNP genotyping system
LC-MS/MS	Liquid chromatography and tandem mass spectrometry
LD	Linkage disequilibrium
LVS	Low-volume spectrophotometry
MAF	Mean allele frequency
MDCT	Multi-detector computed tomography
MDR3	Multidrug resistance protein 3
MHRA	Medicines and Healthcare Products Regulatory Agency
MICA	MHC class I polypeptide-related sequence A
MIP-1β	Macrophage inflammatory protein-1β
Mir	microRNA
MLR	Multiple linear regression
MMP	Matrix metalloproteinase (enzymes)
MRC	Medical Research Council
MRCP	Magnetic resonance cholangiopancreatography
MRI	Magnetic resonance imaging
MRP2	Multidrug resistance-associated protein 2
MS	Mass spectrometry
M-W U	Mann-Whitney U test
NG	Nasogastric tube
NGAL	Neutrophil gelatinase-associated lipocalin (also known as Lipcalin-2)
NK	Natural killer (cell)
NKG2D	Natural killer cell receptor G2D
NoPSC	Norwegian PSC centre, Rikshospitalet, Oslo
NPV	Negative predictive value
NR1H4	Nuclear receptor subfamily 1, group H, member 4 (also known as FXR)
OLT	Orthotopic liver transplantation
OR	Odds ratio
OV	*Opisthorchis viverrini*
PCA	Principal components analysis
PCR	polymerase chain reaction
PDT	Photodynamic therapy
PET/CT	Positron emission tomography and computerised tomography
PFIC1	Progressive familial intrahepatic cholestasis type 1
PFIC2	Progressive familial intrahepatic cholestasis type 2

PPV	Positive predictive value
PSC	Primary sclerosing cholangitis
PTC	Percutaneous transhepatic cholangiogram
PTD	Percutaneous transhepatic drain
PC	Phosphatidylcholine
R0	A resection with margins negative for cancer
REC	Research Ethics Committee
RF	Radiofrequency (signal)
ROC-AUC	Receiver operator characteristic – area under the curve
ROI	Region of interest
ROX	A fluorescent dye marker used to label molecules
SARS	Severe adult respiratory syndrome
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SELDI-TOF	Surface-enhanced laser desorption/ionization time of flight
SNP	Single nucleotide polymorphism
SNR	Signal-to-noise ratio
SOP	Standard operating procedure
SPA	Sinapinic acid
SST	Serum separation tube
TE	Echo time
TNM	Tumour-node-metastases (criteria)
TR	Time to repetition
UCL	University College London
UCH	University College London Hospitals NHS Trust
US	Ultrasound
VIC	A fluorescent dye marker used to label molecules
WHO	World Health Organization
WTCCC	Wellcome Trust Case-Control Consortium
^{90}Y	Yttrium-90

Table of contents

Index of Tables

Index of Figures

1. Introduction

Cholangiocarcinoma (CC) was first described in 1840. Dr Max Durand-Fardel, a French physician, published a monograph on four patients with jaundice who rapidly succumbed to their disease. In three cases, *post mortem* examination revealed tumours of the pancreas obstructing the bile duct. In the final case he reported *'a soft, whitish, cancerous material, completely filling the bile duct that did not extend beyond the wall of the duct'*. Durand-Fardel concluded that this represented a primary cancer of the extrahepatic bile duct.[1] CC is a cancer that arises in the epithelial cells of the bile ducts and has features of cholangiocyte differentiation.[2]

1.1 Epidemiology

Worldwide, CC is the second commonest primary liver tumour, after hepatocellular carcinoma (HCC). CC is the commonest cause of death from primary liver cancer in the UK, with over 1000 deaths per year.[3] It occurs in both sexes with a slightly higher incidence in men.[4] In Western European countries, such as the United Kingdom, the median age at diagnosis of CC is 65 years and it rarely presents before the 5th decade, except in patients with primary sclerosing cholangitis (PSC) or uncommon genetic diseases of the biliary transporter proteins.[5] CC incidence varies hugely between regions and countries of the world. Incidence in Australia is relatively low, at 0.2 cases/100,000 men. The population of northern Thailand, and of other areas along the course of the Mekong River from the Yunnan Province of China to the Lao Peoples' Democratic Republic, has a very high rate - 96 cases/100,000 men (100-fold that of the UK). This very high rate has been clearly associated with endemic liver fluke infestation, with an earlier age of presentation than in Europe.[5] Although geographical variation in CC incidence is largely attributable to the prevalence of environmental risk factors, it is likely that ethnic or genetic factors may also contribute. Studies of the US population have shown wide variation between different ethnic groups with rates of 1.22/100,000 in Hispanic Americans and 0.3/100,000 in African Americans.[6]

Since 1968, there has been a 15-fold increase in mortality from CC in the UK, with a similar global trend across industrialised countries.[7-9] In most epidemiological studies, this is ascribed to increasing incidence of intrahepatic CC (CC developing proximal to the liver hilum), with extrahepatic CC rates remaining stable. A proportion of this rise in IHCC has been attributed to misclassification of hilar CC as IHCC.[10] However, it is likely that this accounts for only part of the increase in IHCC. Increasingly sophisticated diagnostic tools, and a global increase in the prevalence of IHCC risk factors, such as cirrhosis of any cause, may have also contributed to this increase.[11]

The mortality rate of CC is equal to its incidence. Mortality from CC is so high as most patients present too late for complete surgical resection, the only cure, and the cancer is generally resistant to chemotherapy and radiotherapy. The prognosis of CC is short, with median survival of 3-6 months in patients receiving palliative care.[12] Overall, median survival in developed countries is less than two years.[12, 13]

1.2 Risk factors

Most cases of CC are sporadic, with no known predisposing risk factor. The prevalence of risk factors, and hence their influence on CC incidence rates, varies with geography and ethnicity.

1.2.1 Inflammatory risk factors

Known risk factors for CC in Western populations include PSC, chronic secondary cholangitis and choledocholithiasis.[14-16] PSC is the commonest known cause of CC in the West but still only accounts for around 10% of CC cases.[17] Worldwide, chronic infestation of the biliary tree with the parasitic liver fluke *Opisthorchis viverrini* (OV) is a major risk factor for CC. In Northern Thailand a 5 to 15-fold increased risk of CC is described in OV infested individuals.[18, 19] Hepatolithiasis, in which calcium bilirubinate calculi form within the intrahepatic biliary tree, has been shown to increase risk of CC substantially in Western and Asian populations.[20, 21] It is rare in the West but has a high prevalence in some Southeast Asian countries such as Taiwan (20%).[22]

Hepatolithiasis has proven to be a major risk factor for IHCC in Taiwanese, Chinese and Korean studies.[20, 23, 24] Anatomical anomalies in the biliary tree caused by Caroli's syndrome (discussed below) and unresected choledochal cyst carry a lifetime risk CC of around 15%.[25, 26] All of these conditions cause structural bile duct abnormalities including dilatation and irregularity. These are thought to lead to biliary stasis, reflux of pancreatic juice into the bile duct, deconjugation of biliary toxins and predisposition to chronic secondary bacterial cholangitis. Each of these mechanisms, or a combination of them, may result in biliary epithelial injury and inflammation.[26-28]

The viral hepatitides have been investigated as potential risk factors for CC. Hepatitis C infection has been examined in multiple case-control studies and found to confer a relative risk of CC between 2.55 and 9.7, well above the background population risk.[14, 15, 29-34] A pathologic study of 1058 explanted livers sought evidence of dysplastic, precursor intrahepatic biliary lesions in patients with chronic hepatitis C (n=511), alcohol (n=112) or both (n=85) as well as chronic liver disease of other aetiology (n=216) and acute liver failure (n=134). A total of 1058 explanted livers were examined and 19 (1.8%) were found to have dysplastic biliary lesions. All the dysplastic changes were found in patients with hepatitis C (10/19), alcohol abuse (5/19) or both (4/19). In all cases, the dysplasia was multifocal.[35] Hepatitis B has also been investigated and, in the majority of published case-control studies, associated with an increased risk of CC.[15, 23, 24, 33, 34] HIV positivity was associated with a modest increase in CC risk in one study, but the strength of this finding is limited.[5] Non-alcoholic fatty liver, as well as diabetes and obesity without demonstrable liver disease, have recently been linked with increased risk of CC.[22] Cirrhosis of any cause increases the risk of CC, with a number of case-control studies producing relative risk estimates of between 5 and 27.[14, 15, 24, 32, 36] Causative mechanisms have not been elucidated for these risk factors, but chronic inflammation of the liver is a feature common to them all.

1.2.2 Toxic risk factors

Pro-mutagenic DNA adducts have been identified in CC-adjacent human tissue, indicating prior exposure to DNA-damaging exogenous toxins.[37] A variety of toxins and occupational risks have been implicated in CC, with varying strength of evidence. Thorotrast, an α particle-emitting radiological contrast containing thorium oxide, was used and subsequently proscribed in the middle of the last century. Thorium oxide has a long half-life and is concentrated in the reticuloendothelial system. It has been incontrovertibly linked with CC and a number of other malignancies, with a latency period of 20-35 years.[38, 39] Thorotrast exposure increases the risk of CC to 300 times that of the general population.[5] Epidemiological studies have implicated tobacco smoking and alcohol use, although conflicting evidence also exists.[14, 16, 33, 40, 41] Occupational exposure to the manufacture of rubber, petrochemicals and varnishes has been identified as an independent risk factor for CC.[42] These associations have been ascribed to exposure to chemical by-products, including dioxins and nitrosamines.[42] A proposed common pathway of these environmental risks is exposure of the biliary epithelium to oncogenic substances, which are concentrated in the bile, causing a sequence of genetic mutations and thereby initiating and promoting neoplastic change.[37]

1.2.3 Genetic risk factors for CC

1.2.3.1 Genetic diseases as risk factors for CC

Lynch syndrome (HNPCC) and biliary papillomatosis, which are both genetic conditions, increase risk of CC.[43] Biliary papillomatosis carries a huge (85%) lifetime risk of progression to CC.[44] These genetic diseases appear to offer an epithelial substrate that is highly vulnerable to subsequent oncogenic factors. Caroli's disease and syndrome represent a spectrum of bile duct ectasia, dilatation and congenital hepatic fibrosis, and increase the risk of CC (discussed in section 1.2.1, above). Caroli's disease and syndrome appear to be inherited as autosomal recessive traits. The genetics of these conditions is incompletely understood, but there is a close

association with autosomal recessive polycystic kidney disease (ARPKD). The gene responsible lies on chromosome 6 and has been named *PKHD1* (polycystic kidney and hepatic disease 1). The protein encoded by *PKHD1* is expressed predominantly in the kidney, less so in the liver and in low abundance in the pancreas and lungs. The protein is found on the primary cilia and centrosome complex of renal tubule cells and cholangiocytes and appears to play a role in cell adhesion and proliferation.[45-47] Genetic defects leading to congenital abnormalities in bile salt transporter proteins such as bile salt exporter pump, (BSEP/ABCB11), familial intrahepatic cholestasis protein 1 (FIC1/ATP8B1) and multidrug resistant protein 3 (MDR3/ABCB4) cause cholestatic disease in infants. These defects result in unstable bile content and deconjugation of xenobiotics, previously conjugated in the liver.[9, 48, 49] Very early onset hepatocellular carcinoma and CC have been reported in children with progressive familial cholestasis type II (PFIC II), caused by BSEP/ABCB11 deficiency.[50-52]

1.2.3.2 Specific genetic polymorphisms as risk factors for CC

1.2.3.2.1 Specific polymorphisms and CC risk factors

1.2.3.2.1.1 Genetic risk factors for gallstone disease (GSD)

Cholelithiasis is a very common, but weak, risk factor for CC. In a large Scandinavian twin study, genetic factors were calculated to account for 25% of the gallstone risk among the twins.[53] Polymorphisms in *UGT1A1* (implicated in Gilbert's syndrome) may predispose to pigment gallstone formation. *ABCG8, encoding part of cholesterol transporter,* was unambiguously associated with an increased gallstone risk.[54, 55] Association has also been made between variants of *NR1H4* (the gene that encodes the nuclear bile salt receptor FXR) and *SLC10A2* (encoding the apical sodium-dependant bile acid transporter) and cholelithiasis.[56, 57] Recent studies have demonstrated a number of polymorphisms in other biliary transporter genes, including *ABCB4* and *ABCB11*, which increase the risk of cholelithiasis.[58-61]

1.2.3.2.1.2 Genetic risk factors for PSC

Genetic factors in primary sclerosing cholangitis (PSC), the major risk factor for CC in Western populations, have been investigated extensively. Genetic variation in the human leukocyte antigen (HLA) complex on chromosome 6p21 has been repeatedly associated with PSC.[62, 63] Studies of specific genes have been small and negative to date, including a study of *ABCB11* and *ABCB4* (n= 37).[64] A recent national genome wide association study (GWAS) in Norway demonstrated associations between polymorphisms in *HLA-B*, macrophage-stimulating 1 (*MST1*), G protein-coupled bile acid receptor 1 (*GPBAR1*) and PSC.[65] A large-scale international GWAS in PSC is underway.

1.2.3.2.2 Specific genetic polymorphisms and risk of CC

A number of studies of candidate genes in CC, each examining between 30 and 216 cases, have reported a positive association with the disease. All have yet to be replicated in validation cohorts. These studies, and their principal findings are summarised in Table 1.1.

1.2.3.2.2.1 Xenobiotic metabolism, folate metabolism and DNA repair

These studies have been performed on populations exclusively from the Far East. Polymorphisms affecting enzymes of xenobiotic metabolism that have been associated with CC include glutathione S-transferase (*GST01)*, N-acetyltransferase 2 (*NAT2)* and cytochrome P450 1A2 (*CYP1A2*). Polymorphisms in methylenetetrahydrofolate reductase (*MTHFR*), involved in folate metabolism, and XRCC1, involved in DNA repair mechanisms, have also been implicated in single studies.[66-69] Several of these studies included gallbladder cancer, and even ampullary tumours, with CC cases.

1.2.3.2.2.2 Alpha-1 antitrypsin (A1AT)

In a German cohort, alpha-1-antitrysin (α1AT) Z heterozygosity was found to be over-represented in patients with CC suggesting a potential contribution of aberrant α1AT function or excretion in biliary carcinogenesis.[70]

1.2.3.2.2.3 MRP2

A recent German study, in a cohort of 60 patients, found a polymorphism in multidrug resistance-associated protein 2 gene (MRP2/*ABCC2*) to be associated with CC.[71] MRP2 is an ATP-binding cassette biliary transporter expressed on the apical membrane of hepatocytes and cholangiocytes. The c.3972 C>T SNP genotype was significantly more frequent in CC (39.2%) vs. controls (26.0%, p = 0.022) OR of 1.83 (95% CI = 1.09 - 3.08). MRP2 is a reasonable candidate susceptibility protein as it plays a critical role in the excretion of xenobiotics and polymorphisms in *ABCC2* have been associated with HCC. This is the only published study to date of genetic variation in biliary transporter proteins as risk factors for CC.

1.2.3.2.2.4 Natural killer cell protein G2D (NKG2D)

Genes modifying Natural Killer cell activation (natural killer cell receptor G2D, *NKG2D*) have been implicated in a number of neoplasms. This gene has been investigated in a cohort of Scandinavian PSC patients. 2 of the seven SNPs studied were associated with altered susceptibility to CC in PSC.[72] Polymorphisms in *NKG2D* are yet to be investigated in sporadic CC.

1.3 Pathogenesis

CC results from malignant transformation of cholangiocytes, although some CCs appear to arise from hepatic progenitor cells.[73, 74] In comparison to other gastrointestinal tract malignancies, the pathway of cholangiocyte transformation is poorly understood. There is evidence of a stepwise process, with non-invasive pre-malignant precursor states now described. Biliary intraepithelial neoplasia (BilIN) lesions have been defined, and intraductal papillary mucinous neoplasm (IPMN) appears to be a less common precursor, similar to the same lesions found in the pancreas.[75, 76] Recent work on the molecular pathways involved, including gene expression studies, has been fruitful. There is now a reasonable consensus on some of the molecular events that contribute to cholangiocarcinogenesis. Common features of many disease risk factors for CC are chronic cholestasis and inflammation of the biliary

tree. It is postulated that such inflammation increases cell turnover, overwhelms local tissue repair pathways, and is intrinsically pro-oncogenic. Chronic inflammation and cholestasis have been shown to trigger release of cytokines and growth factors including hepatocyte growth factor (HGF), epidermal growth factor (EGF), platelet derived growth factor (PDGFs), TGF-B, TNF-a and IL-6. Abnormal bile acid signaling transactivates EGFR and induces the COX-2 pathway. The cytokine cascade induces inducible nitric oxide synthase. iNOS up-regulation leads to increased levels of reactive nitrogen oxide species that cause DNA breakage, gene and protein mutation, and suppression of DNA repair enzymes.[77, 78] Cholangiocyte hyperplasia, dysplasia and aneuploidy are the result. Ongoing damage to the biliary epithelium enhances cholangiocyte exposure to endogenous mutagens, such as hydrophobic bile acids and oxygen free radical species, as well as exogenous carcinogens that are concentrated and excreted in the bile.[77] The acquisition of subsequent mutations of genetic and epigenetic pathways that affect cell cycling, apoptosis, proliferation, angiogenesis and invasion contribute to the evolution of an overtly malignant lesion. This cascade is summarised in Figure 1.1, below (adapted from Sirica *et al* 2008).[79] There follows a summary of the factors currently implicated in cholangiocarcinogenesis. Whether these factors and pathways are truly pathogenic, or are bystander events is frequently unclear.

1.3.1 Inflammatory mediators, growth factors and cholangiocarcinogenesis

1.3.1.1 Interleukin-6 (IL-6) and its receptor, IL6R

IL-6 is a cytokine that is released by injured tissue and has pro-inflammatory and anti-inflammatory effects. IL-6 is usually undetectable in the healthy liver.[80] It appears to play a number of critical roles in cholangiocarcinogenesis. It is secreted principally by macrophages and T-cells but also by periportal stromal cells, including stellate cells, and cholangiocytes themselves. It acts in an autocrine loop, promoting proliferation of cholangiocytes and well as inducing nitric oxide synthase.[81, 82] IL-6 activates the signal

transducer and activator of transcription-3 (STAT-3) receptor. The binding of IL-6 to the gp130 receptor triggers STAT3 phosphorylation by JAK2. Phosphorylation of this receptor kinase activates nuclear transcription of a variety of mitogenic and anti-apoptotic factors.[83, 84] IL-6 also acts to down-regulate bcl-2–associated X protein (Bax), thereby inhibiting cholangiocyte apoptosis.[80] IL-6 is markedly elevated in animal models of bile duct injury.[80] It has been shown to stimulate cholangiocyte DNA synthesis, is elevated in the serum of patients with cholangiocarcinoma as well as in the bile of patients with cholangitis.[85, 86] IL-6 also triggers increased expression of HGF and c-met[87, 88], induces COX-2 expression[89] and increases telomere length[90], all of which contribute to choangiocarcinogenesis and are considered in more detail below.

1.3.1.2 Hepatocyte growth factor (HGF) and its receptor, c-met

The HGF ligand and its tyrosine kinase receptor, c-met, are not expressed in normal biliary tissue. Activation of c-met by HGF leads to transduction of numerous pathways including the RAS (scattering, proliferation and branching morphogenesis), PI3K (cell motility, adhesion, cytoskeletal reorganisation and cell survival) and STAT3 (discussed above) pathways. Activation of HGF/c-met is normally limited to stem and progenitor cells, or healing tissues. It is upregulated in bile duct injury models.[80, 91, 92] The proto-oncogene, *c-met*, encodes the c-met receptor. Increased c-met expression has been identified in numerous studies of CC tissue, as well as other solid cancers.[93-96] A high level of c-met expression has been associated with particularly poor prognosis and is known to co-activate EGFR, and *vice versa*.[97] Co-expression of c-met and EGFR has been found in a variety of cancer cells lines.[98]

1.3.1.3 Epidermal growth factor (EGF) and its receptor, EGFR

EGF acts by binding with its receptor on the cell surface and stimulating the intrinsic protein-tyrosine kinase activity of the receptor. This initiates a signal transduction cascade, increases the expression of certain genes, including the gene for EGFR. EGF/EGFR activation results in cellular proliferation, differentiation, and survival.[99] EGF levels are elevated in inflamed biliary epithelium and EGFR/ErbB1 induction has

been demonstrated in CC tissue.[99, 100] Activation of EGFR increases COX-2 expression and, in a positive feedback loop, COX-2 ezymatic product prostoglandin-E2 (PGE_2), transactivates EGFR.[101] EGF/EGFR interaction also activates the STAT-3 pathway, already discussed in 1.3.5, above.

1.3.1.4 Cyclooxygenase-2 (COX-2) and its product, prostoglandin-E2 (PGE_2)

COX-2, an enzyme that synthesises prostaglandin species, is minimally expressed in normal biliary tissue but is upregulated in inflammatory states. Cytokines such as TNFα, IFNγ, IL-6 and IL-1β all induce COX-2 expression.[102] As discussed in 1.3.2, above, EGFR and COX-2 cross-upregulate each other, in a positive feedback loop. Hydrophobic bile acids, known to promote cholangiocarcinogenesis in mouse models, are now known to induce COX-2 expression via EGFR.[103-106] Enhanced COX-2 expression has been demonstrated in dysplastic cholangiocytes, cholangiocytes from PSC affected subjects and in cholangiocarcinoma cells.[107-109]

COX-2/PGE2 attenuates the mitotic inhibitory effect of TGF-β[110] and also enhances the pro-mitotic effects of IL-6 and HGF.[111] In vitro, PGE_2 application to CC cell lines promotes cell division and reduces apoptosis.[112, 113] PGE_2 has been shown to promote tumour growth and upregulates myeloid leukaemia cell differentiation protein **(**Mcl-1), an anti-apoptotic member of the Bcl-2 family.[102]

1.3.1.5 Transforming growth factor-β1 (TGF-β1)

TGF-β1 is involved in tumour progression by promoting angiogenesis and suppressing the immune system. TGF-β1 also has a growth-inhibitory effect on epithelial cells including carcinoma cells via the SMAD and DAXX pathways.[114] This anti-mitotic effect is abrogated by COX-2/PGE2 activity in cell line models of cholangiocarcinogenesis.[110] TGF-β1 also exerts immunomodulatory effects on T-lymphocytes and natural killer (NK) cells, reducing their activity in tumour surveillance.[115-117]

1.3.1.6 Tissue necrosis factor-α (TNF-α) and its receptor, NF-kappaB

High levels of TNFα can be identified in the bile of patients with cholangitis.[85] TNFα activates NF-kappaB, leading to upregulation of the Snail pathway, consequent repression of E-cadherin in cholangiocarcinoma, promoting migration and invasion.[118]

1.3.1.7 Inducible nitrogen oxide synthase (iNOS)

iNOS catalyses nitric oxide (NO) from L-arginine. It is induced by a wide range of cytokines and plays a central role in the mediation of inflammation. It is thought to contribute to cholangiocarcinogenesis in a variety of ways. NO directly damages DNA via oxidation and inhibits DNA repair mechanisms. It can nitrosylate thiol and tyrosine residues in proteins, disrupting or modifying their function, and plays a critical role in activation of the Notch pathway, reducing apoptosis in cholangiocytes.[119] NO also upregulates COX-2 expression, as discussed in 1.3.1.4, above.[120]

1.3.2 Somatic genetic changes in cholangiocarcinogenesis

The inflammatory milieu in the bile duct is thought to contribute to initiation of genetic mutations that drive cellular proliferation, decreased senescence, decreased apoptosis and invasion.[121, 122] The innately toxic nature of bile contributes to this vicious cycle of inflammation and also contributes directly to further DNA damage.[37] Molecular studies of dysplastic biliary and CC tissues have demonstrated cumulative genetic mutations in the epithelium. Dysregulation of oncogenes (K-Ras, c-myc, c-neu, c-erbB-2, c-met, CCND1) and tumour suppressor genes (p53, DPC4/Smad4, CDKN2A, APC, $p16^{INK4a}$)[123-125] has been demonstrated. However, the frequency of such mutations varies widely with ethnicity, tumour stage, tumour location and pre-existing biliary pathology.[126] The sequence of genetic changes and their role in initiation and progression of CC in not clear.[127, 128]

1.3.3 Epigenetic changes in cholangiocarcinogenesis

Recent studies have demonstrated a range of epigenetic factors in carcinogenesis. Promoter hypermethylation, histone deacetylation and altered microRNA (mir)

signalling have all been implicated in CC. Methylation of DNA within gene promoter region CpG islands can downregulate or silence associated gene transcription. In CC, this mechanism of down regulation of tumour suppressor genes has been repeatedly demonstrated, including APC, $p16^{INK4a}$ and p73.[129] Histone acetylation also mediates suppression of associated DNA transcription and thereby tumour suppressor genes.[130] Mir modulates the transcription of genes. Many genes that encode mir are located within malignancy-associated regions of the genome. Various studies have associated abnormal mir transcription, and thereby modified transcription of tumour promoting/suppressing genes, with cancer. Over expression of mir-141, mir-200b and mir-21, and downregulation of mir-29b mir-370, have been specifically associated with cholangiocarcinogenesis.[129]

1.4 Histopathology

1.4.1 Histological classification

More than 90% of CCs are adenocarcinomas, the reminder are squamous cell carcinomas. CC can be classified as well, moderately or poorly differentiated. All subtypes have a propensity for local invasion, indolent growth, peri-neural spread and mucin production.[131] Three distinct histological subtypes of CC have been described - sclerosing, nodular and papillary. A substantial majority are of the sclerosing type, characterised by an intense desmoplastic reaction. Invasion along and through the bile duct wall is an early feature.[131, 132] Presentation is generally at a late stage and prognosis is particularly poor. Sclerosing CC has a radiological appearance similar to that of benign strictures of the bile duct, complicating diagnosis and management further.[133]

1.4.2 Precursor lesions

Two distinct pre-malignant biliary intraepithelial lesions have been defined, that are thought to precede invasive intrahepatic and hilar CC.[134] Biliary intraepithelial neoplasia (BilIN) is a flat or micro-papillary dysplastic biliary epithelial lesion.[135] It has

been demonstrated in hepatic explant series of patients suffering from PSC, hepatolithiasis and choledochal cysts.[136] Intraductal papillary neoplasm of the bile duct (IPNB) is characterized by prominent papillary growth of atypical biliary epithelium with distinct fibrovascular cores and mucin over-production.[137, 138] BilIN and IPNB are regarded as biliary counterparts of pancreatic intraepithelial neoplasia (PanIN) and pancreatic IPMN. [75, 139]

1.4.3 Anatomical classification

The anatomical site of any CC is of critical importance as this determines resectability, the nature of any surgery and the likelihood of its success. About 65% are hilar or perihilar (Klatskin tumours), 25% in the distal bile duct (defined as beyond the point at which the bile duct passes posterior to the duodenum) and 10% are within the peripheral liver.[2] Perihilar and distal lesions are classified as extrahepatic while lesions proximal to the hilum are classified as intrahepatic. In planning surgery, further detailed anatomical classification of perihilar CC is made into Bismuth-Corlette categories (Figure 1.2). CC, cancer of the gallbladder and cancer of the ampulla of Vater are now widely considered to be separate entities - in both clinical research studies and clinical practice.[140]

1.4.4 Pathological diagnosis

1.4.4.1 Diagnostic histology

The fibrotic, paucicellular nature of CC makes successful sampling for histology or cytology very difficult. In small or early lesions, percutaneous biopsy often reveals normal or fibrotic tissue only. The intimate relation of most CCs to large bile ducts, and their adjacent blood vessels, means that percutaneous or laparoscopic biopsy carry significant risk of damaging critical structures. Very large, or peripheral lesions may be highly amenable to percutaneous biopsy. Seeding of malignant cells from a CC through a biopsy track is a theoretical possibility and has been demonstrated in studies of HCC.[141] It is therefore avoided in cases where curative resection is thought possible, as is EUS guided FNA.

1.4.4.2 Diagnostic cytopathology

Collection of cytology samples from biliary strictures is possible, using a brush or biopsy forceps at ERCP. The desmoplastic nature of most CCs means that these techniques offer a low yield of material for analysis. The cellular inflammatory response in conditions such as PSC complicates analysis further as atypical cells are seen in both conditions. In clinical studies, the sensitivity of these techniques (or combinations of them) ranges from 40 to 64% and specificity from 90 to 100%.[142-144] These studies included patients with a mixture of diseases, including metastatic and pancreatic cancers compressing the bile duct. Subgroup analyses suggest that cytological sampling is more likely to yield the diagnosis in lesions that originate in the bile duct, i.e. CC. Differential sensitivities for CC and pancreatic cancer of 80% and 35% respectively have been shown.[142] Cytological diagnosis is fraught with variability in technique, for both endoscopist and cytopathologist. In the studies considered here, gold-standard technique and expert operators were used.[145-147] Real-world performance is likely to be substantially worse.

1.4.4.3 Cytopathological markers

Immunohistochemistry has been unhelpful in the differentiation of CC from other malignancies, as no specific markers have been identified. However, the use of DNA proliferation measures shows promise. Cytokeratin, CEA and mucin staining of biopsy tissue are suggestive of the diagnosis, but are non-specific. Testing of K-ras and p53 mutation in cells collected for cytology has been investigated in clinical studies and shown not to improve sensitivity or specificity.[148-150]

1.4.4.3.1 Digital image analysis (DIA) and fluorescent in situ hybridisation (FISH)

DIA measures the quantity of nuclear DNA in cells whilst FISH labels DNA with a fluorescent marker, permitting visualisation of chromosomal abnormalities. The use of DIA or FISH can improve the sensitivity and specificity of biliary stricture cytology in patients with and without PSC. In one study, DIA demonstrated sensitivity and specificity of 39% and 77% compared with 18% and 98% with standard cytology.[151] In

patients with PSC, DIA has sensitivity of 43% and specificity of 87%. FISH has sensitivity of 47% and specificity of 100% in such patients.[152, 153]

1.4.4.4 Cytopathology in PSC

False positivity of cytopathological analysis is more common in patients with PSC and, in combination with the confounding radiological findings in PSC, can lead to incorrect diagnosis and therapy.[154]

1.4.4.5 Cytopathology in Immunoglobulin G4 (IgG4) cholangiopathy

The absence of malignant cells and the presence of a lymphoplasmacytic infiltrate rich in IgG4-positive cells is diagnostic and can be sought in biliary brushings, ampullary biopsies or FNA samples.[155]

1.5 Clinical features

Symptoms of icterus and pruritus are common, but occur only once bile outflow is almost completely obstructed. A minority of patients report a dull right upper abdominal pain. Non-specific features include lethargy, weight loss and low-grade pyrexia. The long latency before presentation is thought to contribute to the advanced stage at diagnosis and, therefore, poor prognosis.[156] Physical signs of CC include jaundice, right upper abdominal mass or hepatomegaly and only occur when the lesion is extensive enough to cause palpable mass effect or substantial biliary obstruction.[157] Where the patient suffers from pre-existing benign biliary disease such as PSC, progressive jaundice, weight loss and malaise may accompany deterioration of that underlying disease or herald the development of a CC. It is also impossible to clinically distinguish the presenting symptoms and signs of CC from those of hepatocellular carcinoma, pancreatic cancer or metastatic cancer of the liver.[158] Clinical and imaging findings in Immunoglobulin G4 (IgG4) associated cholangiopathy may closely mimic those of CC. IgG4 cholangiopathy is part of a recently described, and increasingly diagnosed, multisystem inflammatory disorder. The disease is non-malignant and generally steroid responsive. Many patients with IgG4 cholangiopathy present with

biliary strictures and jaundice, making it an important differential diagnosis.[155] Differentiating between these conditions and CC is critical for accurate diagnosis, management and prognostication in patients presenting with obstructive jaundice.

1.6 Laboratory tests

1.6.1 Serum biochemistry

Standard liver function tests including bilirubin, alkaline phosphatase (ALP) and transaminases are usually elevated at the time of diagnosis. Peripheral IHCC often causes no elevation in bilirubin, as the volume of liver obstructed by the lesion does not affect the total excretory capacity of the liver significantly.[158] Elevated bilirubin and ALP are associated with all benign and malignant cholestatic diseases. Progressive deterioration of these indices in patients with PSC may indicate an evolving CC.[156]

1.6.2 Tumour markers

Carbohydrate antigen 19-9 (CA19-9) is the most widely used tumour marker in the diagnosis of CC.[140] The defining antibody reacts with mucin glycoproteins that are coated with sialylated blood-group epitopes such as sialyl Lewis^a.[159] Large studies have demonstrated sensitivity and specificity of up to 75% and 85% respectively.[160] However, specificity is greatly reduced in the presence of cholestasis or cholangitis. In studies of PSC cohorts, sensitivity and specificity have ranged hugely; 38-80% and 50-98% respectively.[159, 161] The selection criteria of cases studied and the CA19-9 cut-off value used account for this wide variation. Most importantly, the elevation of CA19-9 has been shown to occur late in the development of CC. In a longitudinal study of PSC patients undergoing screening for CC with CA19-9, only 2/14 patients who developed CC were candidates for curative treatment.[162] CA19-9 is also elevated in gastric cancer, pancreatic cancer, primary biliary cirrhosis and individuals who smoke. 7% of the population are Lewis-negative and are unable to express CA19-9 at all.[163]

Carcinoembryonic antigen (CEA) has been investigated extensively but is only positive in about 30% of patients with CC.[126] Some studies have suggested that CEA can

enhance sensitivity and specificity when used in conjunction with CA19-9 but this has not been validated in prospective clinical studies.[159]

1.6.3 Serum IgG4

Elevated serum IgG4 is a feature of IgG4 cholangiopathy. Many patients with IgG4 cholangiopathy present with biliary strictures and obstructive jaundice, making CC an important differential diagnosis. In a study of 126 CCA and 50 IgG4 cholangiopathy patients, 17 (13.5%) of the CC cohort had elevated IgG4 and 4 (3.2%) had a >2-fold increase. PSC was present in 31/126 CC patients, of whom 7 (22.6%) had elevated IgG4 and 2 (6.5%) had a >2-fold elevation. Of the 50 IgG4 cholangiopathy patients, 39 (78.0%) had elevated IgG4 and 25 (50.0%) had a >2-fold increase. In a validation cohort of 161 CC and 47 IgG4 cholangiopathy patients, the results were consistent with those of the test cohort. The authors concluded that, although very useful, IgG4 elevation alone does not exclude the diagnosis of CC.[164] In the absence of clear radiological or cytopathogical diagnosis, serum IgG4 quantification is likely to prove helpful in identifying the small subset of patients that may have IgG4 cholangiopathy rather than CC.

1.7 Imaging in the diagnosis and staging of CC

A variety of imaging techniques is used in the assessment of biliary strictures and hepatic mass lesions. Imaging is the mainstay of diagnosis and staging of CC and most patients undergo three or more imaging techniques before their lesion is considered fully characterised.[140] Recent advances include detection of perfusion characteristics of benign and malignant lesions, liver specific contrast agents and enhanced signal detector technology. Standard imaging techniques can reliably distinguish malignant from some common benign causes of biliary obstruction, such as choledocholithiasis. However, discrimination between benign and malignant strictures of the bile ducts is more difficult.[165] Cholangiography is the most accurate way to delineate the site and intraductal extent of any stricture. This can be obtained via

endoscopic retrograde cholangiopancreatography (ERCP), percutaneous transhepatic cholangiogram (PTC) or magnetic resonance cholangiopancreatography (MRCP) with equal accuracy.[140]

In the context of background PSC, diagnostic imaging of CC is particularly challenging. Approximately 70% of patients with PSC have cholangiographic evidence of both intra- and extra-hepatic ductal involvement – including stricturing and dilatation.[166, 167] New, malignant strictures may be superimposed on these benign strictures, making accurate detection difficult.[168] This is a major problem - in one study of 125 PSC patients who were followed up for ten years, 45% developed a benign dominant stricture at some point.[169] In a 6 year prospective study of 230 patients with PSC, 10% were diagnosed with CC. The use of CA19-9 alone was less sensitive than a combination of CA19-9 and cross sectional imaging. However, the addition of US or CT imaging did not improve specificity, positive predictive value (PPV) or negative predictive value (NPV). The addition of cholangiography (as MRCP or ERCP) did improve specificity and PPV.[170]

Even with the full range of imaging technologies discussed in the following sections, accurate diagnosis and staging of CC is often impossible.

1.7.1 Ultrasound (US)

1.7.1.1 Transabdominal B-mode US

Transabdominal ultrasound (US) is a quick, widely available and safe test. Transabdominal B-mode US is highly sensitive for the detection of bile duct dilatation and, in this context and skilled hands, delineating the site of any obstruction.[171-173] However, US is operator dependent and has low sensitivity and specificity for CC detection and is inaccurate in estimating local tumour extent. Sensitivity for detection of hepatic masses of any type ranges from 37-87%.[174, 175]

1.7.1.2 Contrast enhanced US (CEUS)

Developments in ultrasound technology, and associated ultrasound contrast agents, have led to major advances in characterisation and detection of liver lesions. These contrast agents are microbubbles, and contain an inert gas (typically a perfluorocarbon compound) with a stabilising outer phospholipid shell. The bubbles are 3-5μm in diameter and remain intact in the circulation for up to 20 minutes, but only have an optimum window of use of 5 minutes sonographically. When exposed to US energy, these bubbles generate specific harmonic signals that can be distinguished separately from the returning fundamental signals. The two signals are used to generate a separate side-by-side image or are displayed as an image overlay. These agents remain purely within the vascular space and thus allow the continuous assessment of the vascular perfusion of liver lesions from the arterial and portal phases right through to the late phase.[176] This technique has significantly improved the ability of ultrasound to accurately characterise liver lesions, including small IHCCs. On CEUS examination, IHCCs are typically hypovascular in the arterial and portal phases and remain as a defect in the delayed phase after contrast administration.[133] This characteristic is shared with metastatic liver cancers, but a discriminating factor is that CC masses tend to expand the intrahepatic ducts. This is not entirely reliable, as mass-forming types of CC appear identical to metastases. Finally, this hypovascularity allows CC masses to be distinguished from HCC, which enhance avidly in the arterial phase, or benign vascular lesions. Unfortunately, CEUS is less accurate at diagnosing the commonest, extra-hepatic infiltrating form of CC, especially in patients with PSC.[177]

1.7.2 Computed Tomography (CT)

New multi-detector CT (MDCT) has reduced the time required to image the whole liver by a factor of three, imaging the whole body in less than 1 minute. This permits repeated imaging through the different vascular phases and multiplanar reformats, thereby improving lesion characterisation. It can be difficult to distinguish duct dilatation due to a small CC or a calculus and therefore a pre contrast scan is extremely useful.

As with CEUS, CCs do not enhance in the arterial phase but this phase is important in delineating the anatomical vascular supply to the liver. In the portal and delayed phases, CCs typically shows a higher enhancement than the adjacent normal liver thought to be related to the abundant fibrous stroma of this type of tumour.[178]

Small hilar lesions can be difficult to detect on CT. Although CT is highly sensitive for biliary dilatation and identification of the site of biliary obstruction, the accuracy of characterising sub centimetre hypoattenuating lesions is poor.[179] Older studies have suggested that CT has low sensitivity and specificity for CC, is inaccurate at estimating local tumour extent and predicts resectability with accuracy of only 60%.[180] However, the advent of MDCT allowing improved resolution has been reported to substantially improve on these disappointing results.[181]

Although MDCT does not have a very high specificity for distinguishing CC from other liver malignancies and pathologies it remains the modality of choice for disease staging, particularly in the detection of distal metastases in the chest, abdomen and pelvis and also in assessing vascular invasion.[140, 179]

1.7.2.1 Positron emission tomography and CT (PET/CT)

Several studies have considered the use of *P*-[^{18}F]-2-deoxy-glucose PET/CT to resolve indeterminate cases as benign or malignant disease.[140, 179] In one study, PET/CT had sensitivity of 93% and specificity of 80% for IHCC. For the more common, extrahepatic, CC this fell to 55% and 33% respectively. Sensitivity and specificity for regional lymph node metastases were 12% and 96%.[182] Another study found sensitivity of 85% for nodular forms of CC but a dismal 18% for the commoner infiltrating type. Sensitivity for previously undetected extrahepatic metastases was 65%.[183] PET/CT was found to alter management significantly in the majority of patients in a study of a non-PSC CC cohort.[184] False positivity is common in patients with inflammatory biliary diseases.[185] Despite disappointing performance of PET/CT in the clinical management of CC, it maintains a limited role. Its positive predictive value in regional nodal disease of non-

PSC patients being staged prior to curative resection (or transplantation) may avoid inappropriate surgery for incurable lesions.[184]

1.7.3 Magnetic resonance imaging (MRI) and magnetic resonance cholangiopancreatography (MRCP)

MRI is the cross-sectional imaging modality of choice in assessing the biliary tree as it provides superior soft tissue resolution when compared with CT and US. MRCP provides a good overall anatomical assessment of the biliary system and, in addition, contrast-enhanced images of the liver parenchyma allow improved detection and characterisation of smaller lesions as well as periductal extension of tumour.

Typical Liver MRI normally includes T_1-weighted and varying heavily T_2 weighted sequences as well as n-phase/out-of phase spoiled gradient echo sequence. This is usually supplemented with dynamic fat-saturated T1-weighted gadolinium enhanced sequences to allow tumour enhancement pattern characterisation and estimation of extent. This, together with the MRCP sequence, is very helpful in assessing periductal infiltration as well as distinguishing benign from malignant strictures. CCs are typically low signal on T1 weighted sequences and moderately hyperintense on the T2 weighted sequences. They demonstrate an enhancement pattern similar to that seen on CT and CEUS but it is the periductal high signal oedematous change that is helpful in distinguishing ductal from other pathology. Therefore high resolution and varying T2 weighted sequences are employed.[133] MR is also particularly useful at detecting satellite intra-hepatic liver metastases. Studies have reported that extra-hepatic CCs show slow homogeneous enhancement into the late phase of scanning, while intra-hepatic lesions tend to enhance peripherally initially, with prolonged retention of contrast in the late phase.[186, 187] Contrast MRI assessment of vascular invasion is comparable to traditional angiography.[188]

The use of specific MR liver contrast agents such as ferumoxide has also been reported to improve the sensitivity of MRI studies for CC. Liver specific MRI agents consist of hepatocyte specific and reticuloendothelial agents which are taken up by

Kupffer cells. Older hepatocyte specific contrasts were manganese based agents such as Teslascan but more recently the development of bimodal gadolinium agents such Multihance/Gd-BOPTA (Bracco) and Gadoxetate, Primovist (Bayer-Schering), not only permit assessment of arterial and portal phase enhancement characteristics but also undergo biliary excretion and provide additional detail of the biliary tree.[189] Reticuloendothelial based or superparamagnetic iron oxide (SPIO) agents are very helpful in distinguishing benign from malignant lesions and thus improve the accuracy of MRI. These agents, for example Ferumoxide (Guerbet) or Resovist (Bayer-Schering), shorten the T1 weighted signal in normal liver which demonstrates Kupffer cell activity thereby appearing as low signal while malignant lesions such as CC return a higher normal signal when compared with the adjacent liver post SPIO administration.

Although contrast enhanced sequences are useful it is the MRCP that is key. MRCP provides high-resolution cholangiography, is non-invasive and rivals ERCP. It requires no contrast and utilises MR imaging data accrued with heavily T_2-weighted sequences. The high water content of bile acts as an endogenous contrast medium. On these sequences, CCs are seen as shouldered stenotic or sclerosing ductal lesions. Subtle duct wall thickening, nodular or papillary filling defects may be seen. MRCP is highly sensitive for biliary obstruction and its localisation but is slightly less accurate than ERCP for differentiating benign from malignant biliary strictures.[190] Although MRCP can detect proximal ductal extent more accurately than ERCP, it can underestimate the extent of hilar and distal ductal lesions.[191]

Recent developments in MR sequences include diffusion-weighted imaging (DWI) which utilises Brownian motion of water molecules and their diffusion characteristics within biological tissue *in vivo*. This has been shown to help distinguish benign from malignant lesions where cancerous lesions typically show a restriction in movement of fluid, likely due to the high cellular density and consequent effect of cell membrane restriction of water movement.[192] This degree of diffusion is calculated as an apparent

diffusion coefficient (ADC) and can be displayed graphically as well as numerically. However, the sensitivity changes with differing diffusion sensitivity coefficient (b value) scan parameters.[193] A recent study by Cui *et al*, assessing CCs with DWI and histopathological correlation, concluded that the b-value of 800 s/mm^2 was the best in DWI of extrahepatic cholangiocarcinoma and that the lesion ADC value declined as the degree of cancerous tissue differentiation decreased.[194] Overall, the combination of cholangiography and extra-ductal information provided by a MRI/MRCP, with all the additional information with contrast enhancement and DWI, makes it the investigation of choice for diagnosis and local staging of CC.[133] Of course, MRCP offers no opportunity for tissue sampling, or therapeutic intervention (*cf*. ERCP).

1.8 Endoscopic techniques in the diagnosis and staging of CC

1.8.1 Endoscopic retrograde cholangiopancreatography (ERCP)

ERCP is still considered the gold standard for cholangiography and is widely available in developed countries. After passing a duodenoscope to the ampulla of Vater, the bile duct is cannulated and contrast is instilled under endoscopic and fluoroscopic control. High-resolution cholangiograms are obtained in which biliary strictures can be reliably detected and differentiated from other obstructing pathologies such as choledocholithiasis. Differentiation between benign and malignant strictures can remain problematic. Features that favour malignancy include stricture length of >10mm, a ragged contour, abrupt transition (shouldering) or fixed filling defect. Strictures at the hilum are highly suspicious and warrant particularly aggressive investigation for malignancy.[195]

Tissue sampling can be undertaken at ERCP using a variety of techniques including cytology brushing, forceps biopsy and bile aspiration for cytology. Additional devices can be introduced at ERCP, including a choledochoscope or RFA catheter (see sections below). Therapy, such as stent insertion, can be delivered at ERCP. However, ERCP provides only two-dimensional imaging of the bile duct and gives no extraductal

information.[133] It carries a significant risk of serious morbidity, particularly pancreatitis.[196]

1.8.2 Choledochoscopy (e.g. SpyGlass™ system)

Recent technological advances have reignited interest in this technique. The procedure requires insertion of a small, fibreoptic endoscope through the biopsy channel of a standard duodenoscope (or a percutaneous system) into the bile duct. First, second and third order bile ducts can be accessed and visualised. Modern systems permit biopsy of biliary lesions under direct endoscopic vision. Mucosal appearances that are suggestive of malignancy include nodularity, papillomatosis and ulceration.[197-199] These systems can facilitate stent insertion and targeted biopsy of specific strictures.[200, 201]

A number of case-series have been reported that examined the use of choledochoscopy to assess indeterminate strictures. A recent prospective study examined 76 patients with indeterminate biliary strictures after non-invasive imaging. 22/38 (58%) of malignant strictures were correctly characterised by ERCP alone. The addition of choledochoscopy with targeted biopsy increased the sensitivity for malignancy to 38/38 (100%). 7 of the 38 benign strictures were the result of PSC, and one of these cases was falsely categorised as malignant after choledochoscopy and biopsy.[202] A further case series of 255 indeterminate strictures claimed that relevant information was garnered by choledochoscopy and biopsy in 68% of cases, changing the diagnosis in 10%. However, malignancy was only demonstrated by choledochoscopy and biopsy in 10% of the lesions that were subsequently proven to be cancerous at excision.[203] In a study of 53 patients with PSC and indeterminate dominant stricture, sensitivity and specificity for CC both exceeded 90%.[199]

Some centres have employed confocal microscopy, which is established in the endoscopic assessment of gastric and colonic neoplasms, to assess biliary strictures at choledochoscopy. Fluorescein solution is instilled into the bile duct through the channel

of the choledochoscope. This is followed by a catheter-based microscope that allows visualisation of the pattern of cells and vessels in the stricture. Lesions are classified using criteria similar to those used in the confocal microscopic assessment of other GI lesions. In a group of patients with previously indeterminate lesions, sensitivity for malignancy of ERCP alone, with the addition of choledochoscopy or with the addition of choledochoscopy and confocal microscopy were 56%, 59% and 83% respectively.[204] Other groups have examined biliary strictures experimentally using narrow-band imaging technology.[205] These techniques of *in vivo* pathological classification show some promise but remain experimental at present.

Modern choledochoscopy appears very promising but has several disadvantages. The double scope systems are expensive and expertise is currently limited to a few highly specialist centres. The fibreoptic imaging bundle is extremely fragile and the image transmitted, at best, is suboptimal in comparison with the images produced by modern digital endoscopes. Choledochoscopy is invasive and gathers no information beyond the lumen of the bile duct and so does not contribute to staging of the disease. The risks of choledochoscopy exceed those of standard ERCP as the catheter is large and relatively inflexible, increasing risk of bile duct injury and cholangitis.

1.8.3 Endoscopic Ultrasound (EUS)

EUS requires a modified endoscope, with an ultrasound probe built into its tip. The bile duct can be imaged reliably to the bifurcation of right and left hepatic ducts, through the wall of the stomach or duodenum. EUS permits sampling of suspected CC by fine-needle-aspiration (FNA). The two largest case series examining EUS + FNA reported sensitivities of 86% and 89% for CC. However, this strategy is avoided in the assessment of potentially resectable lesions, as there is a risk of tumour seeding.[206, 207] EUS is highly sensitive for local nodal disease and this has been studied in trials of liver transplantation for CC. Of 47 patients, 17% were found to have local lymph node disease on EUS and FNA, precluding transplant.[206] The exclusion of such patients from inappropriate transplantation has contributed to hugely improved outcomes in

transplantation for irresectable CC. EUS is hampered by limited range so is unhelpful in imaging lesions extending proximal to the hilum and intrahepatic lesions. It is highly operator dependent and requires very costly systems.

1.8.4 Intraductal ultrasound (IDUS)

IDUS requires the passage of a catheter mounted US probe into the bile duct through an endoscope or percutaneous biliary drain. This allows radial imaging of the biliary tree. Various characteristics correlate with malignancy, particularly IDUS evidence of bile duct wall disruption.[208] A small (n=30) retrospective study demonstrated sensitivity and specificity of 92% in categorising strictures as benign or malignant – superior to ERCP in this particular study.[209] Use of IDUS to enhance cytology yield, in combination with ERCP, may improve overall accuracy to >90%.[209, 210] Some centres are now using contrast enhancement techniques with IDUS.[211] However, these studies on indeterminate biliary strictures have been on a heterogeneous group, with only small numbers of PSC patients included. Current IDUS technology has only limited ability to evaluate the layers of the bile duct and infiltration of surrounding structures such as lymph nodes.[212, 213] IDUS is only available in a very limited number of centres, and is highly operator dependent. The catheter based system is very delicate and expensive. Introduction of the catheter requires ERCP or PTC, with the attendant risks of these invasive procedures.

1.9 Screening for cholangiocarcinoma in high risk groups

Even though groups at high risk of CC may be readily identified, such as PSC patients and populations living in areas with high levels of endemic OV infection, there is currently no proven benefit in screening for CC in these groups. Criteria for an effective formal screening program were published by Wilson and Junger in 1968 and are retained by the WHO today. Two of these are not currently met in the case of CC risk groups; *i) that there should be an accepted and effective treatment available for patients with the disease ii) that the screening test detects disease at an earlier stage,*

when treatment is more effective.[214] Patients with advanced benign liver diseases (such as PSC, hepatolithiasis or chronic OV infection) patients are especially unlikely to be candidates for curative resection of CC. In most healthcare systems, the diagnosis of CC contraindicates liver transplant. A number of longitudinal studies of formal screening protocols in PSC have utilised CA19-9, ultrasound, MRCP and ERCP with cytology.[170, 215, 216] Although some protocols have achieved respectable sensitivity and specificity, outcomes remain unaltered. Many centres in resource-rich systems do carry out informal screening for CC in PSC – varying from regular LFT measurement to regular CA19-9 estimation to US or MR liver imaging. Such approaches are not currently evidence based. Areas of highest *OV* prevalence are generally healthcare resource constrained and the cost of imaging based CC screening programmes would be prohibitive. More cost effective public health strategies may include population screening for *OV* infection with faecal antigen testing and subsequent eradication, public education against high-risk dietary habits and efforts to eradicate animal reservoirs of *OV*.[217-219] In the future, the development of a clinically useful vaccine against *OV* would be of huge benefit, although this currently remains an aspiration.[218] Until refined diagnostic tools permit earlier diagnosis, or new treatments that improve outcome become available, formal screening programs for CC in those subpopulations at high risk of CC are unlikely to lead to improved survival.[220]

1.10 Staging of cholangiocarcinoma

Staging of CC is currently poor. Even in cases where curative resection is attempted, survival rates are low as CC is a highly invasive tumour with early vascular and perineural invasion. Imaging is the mainstay of pre-operative staging, supplemented with staging laparoscopy in indeterminate cases.[140] Perihilar lesions are particularly difficult to stage correctly and often resectability can only be determined at surgery.[221] Survival after diagnosis with CC is dependent on complete surgical resection with clear margins, known as R0 resection.[222-224]

Accurate staging is principally required to determine resectability, and thereby allocate individual patients to the most appropriate treatment pathway. Staging is also important for the design and conduct of clinical trials and in prognostication. A number of CC staging systems have been used, these are generally based on tumour-node-metastases (TNM) criteria, the American Joint Committee on Cancer (AJCC) TNM criteria being the most widely used. Such systems are based on some parameters that require pathological examination of an excised lesion for precise classification, although pre-operative imaging, biopsy or operative staging findings may be used to define a provisional stage prior to resection. It has become clear that different staging systems are required for intra-hepatic, peri-hilar and distal CC due the different patterns of tumour extension, prognosis and therapeutic approaches.[225] Historically, intra-hepatic CC (IHCC) has been staged according to the same criteria as hepatocellular carcinoma (HCC) using the 6th edition of the AJCC TNM system. The 7th edition of these criteria refined these for IHCC by adding vascular invasion, tumour number and lymph node status. Other systems, such as those of the National Cancer Centre of Japan (NCCJ) and Liver Cancer Study Group of Japan (LCSGJ), are also in widespread use.[226, 227] Unmodified, these systems are not appropriate for staging of extra-hepatic CC, as they do not take account of features of extra-hepatic CC that inform management such as precise ductal distribution of hilar CC; hepatic function; hepatic atrophy and vascular invasion. The Memorial Sloan-Kettering system adds biliary extent, vascular encasement and liver atrophy to the TNM criteria. Correlation with R0 resection, nodal spread and survival is somewhat improved using this system.[228] The 7th Edition of the AJCC/UICC system also incorporates these criteria and now addresses hilar extrahepatic and distal extrahepatic disease separately in its staging algorithm. All current staging systems for CC have their weaknesses and have proven imperfect in determining prognosis and response to therapy in patients with CC. As new prognostic factors emerge from multivariate analysis of different CC subtypes, further refinements of these imperfect systems are proposed. Expert groups are

currently refining and validating systems that include factors such as depth of invasion, jaundice and performance status.[229, 230] Even when staged as operable, clear margins are achieved in only 20-50% of patients undergoing curative resection for CC. In those patients with clear margins, the 5-year survival rate is 19-47%. Those undergoing resection with positive margins have 5-year survival of 0-12%.[222, 231] This indicates that that even with high quality expert imaging, current preoperative staging is poor and does not predict resectability of CC or survival.[232, 233]

1.11 Treatment

1.11.1 Surgery

Surgery is the only therapy that can cure CC. Survival correlates strongly with negative margins (R0 resection), absence of vascular invasion and absence of lymph node involvement.[158, 234-240] Resection is usually only attempted with curative intent, as palliative or debulking resections confer no survival benefit.[228] Contraindications to surgery include proven distant metastases, malignant ascites, bilateral vascular or biliary involvement, advanced cirrhosis or PSC. In a large case-series, resection was feasible in distal, intrahepatic and perihilar lesions in 91, 60 and 56% of cases respectively.[158] Mono-focal intrahepatic CC can be resected in a segmental fashion with 5-year survival rates of 22-44%. Surgery for peri-hilar lesions is determined by Bismuth-Corlette criteria (Figure 1.2). Patients with Bismuth I and II lesions undergo en bloc resection of the extrahepatic bile ducts, gall bladder and hilar lymph nodes with formation of a hepatojejunostomy. Type III and IV lesions require lobectomy or hemihepatectomy of the corresponding liver.[241-244] Distal lesions are resected with a pylorus preserving pancreaticoduodenectomy (Whipple's procedure).[158] R0 resection of peri-hilar CC carries a five-year survival of 11%-41% and for distal CC 27%-37%.[158, 223, 228, 245-247] Overall 5-year survival for R0 resections remains <50%.[245] It is likely that some of these studies underestimate current outcomes as survival rates after CC resection have improved in the last decade - most likely due to improved staging and patient selection.[248]

1.11.1.1 Surgery for PSC patients with CC

Resection of CC in patients with PSC has higher associated morbidity and mortality than in patients with sporadic CC in a normal background liver. The underlying liver function is generally inadequate to withstand resection of the required volume of liver. Resection for extrahepatic lesions requires the formation of a biliary-enteric anastomosis, which can exacerbate the on going cholangitis, and is an additional risk factor for CC.[249, 250] Finally, PSC patients with CC are likely to have multifocal CC, or at least multifocal biliary dysplasia, that makes early recurrence or development of a new lesion common. Outcomes for patients with PSC who develop intrahepatic or perihilar CC and undergo resection are dismal.[251, 252]

1.11.1.2 Orthotopic liver transplantation (OLT)

Historically, orthotopic liver transplant (OLT) for unresectable CC achieved appalling outcomes, with early tumour recurrence and very short survival the rule. In the last two decades, a handful of centres in the United States of America (USA) have attempted liver transplantation in highly selected patients with CC. Protocols comprise extensive preoperative assessment, including EUS with lymph node FNA and a staging laparotomy with lymph node biopsy. Suitable candidates progress to preoperative chemotherapy, external beam radiotherapy and intraductal brachytherapy. Patients with irresectable disease confined to the liver, and who tolerate the neoadjuvant regime, undergo transplantation. In the largest published series, the cohort that actually underwent transplantation had 1-, 3-, and 5-year survival after OLT of 90%, 80%, and 71%. These figures are comparable to those achieved with OLT for other indications.[253] The excellent survival figures are, in part, a result of the meticulous selection of appropriate cases. Even in centres where it available, only a minority of patients with CC actually undergoes OLT. Encouragingly, OLT seems to have been particularly successful in patients with PSC related CC where resection is almost never successful.[253-258] The strict exclusion criteria, the cost of protocols and the small donor pool mean that OLT is likely to remain an option for only a small proportion of

CC patients. Where permitted and available, live related liver donation may mitigate some of this restriction. For now, CC remains a contraindication to OLT in most healthcare systems worldwide – including patients with PSC related CC. Outcome data from the more numerous USA and European centres that are now using OLT in CC is emerging, demonstrating comparable outcomes.[259]

1.11.1.3 Neoadjuvant therapy

Neoadjuvant chemoradiotherapy is rarely feasible as CC patients are often jaundiced with poor performance status at presentation. Small series of patients have been studied with complete responses in 20-30% and partial responses in the remainder.[260, 261] Cases of irresectable CC being downstaged and then resected successfully have been reported.[261] However, prospective clinical studies are required and neoadjuvant therapy remains feasible in only a minority of patients.

1.11.1.4 Adjuvant therapy

Adjuvant chemotherapy, in conjunction with R0 resection of CC, has not been proven to improve recurrence or survival rates.[260] Many centres offer 5FU or gemcitabine based adjuvant chemotherapy and this is the subject on ongoing clinical trials. Routine adjuvant chemoradiotherapy has been studied in small retrospective case-series, which have suggested a modest survival benefit.[262, 263] However, a prospective RCT of a mixed group of biliary, ampullary and pancreatic neoplasms showed no benefit of chemoradiotherapy in the CC subgroup.[264] In patients with positive margins at surgery, there is stronger evidence in support of adjuvant chemoradiotherapy. Small retrospective series have demonstrated improved 5-year survival (34 vs. 14%) and median survival (21 months vs. 8 months).[265, 266] Postoperative chemoradiotherapy is therefore generally recommended to patients with positive resection margins.[157]

1.11.2 Palliative therapy

In the 50-90% of CC patients that have unresectable disease at diagnosis, palliative therapy is the mainstay.[158] With full palliative treatment, median survival of 6 months is achieved, versus 2 months with expectant care.[267] The major morbidity is related to

biliary obstruction. The resultant jaundice can be debilitating and obstruction rapidly leads to deterioration of liver function and death.[268]

1.11.2.1 Endoscopic palliation

1.11.2.1.1 Stents

Endoscopic insertion of biliary stents is the principal palliative therapy for CC. Plastic or metal stents can be deployed through strictures at ERCP and can reliably decompress the biliary tree. The choice of metal or plastic stent depends on the security of the malignant diagnosis, the predicted prognosis and local resources. Metal stents do not improve survival compared to plastic.[269] Plastic stents have an average patency of 2 months, with metal stents lasting up to a year.[269, 270] Uncovered metal stents are very difficult to remove once placed, so are only used when unresectable CC has been proven.[271] Increasingly, fully covered metal stents are being placed as the primary endoscopic management of biliary strictures distal to the liver hilum. These maximise biliary decompression, minimise risk of stent occlusion and do not complicate future surgery. They can be removed endoscopically and have proven useful in benign stricturing disease. Covered metal stents have a slightly higher rate of migration that uncovered metal stents and cannot be deployed across the hilum, as contra-lateral biliary obstruction may result.[270, 272]

1.11.2.1.2 Photodynamic therapy (PDT)

After systemic delivery of a photosensitising hematoporphyrin derivative (HPD) such as Photosan-3® or Photofrin II®, high intensity light at a wavelength of 630 nm is applied to the bile duct lumen via a catheter.[273] This induces injury to any exposed biliary lesion to a depth of 6-8mm. PDT in generally undertaken before immediate placement of a metal stent. Improvement in duration of stent patency, and patient survival, has been demonstrated.[274, 275] Side effects include a higher rate of cholangitis than is seen in standard ERCP and the unwanted dermatological sequelae of photosensitisation.[13] The HPD drugs have become increasingly expensive in recent years, limiting their use in some healthcare systems. The most recent large clinical trial of PDT in CC,

published only in abstract form to date, had disappointing results, with earlier mortality in the PDT group compared to the control arm. Overall survival was 5.6 months for PDT plus stenting versus 8.5 months for stenting alone (HR 1.82, p=0.027). Although overall survival was improved among those who had chemotherapy compared to those who did not (11.1 vs. 4.8 months, p=0.001), adjusting for this only reduced the PDT plus stenting hazard ratio from 1.8 to 1.6, suggesting that failure to receive subsequent chemotherapy did not fully explain the excess risk in the PDT cohort. Some commentators suggest that flaws in the study methodology may have resulted in a spurious finding of inferiority in the PDT group but, in the absence of further RCT data, PDT should be regarded as of doubtful benefit in CC.[276]

1.11.2.1.3 Intraductal radiofrequency ablation (RFA)

Recent technical advances, led by engineers and clinicians at Imperial College London, have permitted the direct delivery of radiofrequency ablation (RFA) into the bile duct. This is undertaken at ERCP and requires the transpapillary introduction of a specially designed electrode catheter. RF energy is delivered, causing heating and coagulative necrosis of adjacent tissues, to a depth of 5-8mm. Initial studies have confirmed utility and safety of this approach when combined with stent insertion.[272] Unblocking of obstructed metal stents has also been demonstrated. There is a suggestion that stent patency and duration of patient survival may be improved.[277, 278] Randomised controlled trials are awaited. The principal advantage of the RFA approach over PDT is the avoidance of systemic photosensitisation, with the associated costs and side effects.

1.11.2.2 Radiological palliation

Percutaneous transhepatic cholangiography (PTC), with placement of drains (PTD) or metal stents, is frequently necessary to relieve jaundice. This is generally indicated when ERCP and stent delivery is impossible due to anatomical factors, such as duodenal obstruction, or patient factors, such as cardiorespiratory instability.[140] PDT and intraductal RFA catheters can also be deployed via the percutaneous route. Direct

percutaneous RFA of peripheral IHCC is also feasible, as in HCC management. Radiologic embolisation of extensive intrahepatic disease with chemotherapeutic or ß-particle emitting microspheres (yttrium-90) has also been reported.[279]

1.11.2.3 Surgical palliation

There is no role for surgical debulking.[242] Laparoscopic biliary bypass surgery is sometimes required for palliative relief of jaundice. However, ERCP or PTC decompression is generally preferred as the associated morbidity is lower.[280]

1.11.2.4 Oncological palliation

Chemotherapy with gemcitabine and cisplatin has utility in patients with unresectable CC, with prolongation of median survival from 8 to 26 weeks.[267] Studies have incorporated intrahepatic CC, hilar CC, gallbladder cancer and extrahepatic CC as one entity. This is a flaw, and expert groups are now advocating separate clinical studies in each of these groups. Palliative external beam radiotherapy, with or without chemotherapy, may reduce biliary obstruction and pain but does not confer any survival benefit.[262, 281-283] Small studies of high dose radiotherapy with chemotherapy, including intraductal brachytherapy, have shown increased 2-year survival (48 vs 0%).[284] However, prospective studies have not been performed and complication rates with high intensity chemoradiotherapy are high.

1.12 Published guidelines for the diagnosis and management of CC

Detailed guidelines for the diagnosis and management of CC were first published by the British Society of Gastroenterology (BSG) in 2002.[140] A summary of the diagnostic criteria for CC included in this guideline is provided in Table 1.2. An updated version of these guidelines was published in 2012. Other guidelines have been published by learned societies including the European Society for Medical Oncology (ESMO) guidelines on the diagnosis and management of CC (2011) and the IHPBA, AHPBA and EHPBA consensus documents on CC published in 2008.[285-287] The use of such

guidelines is helpful, but clearly local policies, patient needs and healthcare economics must be considered in their interpretation.

1.13 Summary: Unmet needs and thesis aims

Current understanding of the aetiology of cholangiocarcinogenesis is limited. Our knowledge of the aetiology of *sporadic* CC is particularly poor. Diagnostic and staging tests are inadequate. These factors combine in the current appalling prognosis of this disease. My overarching aim is to contribute to the knowledge of the aetiology of cholangiocarcinoma and to enhance its diagnosis and staging. In my thesis I will explore and elucidate the following three specific areas.

1.13.1 Germline genetic risk factors for CC

It is likely that genetic factors affect risk of CC development. Incidence of CC varies with ethnicity – independent of geography or environment. Genetic diseases such as Lynch syndrome and bile salt transporter defects carry a high risk of CC development. A number of genetic factors have been strongly implicated in the pathogenesis of PSC and cholelithiasis, the two major risk factors for CC in the West. NKG2D has been shown to be associated with CC risk in PSC patients. Greater understanding of relevant polymorphisms may clarify the pathogenesis of CC, aid diagnosis and perhaps permit identification of an at risk subpopulation. In Chapters 2 and 3 of my thesis I aim:

i) To investigate the role of genetic variation in biliary transporter proteins as a risk factor for sporadic cholangiocarcinoma

ii) To investigate the role of genetic variation in the NKG2D receptor as a risk factor for sporadic CC

1.13.2 Plasma protein markers of CC

CC is notoriously difficult to diagnose, particularly on the background of chronic bile duct inflammation, the major known predisposing factor. Liver transplantation is accepted therapy for HCC or severe PSC. Conversely, once CC has developed (as it does in up to 30% of PSC cases) transplantation is contraindicated. Therapy for these

diagnoses is radically different, potentially involving surgical resection, liver transplantation, endoscopic intervention, chemotherapy or palliation. Serum tumour markers CA19-9 and CEA have poor specificity and sensitivity and are usually only positive once the disease has become unresectable.[140] Therefore, there is a need for sensitive and specific biomarkers that become positive at an early stage of the disease and can distinguish CC from benign biliary disease, as well as from other liver tumours. In Chapters 4 and 5 of my thesis I aim:

i) To establish protein profiles of plasma from patients with cholangiocarcinoma and define potential new biomarkers

ii) To investigate the potential of neutrophil gelatinase-associated lipocalin (NGAL) as a plasma biomarker of CC

1.13.3 Deficiencies in the imaging diagnosis and staging of CC

Diagnosis of CC currently relies on imaging with CT or MRI.[140] However, these modalities cannot reliably distinguish benign from malignant biliary strictures, or differentiate CC from other liver tumours, such as metastases or HCC. Current imaging technology is inadequate in the staging of CC, leading to some patients with irresectable disease undergoing futile radical surgery, others undergoing inadequately radical surgery - reducing the chance of R0 resection. In Chapter 6 of my thesis I aim:

i) To develop, validate and optimise a miniature detector coil for intraductal MR imaging of the biliary tree

ii) To develop, validate and optimise a MR compatible duodenoscope, with integrated MR detector coils, for imaging of the biliary tree

iii) To integrate 1 & 2 in an imaging system and demonstrate utility in imaging *ex vivo* specimens

1.14 Chapter 1 Tables and Figures

Table 1.1: Summary of studies of genetic polymorphisms as risk factors for CC

Gene polymorphism	Study	Function	Country	Cases - n	Controls - n	Odds ratio (95% CI)
GST01*A140D	Marahatta *et al*, 2006[66]	Detoxification	Korea	30	98	8.5 (2.1-37.8)
NAT2* 12,*6B,*7A	Prawan *et al*, 2006[69]	Detoxification	Thailand	216	233	0.38 (0.1-0.9)
CYP1A2	Prawan *et al*, 2006[69]	Detoxification	Thailand	216	233	0.28 (0.1-0.9)
MTHFR 677CC/ TSER 2R^{+}	Ko *et al*, 2006[68]	Folate metabolism	Korea	47	204	5.38 (1.2-23.6)
XRCC1 194W	Huang *et al*, 2008[288]	DNA repair	China	127	786	1.9 (1.1 – 3.5)
XRCC1 R280H	Huang *et al*, 2008[288]	DNA repair	China	127	786	0.5 (0.3-0.9)
NKG2D rs11053781	Melum *et al*, 2008[72]	NK cell activation	Norway	46 (PSC + CC)	319 (PSC)	2.08 (1.31-3.29)
NKG2D rs2617167	Melum *et al*, 2008[72]	NK cell activation	Norway	46 (PSC + CC)	319 (PSC)	2.32 (1.47-3.66)
MICA 5.1 allele	Melum *et al*, 2008[72]	NKG2D ligand	Norway	46 (PSC + CC)	319 (PSC)	0.3 (0.20-0.95)
***PTGS2* Ex10 +327T>C**	Sakoda *et al*, 2006[289]	COX-2 enzyme	China	127	786	1.8 (1.2-2.7)
α1AT rs28929474	Mihalache *et al*, 2011[70]	α1AT function	Germany	182	350	2.46 (1.14–5.32)
MRP2 c.3972C>T	Hoblinger *et al*, 2009[71]	Biliary xenobiotic excretion	Germany	60	73	1.83 (1.1-3.1)

Table 1.2: Summary of clinical features and recommended investigations for diagnosis, BSG Guideline on CC 2001[140]

Clinical Features	*Serum*	*Radiology*	*Invasive Investigations*
Jaundice	Bilirubin	USS	ERCP
Pale Stools	ALP	Contrast triple phase helical CT	PTC
Dark Urine	GGT	MRI	EUS
Pruritis	ALT/AST	MRCP	Histology
Malaise	CA-125	(PET)	
Fatigue	CA19-9		
Loss of Appetite	CEA		
Weight Loss	(IgG4)		

Figure 1.1: Molecular events in cholangiocarcinogenesis

Legend: Chronic biliary inflammation and cholestasis lead to cellular changes leading to cholangiocarcinoma development and progression. Inflammatory cytokines induce iNOS (inducible nitric oxide synthase) in cholangiocytes, favouring mutagenesis, impaired DNA repair, and COX-2 up-regulation. Aberrant bile acid signalling contributes to the promotion of cholangiocyte growth via transactivation of the epidermal growth factor receptor (EGFR), and is also linked to activation of the COX-2 pathway. Alterations in various genetic and epigenetic molecular pathways selectively depicted in the figure then combine to promote the features of immortalization, evasion of apoptosis, autonomous proliferation, and invasion and metastasis characteristic of malignant cholangiocytes. MMP, matrix metalloproteinase; PDGF, platelet-derived growth factor; TGF, transforming growth factor; TNF, tumour necrosis factor. Adapted from Sirica et al 2008[79].

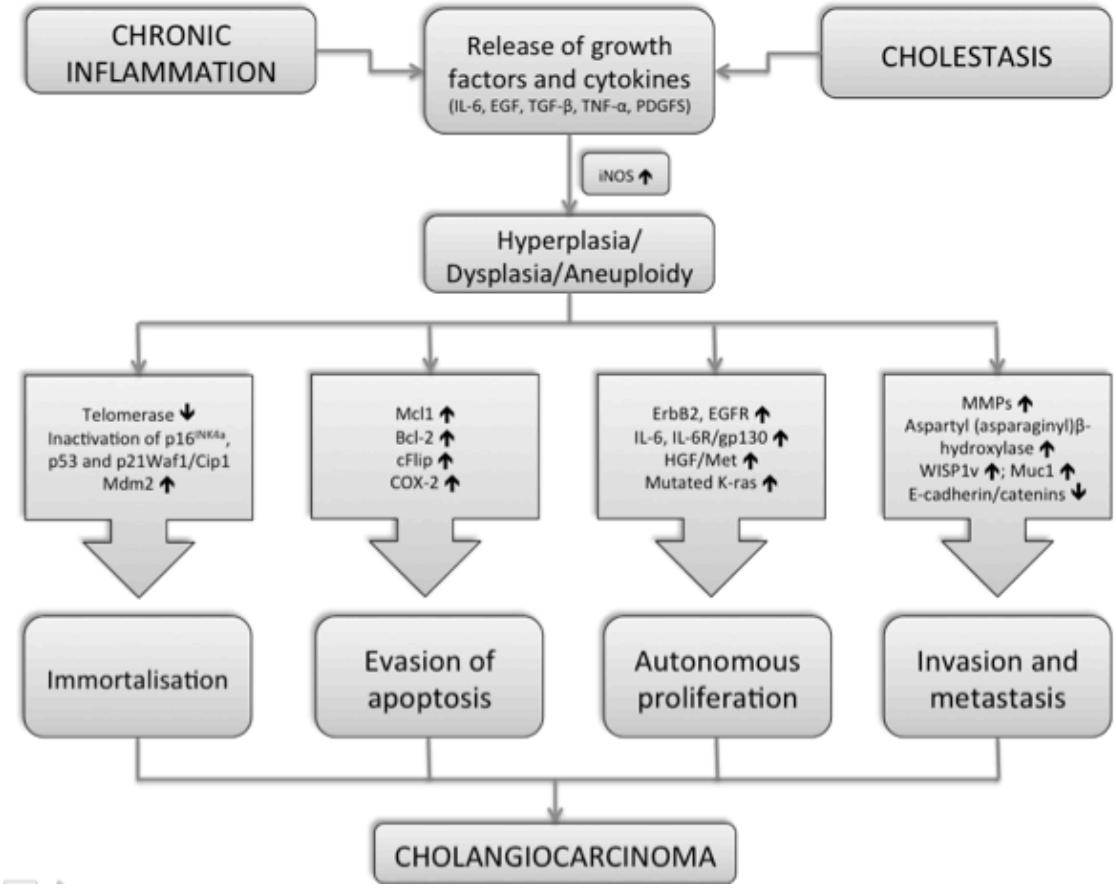
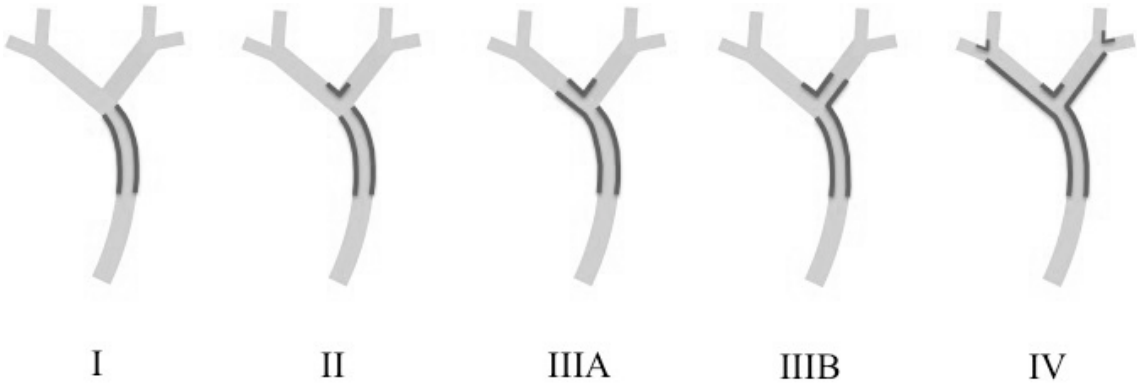


Figure 1.2: Bismuth-Corlette classification of hilar CC

Legend: This categorisation is based on ductal extent of the lesion. **Type I:** limited to the common hepatic duct, not involving the confluence of right and left hepatic ducts. **Type II:** involves the confluence. **Type IIIa:** type II plus involvement of the right hepatic duct. T**ype IIIb**: type II plus involvement of the left hepatic duct. **Type IV**: extending to both right and left hepatic ducts. (Adapted from Sainani et al, 2008[290])



2. Polymorphisms of biliary transporter genes as risk factors for CC

2.1 Background

As discussed in Chapter 1, a variety of different diseases and environmental factors are known to increase an individual's risk of CC. Some of these diseases, such as hepatitis C and gallstone disease, are very common. The hepatobiliary system is the major route of metabolism and excretion for exogenous and endogenous toxins, many of which are genotoxic and potentially carcinogenic. Exposure to environmental toxins has increased in the past few decades, and may well play an important role in cholangiocarcinogenesis.[291-293] Exposure to exogenous toxins is widespread, and biliary exposure to endogenous toxins such as hydrophobic bile acids is universal. Despite these highly prevalent disease and environmental risk factors, only a small percentage of Western populations develop CC. In common with many other complex diseases, it is highly likely that host susceptibility factors, many of which are genetically determined, will modify an individual's response to these stimuli and their risk of cholangiocarcinogenesis. A number of genetic diseases that are risk factors for CC, including abnormalities in bile duct transporter proteins, have already been reviewed in section 1.2.3, where I also discuss the expanding evidence of strong genetic influences on risk of gallstone disease and PSC.

Hepatobiliary excretion of bile salts and toxins is performed by a variety of transporters that are expressed on the apical surface of hepatocytes and cholangiocytes.[294] These transporters, their substrates and functions in the excretion of bile constituents are discussed in the following section and are illustrated in Figure 2.1 (adapted from Geier *et al*, 2006[295]). Their names, encoding genes, known functions and associated diseases are listed in Table 2.1 (adapted from Wadsworth *et al*, 2011[296]).

The function of these transporters governs the rate of bile flow and their dysfunction is known to cause cholestasis and biliary injury. Bile flow and constituents vary between

individuals. Genetic polymorphisms in biliary transporters can lead to reduced expression, or function, or both.[297, 298] A reduction in the flow of bile (cholestasis) may result in increased exposure of cholangiocytes lining the biliary epithelium to xenobiotics and endogenous mutagens, such as hydrophobic bile acids. Hydrophobic bile acids have a strong detergent action that disrupts cell membranes. The formation of bile acids, phosphatidyl choline (PtC) and cholesterol into stable micelles in the bile protects the hepatobiliary system from damage. There is extensive experimental evidence that relative under or over representation of bile constituents, such as PtC and bile acids, leads to unstable biliary micelles and thereby failure of physiological biliary protective mechanisms. The sequelae of this are deconjugation of toxic species leading to the formation of 'toxic bile' and bile duct damage.[294, 299]

2.1.1 Biliary canalicular transporter proteins

2.1.1.1 Bile salt exporter pump (BSEP, ABCB11)

BSEP, encoded by the *ABCB11* gene, is a member of the superfamily of ATP-binding cassette (ABC) transporters. ABC proteins transport various molecules across extra- and intra-cellular membranes. BSEP is responsible for the active transport of bile acids across the hepatocyte canalicular membrane, and secretion of these bile acids is the major determinant of bile flow.[49, 297] Under-expression of BSEP is implicated in progressive familial intrahepatic cholestasis Type 2 (PFIC2).[300] PFIC2 has been associated with very early onset of HCC and CC in children.

2.1.1.2 Multidrug resistance protein 3 (MDR3, ABCB4)

MDR3 is encoded by the *ABCB4* gene and is a phosphatidylcholine floppase that translocates PtC from the inner to the outer leaflet of the canalicular membrane. PtC is critical to the formation of lipids and bile acids into stable micelles in the bile. Polymorphisms in *ABCB4* can lead to a reduction in MDR3 protein function and biliary PtC levels, resulting in unstable micelle formation.[301] MDR3 deficiency has been implicated in PFIC Type 3 (PFIC3).[302]

2.1.1.3 Multidrug resistance-associated protein 2 (MRP2, ABCC2*)*

MRP2 is encoded by the *ABCC2* gene, is a member of the MRP subfamily and plays a key role in drug elimination and drug resistance. It is expressed in the canalicular, apical membrane of the hepatocyte and exports numerous conjugated anionic species into the bile, including bilirubin and many drugs and toxins.[294, 303] A recent study in a German cohort reported an association between a polymorphism in *ABCC2* and CC.[71]

2.1.1.4 Familial intrahepatic cholestasis protein 1 (FIC1, ATP8B1*)*

FIC1 is a member of the P-type cation transport ATPase family and belongs to the subfamily of aminophospholipid-transporting ATPases. It is highly expressed in cholangiocytes and the canalicular membranes of hepatocytes and is also found in the small intestine.[304] FIC1 transports phosphatidylserine and phosphatidylethanolamine from the outer membrane leaflet to the inner, maintaining the correct distribution of these lipids in the membrane and thereby membrane integrity.[305] FIC1 dysfunction is associated with the disease PFIC Type 1 (PFIC1).[306, 307]

2.1.1.5 Farnesoid X receptor (FXR, NR1H4*)*

FXR is encoded by the *NR1H4* gene. It is a nuclear receptor and is expressed at high levels in the human liver and intestine. It is not a biliary canalicular transporter protein *per se*. However, chenodeoxycholic acid and other bile acids are natural ligands for FXR. When activated, FXR translocates to the cell nucleus and binds to hormone response elements on DNA. It thereby up- or down-regulates the expression of other genes. FXR suppresses the production of cholesterol 7 α-hydroxylase (CYP7A1), which is a crucial enzyme in bile acid synthesis. Synthesis of bile acids is thereby inhibited, when cellular bile acid levels are high.[308] Importantly, FXR also modifies the expression of other genes, including *ABCB11* and *ABCB4*.[309, 310]

2.1.2 Rationale for study

Genetic polymorphisms in biliary transporters are known to lead to reduced transporter expression, or function, or both. Subsequent disequilibrium of excreted biliary constituents leads to cholestasis, micellar instability, failure of biliary epithelial

protective mechanisms and the formation of toxic bile – with deconjugation of genotoxic species.[294] Such a genetic predisposition may enhance the effect of other risk factors - known or unknown, genetic, disease or environmental.

2.2 Hypothesis and study aim

2.2.1 Hypothesis

I hypothesised that variation in the genes that encode the biliary transporters BSEP, MDR3, MRP2 and FIC1, and the nuclear receptor FXR, is associated with altered susceptibility to sporadic cholangiocarcinoma.

2.2.2 Study aim

To investigate the role of polymorphisms in BSEP (*ABCB11*), MDR3 (*ABCB4*), MRP2 (*ABCC2*), FIC1 (*ATP8B1*) and FXR (*NR1H4*) as risk factors for sporadic cholangiocarcinoma.

2.3 Materials and methods

2.3.1 Ethics

The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki. I sought and obtained approval of this study from the St Mary's Research Ethics Committee and NHS R&D Department (Ref 09/H0712/82).

2.3.2 DNA from CC patients

I collected blood samples from 172 subjects with CC. Cases were confirmed by a) pre- or post-operative histology or b) multidisciplinary team consensus on the basis of ≥2 imaging modalities, clinical course and serum markers. All cases collected were from Caucasian patients.

2.3.2.1 Prospective collection

I collected blood samples prospectively from 44 consecutive consenting patients with CC. These subjects were recruited at Imperial College Healthcare NHS Trust.

2.3.2.2 Archived samples

I obtained blood samples from 128 subjects with CC from the hepatobiliary biobank archives of Imperial College Healthcare NHS Trust and University College Hospitals NHS Foundation Trust. These blood samples had been collected as part of other recent studies into CC and were stored with appropriate donor informed consent for retention and use in future, REC approved, studies.

2.3.3 Control DNA

I obtained blood samples from 265 control subjects with no liver disease. These were archived blood samples from a prior study, with informed consent for retention and use in future, REC approved, studies. The control blood samples were all from Caucasian patients and I selected them to form a cohort that was as well matched to the case group as possible in terms of sex and age.

2.3.4 Power calculation

I undertook an *a priori* power calculation, using statpages (Statpages.org, Albemarle, USA)[311].

The SNPs selected had a minimum mean allele frequency (MAF) ≈10%. Therefore the expected control group allele frequency was set at 0.1, case group allele frequency of 0.2 (i.e. an estimated odds ratio of 2). A required power of ≥80% to detect any difference, with a p value <0.05, was set. Case and control groups were initially assumed to be of equal size.

Using standard formulae[312], the number (*n*) required in each group was calculated as:

$$n = \frac{[p_1(1-p_1) + p_2(1-p_2)]}{(p_1 - p_2)^2} \times c_{p^1power}$$

$$n = \frac{[0.1(1-0.1) + 0.2(1-0.2)]}{(0.1 - 0.2)^2} \times 7.9 = 197.5$$

Where p_1 is the expected/known proportion in control population; p_2 is the estimated proportion in case population (assuming a difference); $c_{p1power}$ is a constant defined by the values chosen for the desired *P* value and power (defined by a standard nomogram, common values shown in Table 2.2[313])

I anticipated collecting 160 case samples, and had ready access to a larger number of control than disease samples. Therefore, I further calculated for unequal sample sizes:

$$k = n / (2n_o - n)$$

$$k = 197.5 / ((2 \times 160) - 197.5)$$

$$k = 1.61$$

$$kn_o = 1.61 \times 160 = 257.6$$

Where *n* is the calculated required (equal group) sample size, n_0 is the known minimum case group size, kn_o is the size of the required expanded control group, to maintain the same power (80%) at the same alpha (p=<0.05).[314]

So, with a CC group of at least 160 and a control group of at least 258, the study had 80% power to detect the hypothesised difference between cases and controls.

2.3.5 DNA preparation

I used two different DNA extraction techniques on samples from CC patients, depending on the provenance of the sample. The format of collection and storage was different for archived and prospectively collected samples, mandating different approaches to DNA extraction.

DNA was extracted from the control blood samples using a third technique. The format of these samples was different (1-2mL of whole blood) and time pressures meant that I was unable to extract the DNA from these myself. KBioscience Ltd (Hoddesdon, Hertfordshire, UK) undertook these extractions.

All three techniques used are established, validated proprietary systems. As DNA is highly conserved, introduction of any bias due to differing extraction technique is highly unlikely. I undertook a number of DNA quality control (QC) measures, and these are outlined in 2.3.5.4, below. These indices, and the genotyping failure rates, were compared between samples collected and processed with different techniques.

2.3.5.1 Extraction of DNA from prospectively collected blood samples

I collected samples prospectively from CC subjects using a sample collection SOP, which I authored (Appendix 1). The sample for genetic studies was taken into an EDTA tube, in which the blood does not clot. On centrifugation, a buffy coat layer can be readily identified and isolated. I extracted DNA from the blood leucocytes within the buffy coat layer using the QIAamp DNA blood mini kit (QIAGEN Ltd, Crawley, UK) using adapted manufacturer's protocols.[315] This is a kit based, column adhesion technique. It is an economic, high throughput system, but is only applicable to samples that have been anticoagulated at collection.

2.3.5.1.1 Method of DNA extraction from prospectively collected samples

I adapted this methodology from the reagent manufacturer's published guidance.[315] All reagents were obtained from QIAGEN.

i. After collection, each 5mL EDTA blood sample was centrifuged at 2000g for 10 min, before removing the buffy coat layer with a pipette. This was frozen at -20^{o}C in readiness for batch DNA extraction

ii. 30 min before DNA extraction, samples thawed to room temperature

iii. 200μL of buffy coat placed into a centrifuge tube and added 20μL of QIAGEN proteinase K solution

iv. 200μL of Buffer AL added to each sample and vortexed for 15s

v. Sample incubated at 56^{o}C for 10 min

vi. 200μl of ethanol added to the samples, before vortexing for 15s

vii. Resulting solution transferred into the QIAamp mini spin column, held within a 2mL collection tube

viii. The collection tube, column and solution were centrifuged at 6000g for 1 min, drawing the solution through the column

ix. The filtrate was discarded, and 500μL of buffer AW1 added to the top of the column. This was again centrifuged at 6000g for 1 minute, within a clean 2mL centrifuge tube

x. The filtrate was discarded, 500μL of buffer AW2 was added to the top of the column and the column centrifuged at 6000g for 1 min, within a clean 2mL centrifuge tube

xi. The filtrate was discarded. 200μL of buffer AE was added to the column and incubated within a 1.5mL centrifuge tube for 5 min. The column was then centrifuged at 6000g for 1 min

xii. The resulting filtrate, containing the eluted DNA in buffer AE, was stored at -20^{o}C, pending preparation for genotyping

2.3.5.2 Extraction of DNA from archived blood samples from CC patients

The archived samples from CC subjects had been collected in serum separation tubes (SST) - in which blood clots. This initially posed a problem, as DNA extraction from mature clot usually requires manual or ultrasonic fragmentation of the sample. DNA yields from such samples are often low. Moreover, such techniques are unsafe, as they require the use of a scalpel blade on the clot, or nebulisation of the sample with increased risk to the operator of mucosal splash.

I identified a safer, high yield technique in the literature. This required the use of Clotspin Baskets and the Gentra Puregene Blood Kit (QIAGEN). This system allowed me to use a centrifuge and a basket within the sample processing tube to shred the clotted sample. The manufacturer states that the protocol permits purification of high

molecular weight DNA with purification steps that remove contaminants. They state that purified DNA typically has an A260/A280 ratio between 1.7 and 1.9 and is up to 200kb in size. The protocol is optimised for purification of DNA from 5–7mL of clotted blood. I obtained the protocol from the QIAGEN website, to which no modifications were required. I trialled the technique in six samples successfully, and confirmed DNA yield and quality using low volume spectrophotometry.

Once I had demonstrated its utility, I applied this extraction technique to all the archived samples that had been collected in SST tubes.

2.3.5.2.1 Method for DNA extraction from archived samples

I adapted this methodology from the manufacturer's published guidance.[316] All solutions and reagents used in the following protocol were purchased in kit form from QIAGEN.

i) Each frozen, clotted blood sample was thawed out quickly in a water bath at 37°C and then placed on ice. The SST tube was inverted several times to loosen the clot

ii) The blood clot was poured into the Clotspin Basket, held within a 50mL tube

iii) The blood sample was centrifuged, within the basket and 50mL tube, at 2000g for 10 min. This forced the clot through the perforations in the basket, shredding it

iv) 15mL of red blood cell lysis solution was added to the basket to rinse residual clot material through

v) The basket was removed from the 50mL tube and a pipette used to transfer remaining clot material from the basket into the 50mL tube, before discarding the basket

vi) The resultant solution was vortexed for 3 s and then placed on a tube rotator for 5 min at room temperature (15–25°C)

vii) The tube was vortexed again for 3 s, at high speed

viii) The solution was centrifuged again, at 2000g for 5 min, to form a pellet of white blood cells and clot fragments at the bottom of the tube

ix) The supernatant was poured off, retaining the pellet in the tube

x) A further 5mL of red blood cell lysis solution was added to the pellet, and vortexed for 3 s at high speed

xi) The tube was incubated on a rotator for 5 min at room temperature

xii) The sample was centrifuged, at 2000g for 5 min, to pellet the white blood cells and clot fragments

xiii) The supernatant was discarded, leaving behind 200µL of residual liquid and the pellet

xiv) The tube was vortexed vigorously at high speed to disperse the pellet in the liquid

xv) 5mL of cell lysis solution and 25µL of Puregene Proteinase K (20mg/mL) were added and the solution vortexed for 10 s

xvi) The sample was incubated at 55°C overnight, vortexing for 10 s three times during the incubation

xvii) To terminate cell lysis and proteinase activity, the samples were incubated on ice for 5 min

xviii) 1.67mL of protein precipitation solution was added to the cell lysate, and vortexed for 20 s

xix) The sample was centrifuged at 2000g for 10 minutes and then incubated on ice for 2 min

xx) 5mL of isopropanol and 10µL of glycogen solution (20 mg/mL) were pipetted into a fresh 50mL centrifuge tube

xxi) The supernatant from step xix was decanted into the isopropanol/glycogen solution, discarding the pellet of precipitated protein and the original tube

xxii) The solution was mixed by inverting the tube 50 times

xxiii) The tube was centrifuged at 2000g for 3 min, at which point the DNA became visible as a small white pellet at the bottom of the tube

xxiv) The supernatant was discarded and the tube drained onto a clean filter paper, retaining the DNA pellet in the tube

xxv) 5mL of 70% ethanol was added to the tube, with washing of the DNA pellet by inverting the tube 5 times

xxvi) The solution was centrifuged at 2000g for 1 min

xxvii) The supernatant was discarded, again draining the tube onto a clean filter paper, retaining the DNA in the tube

xxviii) The DNA was allowed to air dry at room temperature for 10 min

xxix) 500μL of DNA hydration solution was added to the DNA pellet

xxx) The DNA was incubated at 65°C for 1 hour until dissolved

xxxi) The DNA was incubated overnight at room temperature overnight with gentle shaking before transferring to archive tubes and freezing at -20^{o}C, pending preparation for genotyping

2.3.5.3 Extraction of DNA from archived blood samples from control subjects

KBioscience undertook DNA extraction from the archived control samples using a well-established, robotised technique. The cells were lysed chemically and the resultant solution passed through a column of DNA binding silica particles, mediated by the chaotrope guanidinium isothiocyanate. After rinsing with a sequence of different buffers to remove any contaminants, the DNA was eluted into a low salt buffer. Samples were then transferred to storage tubes, and frozen at -20^{o}C pending genotyping.

2.3.5.4 DNA quality control (QC)

I performed a number of DNA QC measures. These permitted me to compare DNA quality and yield between extraction techniques and prepare correct concentrations of samples for genotyping – reassuring me that the subsequent genotyping step would be successful.

2.3.5.4.1 Low volume spectrophotometry

I performed Nanodrop low volume spectrophotometry (Thermo Fisher Scientific Inc. Waltham, MA, US) on all samples. This system quantifies the DNA concentration and 260/280 ratio – a measure of the purity of DNA. It requires only 0.5µl of sample, so is minimally destructive of the DNA resource. It has a wide range of accuracy for DNA concentration (2-15000 ng/µL).

2.3.5.4.2 Gel electrophoresis

DNA gel electrophoresis permits estimation of DNA quality in terms of fragmentation. Highly fragmented samples are less likely to be successfully genotyped. I subjected a subset of samples to DNA gel electrophoresis. These were selected randomly from amongst the samples within the lowest quartile of concentration and quality on spectrophotometry.

2.3.5.5 Presentation of DNA for genotyping

Using DNA concentration data obtained in 2.3.5.2, I placed samples into 96 well plates with a total volume of 135µl solution per sample at a concentration of 7ng/µl. The required concentration was achieved by diluting an appropriate volume of each DNA sample with a matched volume of Buffer AE (QIAGEN).

2.3.6 SNP selection

I selected the polymorphisms to be genotyped in the 5 candidate genes. I used HaploView (v4.1, Broad Institute, Boston MA, USA) to search HapMap data (v2, release 22, NCBI) from genomic regions of interest within, and 5KB up and down stream of, each candidate gene. Markers with a minor allele frequency (MAF) of less than 5% were excluded. I also excluded markers that were not in Hardy-Weinberg

equilibrium in the HapMap cohort, with a p-value threshold of 0.001. I then used the application tagger within HaploView to select SNPs that captured the maximum genetic variation around each locus with the minimum number of SNPs. Pair-wise comparisons only were used with an R^2 cut-off of ≥0.8. R^2 is an index of linkage disequilibrium (LD) between two SNPs. This resulted in 23 SNPs to be genotyped in *ABCB11*, 15 SNPs in *ABCB4*, 9 in *ABCC2*, 21 in *ATP8B1* and 5 in *NR1H4*. These SNPs are listed in Table 2.3.

I performed primer design by collating the corresponding DNA sequence from the NCBI dbSNP database for each SNP shortlisted. These are listed in Table 2.4. I reverse checked all DNA primer sequences by searching the NCBI basic local alignment search tool (BLAST). I then input these sequences into 'PrimerPicker' (KBioscience). This is proprietary, web-based software that generates designs for the oligonucleotides necessary for the KASPar genotyping process.

2.3.7 KASPar genotyping

KBioscience undertook primer preparation and genotyping. They used a competitive allele-specific PCR SNP genotyping system known as KASPar and which utilises FRET quencher cassette oligonucleotides. The KASPar system has demonstrated superior reliability to Taqman systems and is substantially cheaper.[317] KASPar is a homogeneous fluorescent endpoint genotyping system. It comprises two components - an assay mix and a reaction mix. The assay mix is the SNP-specific component, containing three unlabelled primers. The reaction mix contains a universal fluorescent reporting system (fluorescent dyes FAM, VIC and ROX as a control) and Taq polymerase. The two mixes were applied to each DNA sample and 20 PCR thermo cycles undertaken. A fluorescent plate reader was then used to measure signal intensity from each well. Data were plotted in KlusterCaller software and the genotype determined by the relative signal intensity from each fluorescent marker.

Raw data from the genotyping process was returned to me in comma separated values (csv) format.

2.3.8 Analysis

I managed and manipulated the raw genotyping data with Excel (Microsoft, Redmond, USA). Differences between groups were considered significant when $p<0.05$.

2.3.8.1 Hardy-Weinberg equilibrium (HWE)

I confirmed HWE in all genotyped SNPs using Pearson's χ^2 test in PLINK (V1.07, MAC OS 10.5.8.). I used a p-value threshold of 0.001, in line with standard practice and the HWE p-value criteria set in the tagger algorithm during SNP selection. I determined that any SNPs that breached this HWE threshold in the control cohort would be excluded from further analysis.

2.3.8.1.1 HWE testing in PLINK

PLINK is a programme run from the command line in Unix. I taught myself how to use it using online resources and trial and error. It was necessary to reformat my genotyping data into the prescribed matrices. The first of these is the PLINK-DATA file, containing only the subject labels and allele calls in chromosomal order. And example of the PLINK-DATA file, for *ABCB11*, is reproduced in Figure 2.2. The second is a PLINK-MAP file, listing the rs number of each SNP in chromosomal order. An example of the PLINK-MAP file, for *ABCB11*, is given in Figure 2.3. Finally, I had to write a short command line script to run the programme in each gene. An example script, for the confirmation of HWE in *ABCB11*, is reproduced here:

```
>PLINK --PED PLINKABCB11DATACCAFRICANANDPSCEXCLUSION.TXT --1 --NO-FID --NO-PARENTS --NO-SEX --ALLOW-NO-SEX --MAP PLINKABCB11MAPCCAFRICANANDPSCEXCLUSION.TXT --HARDY
```

2.3.8.2 Allelic association testing (χ^2)

I used a χ^2 test to perform a basic comparison of allelic frequencies between case and control groups using PLINK. Estimated odds ratios (with 95% confidence intervals)

were also calculated at this stage. The same PLINK data and MAP matrices were used as for HWE testing, above. I wrote the necessary PLINK scripts. As an example, the script for *ABCB11* is reproduced here:

```
>PLINK --PED PLINKABCB11DATACCAFRICANANDPSCEXCLUSION.TXT --1 --NO-FID --NO-PARENTS --NO-SEX --ALLOW-NO-SEX --MAP PLINKABCB11MAPCCAFRICANANDPSCEXCLUSION.TXT --ASSOC --CI 0.95
```

2.3.8.3 Cochran-Armitage trend testing (CATT)

I then compared observed versus expected genotype frequencies with Cochran-Armitage trend testing (CATT) using PLINK. CATT is a modified form of χ^2 test, which examines the relationship between variable 1, with two categories, and variable 2, with *n* categories where the *n* categories represent what is thought to be a continuum – e.g. 'high', 'medium' and 'low' doses of a drug in a clinical trial. It is of utility in genetic studies as it permits comparison of aa, aA and AA genotypes. Technically, CATT remains an appropriate test even when HWE criteria are not met.

I was able to use the same PLINK-DATA and –MAP matrices that I produced for HWE and χ^2 testing, above. I wrote the necessary PLINK scripts for each gene. As an example, the script for *ABCB11* is reproduced here:

```
> PLINK --PED PLINKABCB11DATACCAFRICANANDPSCEXCLUSION.TXT --1 --NO-FID --NO-PARENTS --NO-SEX --ALLOW-NO-SEX --MAP PLINKABCB11MAPCCAFRICANANDPSCEXCLUSION.TXT --MODEL
```

2.3.8.4 Haplotype analyses

In complex phenotypic traits, combinations of SNPs may offer a more sensitive marker of modified risk. Each SNP cannot be fully studied in isolation as each is in linkage disequilibrium (LD) with the others. Each may represent a number of genetic polymorphisms. Haplotypes are combinations of different marker alleles on a single chromosome. A combination of SNPs, or the genetic polymorphisms that they represent, may be a more powerful indicator of risk. Haplotype analyses are required to

investigate this. However, in a diploid, human population individual genotyping results in unrelated subjects make it impossible to define an individual's haplotype (unless that individual is homozygous in all the SNPs genotyped). Statistical tools now permit the analysis of indirectly measured, or inferred, haplotypes. These methods assume that all subjects are unrelated and that the haplotypes are ambiguous due to unknown linkage phase of the SNPs. In such applications, the haplotypes are inferred using the assumptions implicit in HWE and application of an expectation-maximization (EM) algorithm. This uses a baseline set of haplotype frequency estimates to calculate a conditional distribution of haplotype pairs that each subject carries. This is known as the estimation step. Based on these conditional distributions, haplotype frequency estimates are updated - the maximization step. The EM algorithm switches between these two steps until the haplotype frequency estimates converge. The Haplo.stats software is an application of this type of EM algorithm and runs within the R statistical environment.

2.3.8.4.1 Haplo.stats

I undertook Haplotype analyses on the set of SNPs genotyped in each gene using Haplo.stats (V1.4.0 in R V2.10.1,).[318, 319]

I reformatted the genotyping data into the required data matrices. One of these was required for each gene. The allelic calls for each SNP were listed in columns, in the order of the SNPs on the chromosome. Each row represented an individual subject. I made these matrices in Excel in a tab separated values format. As an example, part of the *ABCB11* data matrix is presented in Figure 2.4.

The Haplo.stats package (V1.4.4, Mayo Clinic, MA, USA)[320], containing the EM algorithm, was downloaded from the comprehensive R archive network (http://cran.ma.imperial.ac.uk/). I then wrote the required code to perform the analysis within R. I taught myself how to do this using online resources and trial and error. As an example, the code script that I wrote for *ABCB11* is shown in Figure 2.5.

2.3.8.5 Correction for multiple testing

There is an extensive literature on the most appropriate technique for correction for multiple testing in SNP association studies. Bonferroni correction is the most well-known, and is a highly conservative technique. Permutation testing is a well-accepted, less conservative, approach. False discovery rate calculations are less conservative still, and therefore least likely to incorrectly discard a true positive result due to over correction. To explore the significance of my initial findings, I applied each of these techniques in turn.

2.3.8.5.1 Bonferroni correction

I first applied Bonferroni correction to the p-values obtained, using Excel (Microsoft). As 73 SNPs had been genotyped, each nominal p-value was multiplied by a factor of 73 to generate the corrected value. As only 5 haplotype analyses were undertaken, their resulting p-values were multiplied by a factor of 5.

2.3.8.5.2 Westfall and Young permutation testing

This technique initially applies a χ^2 test to the case and control groups. A test statistic and p-value are thereby generated for each SNP. It then randomly reallocates each genotyped subject as a case or control and applies a χ^2 test to test between those allocated case and control status. This is reiterated many thousands of times, generating pseudo-p values and test statistics on the randomly generated cohorts. Assuming the null hypothesis is correct, all such iterations should have the same probability of showing a significant difference between the two groups as a matter of chance. The proportion of resampled data sets where the minimum pseudo-p-value is less than the original p-value is the adjusted p-value.

I performed permutation testing using PLINK (V1.07). I did not perform permutation testing on the haplotype results.

2.3.8.5.2.1 Permutation testing in PLINK

I had to combine the PLINK-DATA and PLINK-MAP matrices from all 5 genes for this analysis, and write further command line scripts. I set a maximum of 10000 permutations per SNP. The PLINK script is reproduced here:

```
>PLINK --PED PLINKABCB11DATACCAFRICANANDPSCEXCLUSION.TXT --1 --NO-FID --NO-PARENTS --NO-SEX --ALLOW-NO-SEX --MAP PLINKABCB11MAPCCAFRICANANDPSCEXCLUSION.TXT --TDT --MPERM 10000
```

2.3.8.5.3 False discovery rate calculation (FDR)

As a less conservative correction for multiple testing, a predicted FDR, which can be interpreted as the estimated proportion of false discoveries at a given p-value threshold, was calculated (R V2.10.1, QVALUE V1.0) with lambdalamda range set from 0.0 to 0.9 by 0.05, with both smoother and robust methods applied.[321] FDR calculations were not performed on the haplotype results, as it is not an appropriate application of this method.

2.3.8.6 HapMap and NCBI dbSNP interrogation for associated SNPs

If any true association exists between an individual SNP and risk of CC, this is highly unlikely to reflect a direct effect of that single polymorphism itself. Instead, the SNP is likely to represent a surrogate for changes in areas of the gene with which it shares close linkage disequilibrium. To explore this, I used HapMap to indentify all other, non-genotyped, SNPs known to lie in strong LD (R^2>0.8) with any genotyped SNPs found to be associated with altered susceptibility to CC. Each of the genotyped and associated individual SNPs, and those found to be in strong LD, were then examined in dbSNP[322] (NCBI, Bethesda, USA, Build 132) to explore potential functional consequences.

2.4 Results

2.4.1 DNA quality

Blood samples were collected from a total of 172 subjects with CC and 256 control subjects. 44 samples were collected prospectively. 128 were obtained from archives.

As expected, the QIAGEN QIAamp blood mini DNA extraction technique was time efficient and effective. Although time consuming, the extraction protocol used on the clotted archive samples proved safe and effective.

Germline leucocyte DNA was successfully extracted from all samples. Low volume spectrophotometry (LVS) confirmed high yields from most samples, and adequate yields from the rest. LVS also confirmed high DNA purity in the vast majority of extracted samples. DNA gel electrophoresis of the 15 least concentrated, least pure samples demonstrated reasonable quantities of large fragment DNA, certainly amenable to genotyping. 73 SNPs were successfully genotyped in samples from all 172 subjects with CC. Of the 256 control samples, 2 failed genotyping resulting in 254 to be included in the analysis.

There were no differences in DNA quality between samples extracted with the QIAGEN™ blood mini kit and those extracted with the QIAGEN™ clotspin and PureGene system. There was no difference in genotyping success between the archived control samples, archived CC samples and the prospectively collected CC samples.

2.4.2 Demographics of case and control groups

8 of 172 CC cases were excluded after genotyping, but before analysis, as the cases had been identified as non-Caucasian (4), as PSC related CC (3) or both (1). This resulted in 164 cases to be included in the analysis. All cases included in the analysis were from Caucasian patients without PSC. The groups were well matched in terms of sex and age (Table 2.5).

2.4.3 Hardy-Weinberg equilibrium (HWE)

HWE was confirmed in all genotyped SNPs in case, control and combined groups. HWE results from the control group, for each SNP genotyped are presented in Table 2.6. As none of these SNPs breached the defined p-value threshold, all genotyped SNPs were included in subsequent analyses.

2.4.4 Association testing

The results of *all* association tests on *all* SNPs are reported in Table 2.7 in Appendix 2. For clarity, the results of allelic χ^2 testing and CATT that yielded statistically significant associations are summarised in Table 2.8.

2.4.4.1 Allelic frequency χ^2 testing

Allelic frequency testing demonstrated an association between 8 SNPs and altered risk of CC. None of the other 64 SNPs reached the p-value threshold of <0.05.

2.4.4.2 Cochran-Armitage trend testing (CATT)

CATT testing revealed a nominal association between genotype status and risk of CC in 7 SNPs and these are summarised in Table 2.8.

2.4.5 Haplotype testing

Haplotype testing revealed an association between haplotypes formed in *ATP8B11* and risk of CC. Haplotype testing in the other 4 candidate genes did not show any relationship with risk of CC. The haplotype findings are summarised in Table 2.9.

2.4.6 Correction for multiple testing

Although 8 of the individual SNPs and 1 of the haplotype analyses were nominally associated with altered risk of CC, it was necessary to correct the p-values for multiple testing.

2.4.6.1 Bonferroni correction

After Bonferroni correction, none of the initially associated SNPs retained significance. The corrected p-values are summarised in Table 2.8. The haplotype finding in *ATP8B1* did retain significance. Bonferroni corrected haplotype p-values are summarised in Table 2.9.

2.4.6.2 Permutation testing

Permutation testing generated p-values consistent with those found with allelic frequency χ^2 testing, with 8 nominally associated SNPs. However, after permutation

correction for multiple testing, these p-values were >0.05 and none of the individual SNPs retained significant associations (data not presented).

2.4.6.3 False discovery rate (FDR)

The false discovery rate (FDR) q-value was calculated as 0.163, predicting <1.4 false positives amongst the 8 nominally significant SNPs and suggesting that the majority of these results are true positives. The respective q-value for each p-value is summarised in Table 2.8.

2.4.7 HapMap interrogation for associated SNPs

I searched for each of the 8 nominally associated SNPs in HapMap and identified 19 other SNPs that lay in close LD (R^2>0.8). Using dbSNP, I was able to identify the functionality of each of these 27 (8 plus 19) SNPs within each gene. Only one SNP (rs2287622) was directly associated with a known functional genetic polymorphism. All of the other 26 SNPs were intronic. These SNPs, and their function, are summarised in Table 2.11, below.

2.4.8 Summary of results by gene

2.4.8.1 ABCB11

Allelic frequency testing in *ABCB11* demonstrated that four individual SNPs were associated with altered susceptibility to CC (p<0.05); rs3770585G>A (OR 0.69, 95% CI 0.52-0.93, p=0.013), rs2287622C>T (1.478, 95% CI 1.11-1.97, p=0.007), rs3770596T>A (OR 1.35, 95% CI 1.02-1.80, p=0.038), rs7605199G>A (OR 1.336, 95% CI 1.01-1.77, p=0.044). CATT also identified an association in these four SNPs, although with slightly higher p-values of 0.013, 0.009, 0.043 and 0.045 respectively.

On searching dBSNP, One of these four associated SNPs - rs2287622 - was found to be directly associated with a genetic polymorphism (V444A or c.1331T>C).

Using HapMap, 15 other SNPs within the *ABCB11* gene were found that are known to lie in strong LD (R^2>0.8) to the four associated SNPs. All of these linked SNPs were reviewed in dBSNP. All are intronic and are not associated with any known functional

genetic polymorphism. Haplotype analysis of all 23 SNPs genotyped in *ABCB11* demonstrated no haplotype associations.

Correction for multiple testing with both Bonferroni and permutation approaches abrogated the statistical significance of all significant findings in *ABCB11*. However, the false discovery rate calculation estimates that <1.4 of the 8 nominally associated SNPs are true false positives, meaning that the majority (>6.6) are likely to be genuine findings.

2.4.8.2 ABCB4

Allelic frequency testing in *ABCB4* identified one SNP that was associated with altered susceptibility to CC - rs2097937A>G (OR 1.46, 95% CI 1.01-2.07, p=0.044). CATT of genotype frequencies also indicated a significant difference in this SNP, but with a slightly higher p-value (0.048). Rs2097937 is intronic. Four other SNPs within *ABCB4* were found to lie in strong LD (R2>0.8) to the associated SNP. Again, all of these four SNPs are intronic. Haplotype analysis of all 15 SNPs genotyped in *ABCB4* demonstrated no haplotype associations. Correction for multiple testing with both Bonferroni and permutation approaches abrogated the statistical significance of the single significant finding in *ABCB4*. The false discovery rate calculation estimates that <1.4 of the eight nominally associated SNPs are true false positives, meaning that the majority (>6.6) are likely to be genuine findings.

2.4.8.3 ABCC2

None of the individual SNPs genotyped in *ABCC2* showed significant association with altered susceptibility to CC. The SNP rs3740066 was not associated with increased risk of CC. In fact, a trend towards association with *decreased* risk of CC was noted (OR 0.7854 95% CI 0.5852-1.054, p=0.09). Haplotype analysis of all nine SNPs genotyped identified no haplotypes to be associated with altered CC risk.

2.4.8.4 ATP8B11

Allelic frequency testing identified three SNPs in *ATP8B1* to be significantly associated with altered susceptibility to CC; rs319454G>A (OR 1.37, 95% CI 1.01-1.86, p=0.04)

rs319448A>G (OR 1.375 95% CI 1.04-1.83, p=0.028), rs12456346A>G (OR 1.334, 95% CI 1.002-1.776, p= 0.048). CATT of genotype frequencies also indicated a significant difference in two of these SNPs, with slightly higher p-values (0.042, 0.032), but did not reach the threshold for significance in the third (rs124563436), with a p-value of 0.055.

One other SNP within the gene was found to lie in strong LD (R^2 >0.8) to the associated SNPs. The 3 associated SNPs are intronic, as is the one SNP that lies in close LD.

Haplotype analysis of the 20 SNPs genotyped in *ATP8B1* demonstrated a significant difference between case and control groups (global-stat score of 28, df 13, p = 0.0093) All 13 haplotypes contributing to the global-stat score finding are summarised in Table 2.10. The haplotype effect was largely driven by a single protective haplotype, reference number 76, identified in 10% of healthy controls and 3% of CC cases (Hap score -3.22, p = 0.0013, OR 0.49 (95% CI 0.22 to 1.1), and a number of other haplotypes associated with decreased and increased risk of CC. Haplotype 76 included the wild-type, lower risk allele at all three loci identified by allelic association studies (rs319454, rs319448 and rs12456346), as did the base (i.e. OR 1) haplotype.

Correction for multiple testing with both Bonferroni and permutation approaches abrogated the statistical significance of the 3 SNP findings in *ATP8B1*. However, the false discovery rate calculation estimates that <1.4 of the 8 nominally associated SNPs are true false positives, meaning that the majority (>6.6) are likely to be genuine findings. The statistical significance of the haplotype finding in *ATP8B1* remained significant, even after Bonferroni correction (p=0.47).

2.4.8.5 NR1H4

None of the five individual SNPs genotyped in showed significant association with altered susceptibility to CC. Haplotype analysis identified no relationship between haplotypes of the 5 genotyped loci in *NR1H4* and CC risk.

2.5 Discussion

I have investigated the relationship of polymorphisms in genes encoding four major biliary transporter proteins and a nuclear receptor, FXR, to risk of CC. This is the first study to examine the role of variation in *ABCB11* (BSEP), *ABCB4* (MDR3), *ATP8B1* (FIC1) and *NR1H4* (FXR). I have also investigated *ABCC2* (MRP2), which has been examined before in a study of just 60 patients and 73 controls.[71]

The case and control cohorts that I assembled were well matched. All subjects were Caucasian and did not have PSC. This minimised confounding factors and maximised the chance that case and control groups were directly comparable. However, this also means that any conclusions that may be drawn are not directly generalisable to non-Caucasian populations, or to PSC patients - who are high risk of CC. Ethnic variation of genome and geographic risk factors for CC mean that entirely different gene polymorphisms may modify risk in other parts of the world. If validated, it would be particularly important to examine the polymorphisms that I have identified in a PSC related CC cohort, and an Asian cohort of controls and patients with *OV* associated CC.

Although very well characterised, the CC cohort in my study did have some inevitable variation in phenotype that, due to subgroup sizes, it is not possible to investigate meaningfully. Such variables include intra- or extra-hepatic CC, operability, response to chemotherapy, rate of spread and survival.

A crucial obstacle to overcome was successful extraction of DNA from archived clot from CC patients. These samples represented 74% of my CC case cohort, and so failure to extract DNA would have rendered my study lamentably underpowered. My lab had not extracted DNA from clot before, and I was not permitted to use blade or ultrasonic techniques due to safety concerns. I expended some time identifying and learning the right DNA extraction method for the archived CC plasma samples. This was worthwhile, as the Clotspin basket method of clot disruption proved safe and highly effective – subsequent DNA yields were massive.

I did not undertake the extraction of the control DNA samples, or the KASPar genotyping, myself. Modern systems are almost fully robotised, and are highly efficient in terms of cost and time. Manually extracting the control samples would have occupied a disproportionate amount of my time. If I had genotyped the 73 SNPs in this number of samples it would have proven costly and time consuming. It is unlikely that I would have achieved as high a genotype call rate as a commercial robotic system.

The statistical analyses that I undertook required the use of several software platforms with which I was unfamiliar. All of these were command line based packages, which need carefully constructed input matrices and programming scripts to run. I have no programming experience, nor did my colleagues. I therefore taught myself how to use them using online resources and trial and error with specimen datasets. I was able to reverse check my results, and they were also reviewed by a genetic statistician who concurred with my technique and results. I am therefore confident that my statistical analysis has been robust.

Eight of the SNPs genotyped are nominally associated with an altered risk of CC. I will consider the implications of these findings, gene by gene, in the subsequent sections. The approach used in this study, with multiple SNPs tested in each gene, means that p-value thresholds for significance must be corrected for multiple testing. None of the nominal SNP associations survived Bonferroni or permutation methods of correction for multiple testing. Both Bonferroni and permutation testing are expected to be rather conservative corrections, and I did observe a considerable excess of nominally significant results compared to the approximately 2 expected under the null. With this less stringent method of correction for multiple testing (FDR), individual SNPs retain significance, indicating a potential association. However, FDR analysis does not identify exactly which of the 8 findings are likely to 'false discoveries'. However, the solitary finding of just 1 SNP in *ABCB4* remains particularly tenuous. Interestingly, the only positive haplotype finding *was* sustained – even after Bonferroni correction. This is discussed in detail, along with the other findings in *ATP8B1*, below.

Apart from rs2287622, a non-synonymous polymorphism associated with a missense mutation in BSEP, all of the associated SNPs are intronic, as are all other SNPs that I identified as being in strong LD (R^2>0.8) to them.

As any gene mutation associated with any of these 8 SNPs cannot be deduced from my data, it is not possible to determine what mutation any influence is due to. Counter intuitively, the 'wild-type' allele in some SNPs is actually less common in the background population than the 'mutant' allele - most notably in rs2287622. To complicate interpretation of risk further, a differential in risk might be due to a protective effect associated with the 'wild-type' allele or a deleterious effect associated with the 'mutant' allele, or even *vice versa*.

It is important to note that the MAF of the SNPs selected for genotyping was intentionally set at a high level. This was necessary, as my relatively small cohort sizes could not be used to identify or exclude an influence of rare polymorphisms. This means that, in genes where I have detected no signal, it is highly unlikely that any *common* variation is associated with CC. In a cohort of this size, and with such a complex trait, testing rare polymorphisms would be futile. Therefore, it is not possible for me to exclude a powerful relationship between rare SNPs in the genes studied, and risk of CC.

2.5.1 BSEP

I identified four SNPs of interest in *ABCB11*, although I demonstrated no haplotypes that modified risk. Of the four, the strongest association was found in rs2287622, which is known to be associated with the V444A polymorphism (c.1331T>C, p <0.007, cor p 0.52). This polymorphism has been positively associated with increased risk of a number of diseases, most compellingly with intrahepatic cholestasis of pregnancy (ICP) and drug-induced cholestasis.[323, 324] However, in my CC cohort, c.1331T>C was found to be associated with *reduced* susceptibility to disease. In ICP and drug-induced cholestasis, c.1331T>C is thought to contribute to reduced function of BSEP compared

to wildtype, reduced elimination of bile acids and therefore intrahepatic cholestasis and jaundice.

It is tempting to conclude that the altered risk associated with this SNP must be related to the V444A mutation, but we cannot do so firmly from these data. Rs2287622 may be in LD with other, currently undefined, mutations in *ABCB11*. To confirm a link with V444A we would have to sequence the relevant part of the gene in the CC cohort and controls and directly demonstrate an association.

It is plausible that reduced BSEP expression or function may be implicated in CC pathogenesis. BSEP is expressed on the apical membrane of hepatocytes and its principal function is the export of bile acids into the canaliculus. It is the major exporter of bile acids and reduction of its expression leads to increased levels of bile acids within the hepatocyte and cholangiocyte and reduced levels of bile acids in the canaliculi, with reduced overall bile flow. Very early onset of CC and HCC has been described in children with genetic disorders of BSEP. The risk of hepatobiliary malignancy was found to be highest as phenotypic expression of BSEP became minimal or absent.[50, 325]

A further possible explanation lies in the 'toxic bile theory'. This has been proposed by a number of authors and supported by studies of biliary transporter gene knockout mice.[299] Under normal conditions bile is already a considerably toxic fluid; toxicity to hepatocytes and bile duct epithelial cells is prevented by a number of physiological mechanisms, including micellar binding of bile acids and exogenous toxins. Bile acids form mixed biliary micelles with PtC and cholesterol, reducing the detergent activity of monomeric bile acids and activity of other toxins in the bile duct lumen. This prevents toxicity of high biliary bile acid and xenobiotic concentrations to cholangiocytes.[326, 327] Disturbances of bile composition may result in failure of this protective micellar system and subsequent hepatocellular and bile duct injury. Such changes may be due to the presence of exogenous toxins – or changes in the physiological constituents of bile.

Any change that leads to increased or *decreased* ratios of bile acid to PtC leads to reduced micellar stability.[294] Changes in this ratio and subsequent biliary damage have been shown in hereditary MDR3/Mdr2 defects, imbalance in BSEP/MDR3 expression following liver transplantation and impaired MDR3 function due to absence of enteral nutrition, or use of total parenteral nutrition.[328, 329] This body of evidence suggests that biliary transporter polymorphisms (or combination of polymorphisms) that lead to disequilibrium of the bile acid to PtC ratio lead to unstable micelles - and bile duct injury.

2.5.2 ABCB4

I identified 1 SNP in *ABCB4* that was associated with an increased risk of CC. Given that only a single SNP was identified, and that even the least conservative correction for multiple testing tends to abrogate its significance, this finding must be interpreted very cautiously. It may well represent a false positive. However, it is worth briefly considering how changes in MDR3 might impact risk of CC. A principal function of MDR3 is excretion of PtC into bile. Given the importance of PtC in stable micelle formation, discussed in 2.5.1, it is unsurprising that MDR3 dysfunction can lead to chronic bile duct injury. This has been demonstrated in progressive familial intrahepatic cholestasis 3 (PFIC3) where *ABCB4* mutation confers absence of canalicular MDR3 protein.[48, 301] These patients develop biliary cirrhosis and cholelithiasis syndromes, both of which are risk factors for CC.

2.5.3 ABCC2

No individual SNPs or haplotypes in *ABCC2* were associated with change in risk of CC. An association between the SNP rs3740066 (c.3972 C>T) in *ABCC2* and CC was previously reported by Hoblinger *et al* in a study of 60 CC cases and 72 controls with an OR of 1.83 (95% CI = 1.09-3.08, p=0.022).[71] Our, larger, study has not confirmed this finding. In fact, a non-significant trend in the opposite direction was detected with an OR of 0.79 (95% CI = 0.59-1.05, p=0.09). With the crucial role of MRP2 in toxin

excretion, it was a highly plausible candidate. However, given my comprehensive coverage of this locus, it is highly unlikely that common variation in this gene is implicated in CC risk.

2.5.4 ATP8B1

Haplotype analysis in *ATP8B1* demonstrated a significant association between specific haplotypes and altered susceptibility to CC. This finding was sustained after Bonferroni correction for multiple testing (corrected p=0.047). The function of ATP8B1, and its dysfunction in patients with PFIC 1 or BRIC 1, is yet to be fully elucidated. A critical role in trafficking phosphatidylserine from the outer to the inner membrane leaflet has been demonstrated. Some authors have hypothesised mechanisms by which ATP8B1 dysfunction may lead to cholestasis. Reduced ATP8B1 function causes disordered phosphatidylserine distribution, leading to membrane instability and vulnerability to lipid extraction by hydrophobic bile acids. Marked depletion of membrane lipids abrogates the function of abcb11 and abcc2 (and potentially other membrane bound transporter proteins) in mouse models, thereby leading to cholestasis, the hallmark of PFIC 1 and BRIC 1.[330] Other authors have proposed a further function of ATP8B1 in the formation of the cellular cytoskeleton in polarised apical cells, where a reduction in ATP8B1 function results in the reduction of microvilli formation in Caco-2 cells. It is proposed that the resultant reduction of the apical surface area of hepatocytes limits the bile salt excretory capacity of the liver in patients with ATP8B1 deficiency.[331] A variety of potential roles in CC risk modification may also be postulated. ATP8B1 is expressed on both hepatocyte and cholangiocyte apical membranes. In the cholangiocyte, decreased membrane stability and increased vulnerability to lipid extraction may increase apoptosis and cell turnover. Abnormality in apical membrane structure may impact the function of other cholangiocyte transporter proteins, leading to reduced efflux, or increased influx, of mutagenic species into the cholangiocyte cytoplasm. In the hepatocyte, reduced ATP8B1 function may result in depressed function of other transporter proteins. The resultant reduction in the principal physiological constituents

of bile may result in reduced bile flow, increasing the concentration of exogenous mutagens in bile and the exposure of the cholangiocytes to such species. Furthermore, such direct and indirect effects might result in unstable micelle formation in the bile, impairing the protective mechanisms of the biliary tract and leading to increased exposure of the cholangiocytes to damaging toxins.

2.5.5 NR1H4

No relationship between NR1H4 polymorphisms and CC was demonstrated. It is highly unlikely that common variation in this gene plays a strong role in risk of sporadic CC.

2.5.6 Conclusion

This is the largest study to date of biliary transporter proteins as susceptibility factors for CC. This study has demonstrated an association between polymorphisms in 3 genes encoding biliary transporter proteins and altered susceptibility to CC. A total of 8 SNPs have been associated with increased or decreased risk of CC. These findings might be accounted for by the over or under expression of BSEP, FIC1 or MDR3 leading to a variety of possible biliary abnormalities.

The functional consequences of these polymorphisms in terms of protein expression or function are not known and cannot be deduced from these data. Given the findings in *ABCB11* and *ATP8B1*, and their biological plausibility as risk modifiers for CC, further study of these genes in a validation cohort is warranted.

A validation cohort would be highly desirable. The genotyping of the 8 candidate SNPs in a new CC cohort would add power to my findings, or refute them. If replicated, deep re-sequencing of the implicated regions of the genome in affected subjects would be indicated. As genetic susceptibility to CC is likely to be highly complex and involve many genes, a genome wide association (GWAS) approach would offer substantial advantages. However, CC is a relatively rare disease and such a study will require a multi-centre, international collaboration. Subsequent to my work in this area, I have visited the cholangiocarcinoma group at the Mayo Clinic, Minnesota, USA. Plans are

well advanced for them to undertake validation of my biliary transporter findings in their own CC cohort. This will be as a prelude to an ongoing collaborative GWAS study into CC.

A range of biliary transporter polymorphisms that cause or modify diseases has already been discovered in progressive familial cholestasis, gallstone disease and cystic fibrosis. Discovery of genetic polymorphisms that modify the risk of CC would offer a major insight into its pathogenesis. If a panel of genetic polymorphisms that modifies risk was confirmed and validated in a PSC or *OV* cohort, risk stratification for development of CC might be possible. Low risk patients could be reassured. *OV* patients considered to be at high risk could be subject to concerted efforts to eradicate their infestation and monitor for *OV* relapse. In very high-risk groups, intensive monitoring and even prophylactic liver transplantation might be feasible before CC develops.

2.6 Chapter 2 Tables and Figures

Table 2.1: Summary of biliary transporters, their pseudonyms, encoding genes, known functions, distribution in the liver and associated diseases

Legend: PFIC, progressive familial intrahepatic cholestasis; BRIC – benign recurrent intrahepatic cholestasis

Acronym	Trivial name	Full name	Gene abbr.	Known function	Distribution in liver	Associated diseases
BSEP	Bile salt exporter pump	ATP-binding cassette, sub-family B (MDR/TAP), member 11	*ABCB11*	Bile acid transport	Hepatocyte apical membrane	PFIC2, HCC, BRIC-2
MDR3	Multidrug resistance protein 3	ATP-binding cassette, sub-family B (MDR/TAP), member 4	*ABCB4*	PtC transport	Hepatocyte apical membrane	PFIC3, Gallstone disease
MRP2	Multidrug resistance-associated protein 2	ATP-binding cassette, sub-family C (CFTR/MRP), member 2	*ABCC2*	Conjugated bilirubin and organic anion transport	Hepatocyte apical membrane	Dubin-Johnson
FIC1	Familial intrahepatic cholestasis protein 1	ATPase, aminophospholipid transporter, class I, type 8B, member 1	*ATP8B1*	Translocation of acidic phospholipids in cell membrane	Hepatocyte & cholangiocyte apical membranes	PFIC-1 /Byler's disease, BRIC-1
FXR	Farnesoid X receptor/bile acid receptor	Nuclear receptor subfamily 1, group H, member 4	*NR1H4*	Transcriptional regulation of ABCB4 and ABCB11	Hepatocytes and cholangiocytes	Cholestasis, Gallstone disease

Table 2.2: Commonly used values for $c_{p,power}$

Legend: Adapted from Machin, 1987[313]

Commonly used values for $c_{p,power}$

	Power (%)			
P	50	80	90	95
0.05	3.8	7.9	10.5	13.0
0.01	6.6	11.7	14.9	17.8

Table 2.3: SNPs selected for genotyping in each candidate gene

Legend: SNP - single nucleotide polymorphism; Ref - study SNP reference number; RS - reference SNP; BP - base pair

ABCB11 **- Chromosome 2**		
Ref	RS number	BP location
1	rs497692	169497262
2	rs3755157	169500417
3	rs6709087	169507256
4	rs853773	169522593
5	rs853772	169522901
6	rs3821120	169525182
7	rs16823014	169525959
8	rs4148797	169526281
9	rs16856300	169526548
10	rs4148794	169529894
11	rs17267869	169531654
12	rs3770585	169532113
13	rs2287622	169538574
14	rs2058996	169542195
15	rs3770596	169548388
16	rs13416802	169550584
17	rs2287618	169551055
18	rs7605199	169564700
19	rs3815676	169578625
20	rs4148773	169582067
21	rs3814382	169597234
22	rs10930343	169598031
23	rs7577650	169599456

ABCB4 **- Chromosome 7**		
Ref	RS number	BP location
1	rs31652	86867623
2	rs2097937	86868839
3	rs31653	86870549
4	rs2373593	86874878
5	rs31666	86890204
6	rs31668	86891142
7	rs8187799	86894112
8	rs31676	86907816
9	rs1149222	86911711
10	rs4148826	86912355
11	rs2109505	86917342
12	ABCB4_indel	86919500
13	rs1202283	86920228
14	rs2302386	86929880
15	rs4148812	86939343

ABCC2 **- Chromosome 10**		
Ref	RS number	BP location
1	rs717620	101532568
2	rs2756109	101548736
3	rs4148391	101556856
4	rs2073337	101557416
5	rs2002042	101577921
6	rs7476245	101584719
7	rs3740066	101594197
8	rs3740065	101595683
9	rs3740063	101600713

NR1H4 **- Chromosome 12**		
Ref	RS number	BP location
1	rs4764980	99409238
2	rs56163822	99411232
3	rs61755050	99450439
4	rs1030454	99469383
5	rs35724	99479509

ATP8B1 **- Chromosome 18**		
Ref	RS number	BP location
1	rs7241054	53464481
2	rs1968274	53465209
3	rs17685876	53468119
4	rs317826	53474943
5	rs4308033	53485104
6	rs4306606	53485293
7	rs317845	53486410
8	rs317838	53489654
9	rs317837	53490222
10	rs11659313	53503387
11	rs319454	53504713
12	rs7236365	53509688
13	rs160993	53510410
14	rs319448	53511146
15	rs319440	53515528
16	rs17686300	53519745
17	rs319457	53519863
18	rs319406	53527210
19	rs319409	53529582
20	rs12456346	53546809
21	rs9676158	53555113

Table 2.4: SNP DNA sequences for oligonucleotide primer design

Legend: SNP - single nucleotide polymorphism, [x/z] - variable allele of SNP in preserved DNA sequence

Gene	SNP	Sequence
ABCB11	rs497692	TATTTGGTCCTTTCCTGGCAGAACCAGGGCTAGATCCCCAAACTTCTGCCTTCAGGTCATCACACCAACCACGCCACCCTGCTCTCTTCCTGTGTGTCTGTCTGGTTACAGGGTGATCTCTGCAGTTGTACTGAGTGC[A/G]ACAGCTCT TGGAAGAGCCTTCTCTTACACCCCAAGTTATGCAAAAGCTAAAATATCAGCTGCACGCTTTTTTCAACTGCTGGACCGACAACCCCCAATCAGTGTATACAATACTGCAGGT
	rs3755157	AATGCATCATCTGGGAATACAAAGACGTGGCTGGAGATCAACCCAACTGGACCACGAGACAATTTTTCAGTTAAATTCAAGTTTCCCTTAATTCTTATACTTGAGAAGTGTTTATGCTAC[C/T]ATCCCTCATTTACCCCCCAAGGACAAG ATATTAATTAGAACACACTTTAAAATGTGTAGCTAAAACCTAAGTGTATCCAAAACCTAAACTGTAGTTCATAATTCTGGTTGGGAAAGACAC
	rs6709087	AAATGTATTAAAATAATTAACATCCAGAAATAGATAAGAGAATTTCGAAGAGGAAAAATTAGTGTGCAGGGATTTGCCTTACCAGATACTAAAATGTATCAAGTTATAAGAATTAAAACA[A/G]AGGGTGCAAAAAAAAAAATAGAAAGA ATTAATAAGACATACTATTTGATAGCACAATAGAATGACTATAATAAATAATTTAATTGTACACTTAAAAATAACTAAGAGTGTAATTGGATTG
	rs853773	ACACAGCTGGACTAGCTTTCTTCTTCTCATACCTTTTATTTTATTATTTTGATAAGTCCGCTCTTTGTTTTCATGTATTTAGCTCTTTTCTGTTAGTTCATTTTTCTCATCTCTAAAGAA[C/T]GAAAAATTTCCTAAGCCTCCATCACACAGA GGGTAAGAAGGCAAAGCTGAGAACTCTCATTTGGGCTCGCAGGGCACTCAGACAGTCATTGTACCTATAAAAAAGAAAATGGTACTGTCA
	rs853772	CAACCATATCCCATAGACATTTGAGGTCACTCATAACTAATCAAACTATGACTCTTAAATCTGTGAATGCCAAAGGATCTGCTGCAAAATGTCTTGAGTACATTTAGATGATTAATATAA[A/C]CCTCTCTCTGCTCAAGGACAAGGACAT TCCTGTGCAGGAAGAAGTTGAACCTGCCCCAGTTAGGAGGATTCTGAAATTCAGTGCTCCAGAATGGCCCTACATGCTGGTAGGGTCTGTGGG
	rs3821120	GATGTGTCAGCTGTGCTGTACGTACCATCAAAGGGTAGGCTTTGGAACTTGTGAAGACTATTAGCACTCACTAAATTCTGCCAGCTTCCTTGTGACATTAAATACCAGCATTTTCTGGAAC[C/G]ATTGCCCCACAGGATGGTCCACCCA CAGCCTGACATGATAGAAATTTTCCCAACCTTCTGATACCTGAAGTTATCCTTAACTCTAAACAACAACACTTAGGACATTTCAAAGCGATACAA
	rs16823014	TACAATCGATGTTTATGGGATAAAAGAATGAATGAAACTTTTATGTCCATGCTGCATACATGCAACAAGAGATTTATTTAAAATTGCTCATCTACTATGATGCATTCATAGTTCAATGCTA[A/G]ACACTGGGCTTTTCTGTTTCCTTTCTT ACATCTTCTGTGTAGGCCAGAGGAAAAATAACCTCCGACTTAAATCTTAGGTCTTGCTTTCTATACTTTGTTACACTAATTTCTGCCTTCCAT
	rs4148797	ACAACATTAAGTTTAGGATTAGGGAAACCCACAGGCTAAAGGTTTTAGTCTCTAATATTTAGGCAATCTCTTTTGGTGAAACTAAAACACAAGGCAGAACATGTATGAGATCAAGTCAGG[A/G]TCAAATATTAGACACCCATAACTGCA TCACAGATAACTTATGTGATAGGATTTAGGCATTGGTTAAGATAAGACAGAGATGGGACCCATAAAAATGGCTATCAGTATTAGTTCATTTGAA
	rs16856300	GTTTACAGGAGCTAAGAGTACTGCTTGGAGAGCAAGGATAAATGGGTATAGTCTCATCTCTGTCACCAACTAACCATGAGACCTTGGGTGAAAGCACAGACTTCTCAGGAAATGGTTCTA[A/C]GTTTTAGAGTCAATACTCCACAAGA CCACAAGGTGGCGATAACTCCTTTGTGTGAATCGACCTTTAGCACCGTAAGAGAGTATTTCCCAATCTACAGAAAGTAAGGTTAGGCCTTTTCCA
	rs4148794	TGTACTCAGTATATTCAGCAAGCAACCCTGTCCTCCCCTGTCTCCATGCTTATGTTTATACATTCTCTCAGTCAGAAATGCTCTTTTCTATAGTCTACAAATCCCCATATCCTACCTTAG[C/T]CATCTCATCCATGAATCTTGCCTTGGTAA CACCTGCCATACATTTCTTTTTCCTTTGACTTTCCATACTCTTTATTCAAACCTGCCTAATAGCACTTACCACATTCTTCTTTACCTTAGG
	rs17267869	ACTTTCATACTCAAATGTACAAATCTTTCTCTGCAGGAGGTTAATAGCTACTTTCTGTGTGTAGATCGTATTTCTGGATTTCTCATCCTAATCACCAAAA[C/T]GTAACATTTGTGAGAAATTATAGCCAAATTAATGTTCACATTAAAATG GTTAACTAGTTCTAATAATAAACTGCAAATAAACTACTTAACATGAGTAGAA
	rs3770585	ATACCTTTTACATGACAGAGTAATGTGAAGCTATTGCTTTGTTATTTGCCAAATTCAGTAAATTCAGTCAAACTAGCATTTCTTAAAATTTGTTTGAATAAACTTGTTCTCAAGAAGAAA[A/G]ATAGGAAAACCCAAGTCTGAAAATTCAA ACTGATAAGGAAAAGATATACTAGGGCCTGAAGTATTTATCTACATCTGAAAATAAATACACTACTTTTTGTGAGCTATTCTTTGAATATTT
	rs2287622	ACAAAGGCATCTGCACCTGTAGCCTAATTGTTTGATTTGGTAAACAGTGACAATCTGAACTTTGCTGTTTTCTTTCTTCTCCAGATTCTAAATGACCTCAACATGG[C/T]CATTAAACCAGGGGAAATGACAGCTCTGGTAGGACCCAGT GGAGCTGGAAAAAGTACAGCACTGCAACTCATTCAGCGATTCTATGACCCCTGTGAAGGAATGGTGGGTGTCTTCCAGAA
	rs2058996	AGATGATGCTGGAGTGTTCTGGTAGCAGGAAGAGGCCCTCACTTTTCTCTGATTAACCAGAGTGGTTAAATTCTAGATTGGGCTTGTTGCTAGCAGAAGAGGTGAAATTCAGGACAACA[A/G]TTGAAAGTTGGTAGCATTTGAACACA TCATTTAGGAACACTCAGTTTTAATCATCATCTACCATGGATACTCAAAAGAATACCCTCGGACCCCCGATTGGTTGACTAGTTGCCTGGTCTT
	rs3770596	AAACACATTAACCTGTATTATCATATTTATTTTCTTTAACAAAGTTCAAAAGATGTCAGCTGTTCATGCTGCCAAGATAAACTGAATTGAAAAAGTTGTTCTTTGTGTATTATCATGTTCTTAATGAAAT[A/T]AAGTCCGTGGTTTGGAATG ATTATGGTTTGACCTCTGATAATCACTTCCATTTGCAAATCAAAATCCTACTCAATTTTCCCTGGCAATGTATTTATCTTAAAGCAGTAGATTTTGAATTT
	rs13416802	AGGAATGGAAAACCAAATATTGCATGTTCTCACTCATAAGTGGGAGCTAAGCTATGAGGATGCAAAGGCATAAGAATGATACAATGGACTTTGGGGACTCAGCGGAAAAGGGTACGAAGGCAGTGAAGGATAAAAGACCA[C/G]AGC TTGGATTCAGTGTATACTGCTTGGGTGATGGGTGCACCAAAATCTCACAAATCACCACTAAAGGACTTACTCATGTAACTAAATACCACCTGTTTCCCAAAAAAATTATGGAAATAA
	rs2287618	TTTTCTCCCTGAAGCTGCTCTGTGTTTGCGATGATTTTATGATCATTTTGTAAAACCACTGCATCACGGCCTGTTTTAAACATTATAACTTGAGCTGTTTCTGCCGAAATTGACTCAAGC[A/G]TTTTGTCTTCACAGGTATGAGAAAAATC TTGTGTTCGCCCAGCGTTGGGGAATTAGAAAAGGAATAGTGATGGGATTCTTTACTGGATTCGTGTGGTGTCTCATCTTTTTGTGTTATGCA
	rs7605199	TTCTCTGGAATAAATCCCCAAGAGTAGAAATGCTGAGTCATATGGTAGTTGAGTGTTTAGAATTTTTTTAATTACTAAACTGTTTTCCAGGATAGCTATGACATGGTTTGAATGGCATGAGTG[A/G]TGGTAGAGTGATCCAATTCTTCCA CATCTTCATCAACATTTAGGGTTATCACTATTTTTAATTTTAGCCATTCCAATAAGTCTATAGTAATATCTCATTTTGATTTAATTTGCATTTCTC
	rs3815676	AGAAATTTTAAAGTATAGGTGTTTACTCCAAATTTTCTCAGGATGGAGGCATTTCTAGGCAGAAAYTGGCATTCTTCATAGTCACTATCAGTTCAACAAAACAATTTATAGTATTTCTCA[A/G]TGCCCCACACAGTTTATCTACTTGGCAA TGGCATCAATGACAGTGATTATTTTCCCCTAAGTTGTAGATKAGYAGCTAAAAACKAAAGTCAGCTCCTGAAATAAACAGAATGAATTCAGA
	rs4148773	CATGCTTCTTTACTTTAATAAGAGTAAACAGGCATGATAGTGTTGAATGACAAAGCTCCCTAGTGGCTTCCTTACACCCCTGGCTATAATCACTGACTTTCACCTCCTGCCCTGCATCTATTCTGACCTA[C/T]ACTGGGGAAAACAGTAT GTGGTCTCAATCCTATGGCTTCTACTAGTGTAGAAGTGTTAATGACATCTTGTTATTAACATCTTATTGTTAATTTGTGGTCTATATTTTAAACAGATAAATT
	rs3814382	GTCAACTCAGAAGTCAACCCATTCAGTTTGCATAGAGGAAACATCTAGAAATCTTGCTTTTCTTTGCTGACCTTGATATATTTGAGATTTGAACATATAATAACATATAAAATTATATGT[C/T]ATAATTTTGGGTTTTATGGGCTAAGTCAT AAACCATCTTATACATAAATTCCAATAGAGAAAAAATGGTGGATGCTGAATTTTAATAAAAAATTTATGACAGAGAAACCTAAAATTGAGA
ABCB4	rs31652	TTTGTAAAAGAGTGGGCACAGAGGGCATGATTATGGGTACTATAGCAATCTGCCAGACATTTTAATATATGTAAAGATTACATCATACTCAGGGTACTGATTCTATACATCCCTTTTAA[A/G]AAACATGTACGCTCTGAATTTGATAAGA CTGGTTTATTGCTAACAATTAAGCCCTTTACACTGTGCCAGCCTCTGTCCTGAAAAGCATCAGTACTATTTCTCTGTGAAAACATTCAGAGT
	rs2097937	TAAAACTTTAAAAAAATACTTAATGATAGGGCTCTTTCTTGCTTTATGGAACCGATATGTGCTTTCAGAGTTAGCTC[A/G]GTAGGGCAATTTCATTTTCTCTTGATGTTGGATGATAAAGCTGGAAATTCATTCATTTAAAAAACTTTGAT TCCACAATAGAATGATTTAGAAAATTACCAAGAGAGATGAGGCTATTGA

Gene	SNP	Sequence
ABCB4 cont'd	rs31653	CAGGAGGATTTGAGGTTGTCTGATGAGGGCTCGGGCAATAGCAATCCTCTGTTTTTGACCTCCTGAGAGCTGAGTCCCCTTATCTCCCACTCTTGTTTCATATTTCTGCAAGTTAACCAA[A/G]TTATAAATATGTTGAATGAACTGTTGC TTAACAGTTGACAGTTCTATTTACAAGAAAAACATTTAGAGTCAATAACAGTTTCTATAACGTTGTTTCAACTTAACAGAGTGTTTATTTTCA
	rs2373593	TTATAAAGGTGAACTACAGACTCCTGCCCAAATTCAAAGCTGACTTCCAAAACAGGGTGCTACCTGCTTATTGCCTCCCCACCGCTTTCAGTTTTGCTGCCTTCTCTTACCTTGAATTCT[A/C]GGCTCTCGAACTTTGACTTTGATCCTC AACCTCCAGCCCTTTCATATTTGGCTTTTGTCATTGATTTTCTAGCATGTAGCCCCGTGGCTCTCTTCCTTGAGTGAAGACCAACTCTTACTG
	rs31666	TTTGTTTGTCCATTGCATGTCTTCCCCGACAGAAACCGCTGTTTGAGGGCACAACTTGGTTTATTTGTGTACCACTGTATCTCCAGTGCCTGGAGCAGCACTAGGCAGA[C/T]AGATGCCCCAAAATATATAGTCCACATTGCCTTGAAT TGCAAGAGAGCCAGAGCTTGCTTTGTTTTGAAACTTTAAAAGATTTGGAGGAGGATCTTTTCTAATTTATTTATTTTTTGGCT
	rs31668	AGGACTTTGGCTTAGTTTAATTTAAGGCTGCTTAATCCCAGAATGGAGCCAGTCAGTGAGGTTGGGAGAAGCAGCAGCTGATGAATTGATCTAATTCAGGCTGTTATGTG[A/G]TGTTTGCAAACTTACCGCTATGATCTCTGAGAATA TGACTGAAAATGCCGGCTGAAGCCCCCCATTGGCAATGGCACATACTGTTCCCACGACAAAGTAGGGCCATTCTGTTTTATTCAG
	rs8187799	ACACACACACACACACACACACACACACACACACACTTTAATGTCTTGATATTCTTTCAGACATCAGGAAGCCAGATCCAGTCAGAAGAATTTGAACTAAATGATGAAAAGGCTGCCACT[A/G]GAATGGCCCCAAATGGCTGGAAATC TCGCCTATTTAGGCATTCTACTCAGAAAAACCTTAAAAATTCACAAATGTGTCAGAAGAGCCTTGATGTGGAAACCGATGGACTTGTAAGTTGTT
	rs31676	CATTCATGAAAAACATCCTACCTAATATTTAACTTGGTTCAGTGTTTAGAAATAATAATTAGAAATAAACAGCAATATGCAAGAACAAAAATAAAAAGCT[C/T]ATGCTCTCTCCTGATATTTATTTTCATATATACAGTTTTTTCGTGGAAG GATTATTTGCCATAAACAATTTTGGGTTTTATGTTTTTGGCATACCATTATTTTTATGTTCC
	rs1149222	ATATTTGAATGCAGTGACTTACTTTTTTACGTCAAAACTTTTTCTTTTTCACTGTAGTCAGTTTATTTAATGTAGTTTATTTAAAATCATACCAAACATGCAGTGATAATTTTTTGGAGA[G/T]GGGGAACCACAGCAAAATATACTGTCAAT TTTGGTCCCACCGATAATTTATCTACCCCCCATCCCATATAGAAATCATAAGATTCTCAGTCTGGCTACAACATGAAAGTGAAAAAGATAA
	rs4148826	AAATAGGTTTGAAAAATCTTTTTAACTCTGTATGTATTCAGGAGAACAATAGGTAGTTTGAGAAAATAGACCCACTCAGGCAATAAAATGCTATATTAGCCTTGCAAATGTTGCTCTTCC[A/G]CAAAAAAAAAAGGAAAGGATAAACCTA AACTTAATCCTTTATGTACTATGTTACATATACATCACATTTTAGATAATAATGAATGCCAGAATGTGACTTAAACTTTTCCTCATGTTATTT
	rs2109505	TTCCAGCTCTTTGTTCTGGCCCCCGAAAGCTATCACAGTCCTGATGGCCCCCAGAGCCTCTTCTGCCACGGCGCCTGCTTTTGCATAAGCAGCTAGTTCTTTGTCACTAAATGCCGAGAG[A/T]ATCTGGACAGAAAAGAAACAGTGAT CACTTTTGTATAGGGAGAAAAGTTTAAAGGCACTCTGGTCAACCCTTTAACCTACTTTTTCTTCCCAAACCCTCTCCTTTAACAAATATTAAAGT
	ABCB4_indel	CGGGAGGCTGAGGCAGGAGAATCACTTGAATCCAGGAGGCGGAGGTTGCAGTGAGCTGAGATGGTGCCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCAGTTTC[-/AGAAA]AGAAAAGAAAAGAAAAAAGAAAAGAGAAGGTATTTAATAGAGCCTTTCTCTTTTTTTTCAGATATGCATATTACTACTCAGGATTGG
	rs1202283	AACATTTTTCATTTAAAATCTTTCCTTGACATATTTTCACACAGTAATTAATTTCTATATGAAAGTGTGACATTAACAATGTACCTACTCTGTTAGCCGCGTATTGAGTTCAGTGGTGTC[A/G]TTGATGTCAAACCATCCTATTTCCTGTCG TAGAATAGCATGAAAAAACTTCTGCCTAATTTTCCTGATCTGTCGACCAGCTGCCAAAGTCCAAAATGAAACTTGTATATAGGCAGCAACA
	rs4148812	CTGTCTGCAGACAGCAATAAATTGTCCTTGCTGCACCTGTGCTGTATAAGTTTGGCTTATTATAGTCAATGACCCACATTCTTAAGATGGATGATTTTTTAAAAAGTAGAGACAGGCTGGCGAT[C/G]ACACCGGACAGAGCTAACTGGT GTGCAAGGAAAGTGAAAATGGGAATTGTGTTAGTAGCACAATGCAAACTCAACAATTTCAAGAAAATGCATTTCTGCATTTGCTTCACAGACCCCCAC
	rs2302386	ATTTCCTTCACCTGTATTAAAAGTCAAACAACAATACACATATTTCTTTAAATTTAAACGAGTATTTGAACAATTTTCTGAGAATTTTACAACTCCTTCT[A/G]TGATATCTGGAGTCAACCAGATATCCAAATCATTAAATAAGATGGTAAT GAATAGCAAAATCAACTCCCAAATTTTTACCCAGTTAAAGAAATGCTATGGATTTTTTTAAAAGGTAAAGA
ABCC2	rs717620	TAACTAACTACCACTTGTTCTGAGTCTGAGAAGAGTCAATATGAAGATAAATGTAATTTCTCACCCAAAAAGTAGAGTTGCAGAACTTCTCCAGCATGATTCCTGGACTGCGTCTGGAAC[A/G]AAGACTCTTCTATTAATATGATTGTGT TGTTTCTTCTTTACTTGTTTCATCAAAGAAAGTGGAACAGGAATTATCTACCATTCTGATGTTCTCCGTAAAGGATGACCTTTCATCCCAACC
	rs2756109	TGAAAAGAACAGAAAGGACCTCTCAGCATCATGCCCAGGCCCTTCACTCTCAGGACTGCCCCTCAGGCCCCCACGCAGATGGGTCTCTCTTTTTACATGGGACTCTTACCAGCTTAGTT[G/T]CCTGGTTTTCTAATCTAAAATAGATTA CAAGTATCAGTTAGGAGCGGGAGAGTGGAACAGTGAAGGCAGCTAGAAGGGCAGAAGCAACTAGGCCAGGGAGAGATGATCAAAACCTGTACA
	rs4148391	GGCAAGGTTCCCTTGCCTCTCTCAACTTGTTTCCTTATAGTTTAAATGAAGATAATAACTCTACAAATGTAAGGTGACTAACTCAACTCAATTTGCCCAGAACTTTTTAAGTCTTAAGAC[A/T]GGAAGGCCTGTGTCCTAGGCAGCCCTT TAGTCCCAGGCAAACTGGAACAGTTAGTCATACTAAGGTTGTTATAAAGTGAGATAATGTATTTAAAATGCTTTGTACATAGTAGGTTGTCAA
	rs2073337	TCAGCTTCCTGAAGAGGAAGTTCACTTTTGATAGACTAGCTGGCATCTCTAGGCCTGACCGCAAGATCCAGGGGACTTGTCCCTCTCACCGCCCCATGCC[A/G]CTTTTCCTCCTTTGTACCATACCCCCAGCTCTAGTTTCACCTTCGT GTACTGTACGATGGTCTGGCTCCCAGAGCAACTTTTCAGGATGCACAGCAAGAATGTGGTATCAGGAAGGATTG
	rs2002042	TCAGTCAACCACATCACCATGGTTACCGTAAGAAAATTGGGCAGTTTTCCCATACTAATGGTAAACATTTTTTTCTGATTAGCCTGGATCTGAGTTTCTGGATTCTGTTATAAACCA[C/T]TTCTCCTTTGAAGAATCCTGTCTTTAATGTA CAAAGAACAGTTTGCAAAACATCCCCGTAAGCCCTTTTTCTGGTGAACACTGAGAGAACATTTCTGAGCAGTTTTGTAAGCTCCAGAGC
	rs7476245	TTGCAGCTCCATAAGAACCCCCATGACGATTAGGTCGGTTCCAGAAGAGCCATGCTATGGTGCTTCCTTGGTTGACTTTGGACAGGCATAGCTGCTTGGAGATGTCTCTCAGCAGCAGCT[A/G]GGTCTTTCTTTTTACTTCGGCAAAG TTCTAGTTTCACTGGGGTGTCTGTGGGTGGTTTAGGCTCAATTCCAGTGTATTCTTCGAGGGAATGTGCAGCTTGAAAGCAAAACAGGGAAACTG
	rs3740066	CCTCCATCCAAATGATGAAGGCTTAGGGCCCAGCTGCTCTCCACTCTGTCCAATTGTTGTTTGATCACAAGGCCTCCCATCCAGGCCTTCCTTCACTCCACCTACCTTCTCCATGCTACC[A/G]ATGTCACAAGTGATCCCTCTGAGGAC CAGATCCAGCTCAGGTCGGTACCGCACTTGGTAGTTGTTAAACTGGATCTTGCCTTTGCTGGGCCAATCTGGCGGAGGCCTCTTATCAGTCACC
	rs3740065	GGCTATAGCAACGCTAAAGAATCGTCTGCTGTACTTGACTTCTGTTGTGTTGCTGCCACCTAGTGGTCAGATACAGGCAGCCACAAATGCATATTACCAAACATTTCAAAGTTCTGGAAA[C/T]GACGTGTCCATCCTAGGGAACTAGCT CAAGTATCCCGAGTAGACCGTGGAATTGACAGTGCTGACTGAACAGACATCAAGAAGAAGCTGCATATGTTTCCTGTAAAACAAGCTGCTCCCC
	rs3740063	ACATCTGCCTCTCACGGCTTGAGCTTCCACCTCCCCATACCCAGTTAATTGTTACAACACACAGTTTCCTATTTTGCCCTCAAGATATGAATAGTTCCCTTAACCTTGTTCGTTTTCATT[C/T]GCGTGATGTAAAATTTTGGCCAGATTAC TTGACATCTCTAAGTCTCAGTTTCTTCATCAGTAATGTGGTGGTGGTTAAACCTTCTGCCATCAGGTGTTTCCCGGCTGACACTGTTACTGT
	rs2756105	AAATGCAATTCAGTTTTTTTCAGTATATTAACACATATGTGCAACCATCACCACAGTCAATTTTAGAACATTTCATTATCTTTAACCCCGTCTCCTCCTTCCCCTTATCCTTAGCAAC[C/T]GCTAAATCTCTTTTCTGTCTTTATAGATTTCC CTGTTCTAGACATTTCATATGAATGGAATTATATAATATGGTTCTTTTTTGATTGGCTTCTTTCACTTAGCACAATGTGTTCAAGGTT
	rs11595888	TAGCTTGATTAGATTTTTTTGTAGATACTTTAACTCCCCCTCTTTTTAAAAGAATTTTAATAAAGCTGAGATAAGAGGCATATTTACTTTTAATTTTACTTTTAGTTTGC[C/T]CCATTATTACCCTAGCTTCTTCCGAGCTCACAAGCTTACC GTACGGCTGACTGTATATGTACTCGGGATCTCTCGTCGACTTGTGCTCAATGACCACGCTCGAGCGTACCTTCACCCTAG
ATP8B1	rs7241054	TTGCCCAGGCTGGTCTCAAACTCCTGGCCTCAAGTGATTCTCCTGCCTCAGCCTTCCAAAGTACTGGGATTACAGGCGTGAGCCACAGTGCCCGGTCCTGATCACCTTTTCTCCATCAC[C/T]TTTTCCCTAACACTTAGTTTCCCTTTCT CCATCACTTCTTCCCTAACACTTAGTTTCCCTGTCCCCGACTTTCTCACTACTAAAGGACCCCTACATTCCTCTAGAAACCTTGTGGGACTA
	rs1968274	CAATGGTAAAAAACAGAATTCTCGAACTATGGAAACTAGTGTGTGAAAGACACAGAATGAGAAAAATGTCTGGAAAAAAATTACTTTCTCCTCTTTGAGGAGTTTCCATAATTACCTAAAGT[C/T]ATCTTCAAATTTAAAAGGACACACC ATTCCCCAGTCTTAGGAATGTGGCTGCCAATAACCTTTTCTCTCGACATAAGACAAGTGAAAAAGAGTGAGGTCTTTGTTCAAGTCCCTGATAAA
	rs17685876	TGATAGGTCAGGAAGTGATACCCAGAACTGCTCTGGCCATCTTCCAGCATCCATATGAAGCTAATAATGAAGGTTTCTGAGAGAAGAGATGGCAAGAACC[A/G]GCCCAATCCTTTCTTGACAACATTAAATTGCTCAATCAACTGGGC CTGAAGCCTACTAATGATTAATGATGACGCACTACTGTAGAGATGTTTTGAGTCA

Gene	SNP	Sequence
ATP8B! cont'd	rs317826	TTGTCCAGGCAGGCCTTGAACTCCTGAACTCAGGTGATCCACCTGCCTTGGCCTCCCAAAGTGCTAGGATTACACACATGAACAACTGTGCCTGGCCGAAATAGATATATTTCTTCCTGAAGATTTCTAAGGG[A/G]CAGAGAAATAAA GTAGTGGGGAGAGAGGGGTGGGGAGGAGGGAGAGAATGGAGGGAGAGAATGGGGGGAGAGAAGAGTGGGGAGAGAAGAGTGGGGAGAGAAGGGAGGGAGGGGAGAGAG
	rs4308033	ATATGAAGAAATAGGGGTTACGATGTGTCACTTGGGGCAGTTTCCTCATGTTTAAATGAGGCAAGACAACAACAGGCAGTTCCAAGTAATCAAGGTGCAGCCTAAATTTGAGCAGATTTTACTTTTAGGGAAACTTGGGG[A/C]CCTTC AGAGGCCATTTTTCTCCATTCCTTCTCAGCTCAGGCCAGGCGCAGTGGCTCATGCCTGTAATCCCAGCCCTTTGGGAGGCCGAGGCGGGTGGATCACCTGAAGTCAGGAGTTCTA
	rs4306606	CTATTGTAGAAACTGCTGAAAATATAGGATTTGCTTGTGAACTTCTGACTGAAGACACCACCATCTGCTATGGGGAGGATATTAAGTGAGTAAAAGCCTGCAGTATGAAATTA[C/T]GTAATAGCTCATCTTTGCAAATGCAAGACTGTC TGTAGAAGATAGCAGATATGAAGAAATAGGGGTTACGATGTGTCACTTGGGGCAGTTTCCTCATGTTTAAATGAGGCAAGACAACAA
	rs317845	TTTTTCTGAGAAACCCTCCTCTGTTTTTCCAGCACATCCTTCTGCATCCTTTCCTGATCCCCCTCAGTAGCAGTTTCTTCCTTATAATTTAGCTCCTGGCCATGTCTTGCTGATTATAAT[C/T]GGTCTTATCTCTCCAGCCAGACGTGGGT TCCCGTCCTGAAAGGGCCATCCCTCTTGTCCCTCTGTCCTTCCAGTGCTTTGACATTTGGTGGGCACAGAGTAGAACAATACAAAGTACTGA
	rs317838	AAGGTGCTGACACTGTTATTTATGAACGGTTACATCGAATGAATCCTACTAAGCAAGAAACACAGGATGCCCTGGATGTAAGTTTCACTTTTATTCTTTCTGCAAGGGATTCTGAACTTA[C/T]TCTCTGATGCTGTCTGTTAAGTTTGCA TGAAAGCTGTATGTTAACTGCTGACCCTCCCTTTGCTTTGCAATAGTTGGGGTGGAAAGAGCAGTAAAGTAGGAATACAGCCACCTGGGTCCC
	rs317837	AGAGCTGACTGTGAAGGAATCTTTTTATATTTATTTCACTATTACATTCAATGTAGATAATGACTTACCAGTTTTGTTGAAGGCAACACTTTTCTAGAGGAACTGATGGTTTTAAGCTAA[C/T]AGTGAATTATGAGATAGTAACATTTCAA ATTAACTTTATAAATGTAGTTTTCTATTAAAAAATCAAACTTAGCTGGGTGTGGTAACTCACATCTGTAATCCCAGCATTTTGGGAGGCTGA
	rs11659313	GCATCCCAAACGATTCTTCGGCTTAAAGGTCTTACCTCTATTTTGTTGTGGTTGTGTTGAGAGGCATCCCGATGGTCCCCTGAGAAACAAACAGGACAAAACAAATTGATTCTACACCTA[C/T]GTCCAGAGGCCCCTCCTTACCTGGCT TTGCTTATATAGTGATGGTGGTAAGACCATATACGGCCTGGACAGGCTAAAATATTTATTATCTGGCACTTTACAGAAAACGTTAGATGACCCC
	rs319454	GGTTTCACTCTGTCACCCAGGCTGGAGTGCAGTGGTGCAAACCCAGCCCACAACAGCCTCGACGTACCCAGGCTCAGGTGATCCTACCTCAGCCTCTCAAGTAACTGGAACTACAGGCACAC[A/G]CAATTTTGCTCACTGCAACCTC CACCTCCCGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCTGAATAGCTGGGACTAGAGGCATCCACCACCACGCCCAGGTAATTTTTGTACTTTTAGT
	rs7236365	TCCTACTAGGAGGTAGTGGCCGGAAAGCACAAGTTCTCCACCTTGGGTGTGGCAGATGATAATAAAGGCTCAGATAAGAGGTGAGGACCACTTCCAAAAGGGAAAACTAAGTTTTTCCAT[C/T]ACCTGCTAAATCTTCCCCTCCTCCC AGTCCCTTTCTGAAGTCTGGGCTTTGGAACCCGGT
	rs160993	TAACCTTACTCAATACTTCATATGATTGAGAATTTAGTAGTCTGAAAAATAATATATGTCAAATAAAAACAGCACTATCATAACGCTTTGAAAGTCTATTTGATTGCTGC[A/G]ATTGACTGATGTATCCTACATCTCCCATAGAAGGTAAG GTTTTGTTCCCAAAATTATCTCTCCAAAGAGAAGGTGGTGTCTTATATTCAAATCCTTGAAGAATCTTATATTACAGGGCAT
	rs319448	AAGTATTGCATGACATTGTAAAATAAGATCACTTAAGAAGCAATATGATAAGTATTATGTTGTAGAAGTTATGAGCATATATTTATGGAAGGAATAAAAAGTCAACCATCCTAGTG[A/G]TACTTGGAGATTCGCAAGTGGGAATAAAATT TCCAGAGACAAATTTAAATTTCTGGACATGGAGTCCAAAAGTGCTATATCTCACAACTTTGTCGTGGAAAGAGCAGAGTGCCAAACAGG
	rs319440	TTGAACCCAGGGGGTGTGGAGGTTGCAGTGAGCTGAGATAGCACCACTACACTCCAGCCTGGGCGACAGAGTGAGACTCTGTCTCAAAAATAAATAAATAAATTGCCTATTGGTAAAGAC[C/T]TTAAAAATCTGTTCCTTTCTCTGAC ACAGATAATAACCTCGATATTATTCATATCTGTCTGGACGCTGCTCAGCAGTTGATAAAGGGCTCAGACACATATGCTCCTGTCTGATCCTTTTA
	rs17686300	CTAATTCCAAGTCTCATCACCGTCATCACATAAGTCTACATCCTGCAGAGTATATAGGAAAGGCTCTTCAAGCATTTGGACAACCTCCTTCTAGTGGTTCTCTCTAGCTTCTGGATACCT[A/G]GAACAGAGCCTAGAGGGCAGTGGCTC CTGAAATTATCTCATTGCAAACTTAACTGGACACCAAGGAAGTTTTGTATTATGATACCTAAAAAAGAGTTGCTCTAAATCTTTTGTGAAAATC
	rs319457	ACTGCACTCCAGCCTGGGTGACAGAGGGAGACTCTGTCTCAAAAAAAAAAAAAGAAAAAAAAAGTTAAGACGAGACTTCTCCCTTGCCACTCCTACCCCCTCACCCCCAGTTAAATATGA[C/T]TTTCACAAAAGATTTAGAGCAACTCT TTTTTAGGTATCATAATACAAAACTTCCTTGGTGTCCAGTTAAGTTTGCAATGAGATAATTTCAGGAGCCACTGCCCTCTAGGCTCTGTTCCAG
	rs319406	GATTACTTGACTGCAGTCAGATTTCTAAAATTGAGTAAAAACGTGTTTGCGGTTCAATTGCAGACAAAGACATACTGGTTACTTTCTCCAGGCTCCCCACAGTATTGGCAGAAGTGT[A/G]AAATTAAGAGATCAAGGTCATTTCTGGTA ACCCCTCACTGCACCTACTCGAGGTCTGATGCCACCCTCTCTGTCTGAATCATCCAAATCAAAATTGGCCAGGGATGAGGGTATAGTGCTG
	rs319409	GTGTGTACATATATGCATTCATATTGTGTATGAATTCTATGCATCACACAGCTGAAAGCACAAGATGGTAGATACCCTTCTGCTTTTGTCTCTTAACAGTGTAATTTGGAGGT[A/C]ATTCCTACCATATTTATGGGGTCACCTCATTCTTT TTTTGCTGACTGCATAGGCTATCATTGTAATCGTGTTTAACTGGCTGTTCCCAGTAGGCATTTGGGTTGTTTCTAGTCGTTCTGA
	rs12456346	TGAAGAGGGCAAAGAGAGGAAGGGAGGAAAGAAGGAGAGCAGACTGGCTGTGAGGGAGGCAGGGCACATGGGAAATGGGAGGAAAGAGGTCTGGCATTTACTGAACATCATCTGTGTGCTTGGGAAGGGC[A/G]CAGAGAGGTA TTAGAATAAGAATTTTTGTAAAAAGTCACAGGCAAAGTCATTTGATATGGGTTTTGAGAAGCGAGAACCACCCAAATCCCCTAGGAATATAAATCACCCTTGGAAAACAT
	rs9676158	CCATCTTGAGTTGACTTTTGTGTATGGTGAGAGACAGGGGCCCAGTTTCATTCTTTAGCATGTGGGGATTTATTTCTGGGTTCCCTATTCTGTTCCATTGGTTTATGTGTCTGTTTTTAG[C/T]CCATATTATGTTTTGGTTACTACAGCCT TGTTATATATATATGTATCAAGGTGACTGTGAAGGTACTATATGGAAACATTTTTTATTGTTGTAAAATATATGTACATAAAATTTACCATT
NR1H4	rs4764980	AACTAAGTCTGCTCTGGCTGTTCCTTCCACTCCATTTTGTGTCCATGGGGCAGAGCAGTGCCCAACCAGTAGTCATGATTCCTCTGCTGGGCACCTCTGCTGCCCAGCCCTCTTGGGTTC[A/G]GAGTTCCACTGTGATATATGGCATA GTTGTTTGAGGATGAGGGTGAGGTTGAAGTGAGTTCTCTCTCTCTTCTGCTTGTTGCATTGGACGGATCTGCCTTATACCTCAATGTTTCCAGAC
	NR1H4_001	CCCCAGCTCTCCTAACCAACCCTTCCCGCATTCCCACAGTCACAAACTATTTATTTTCCTTTCAGGAGTTTTTTTTGAAGACCACCATAAAGAAAGTGCATTTCAATTGAAAAATTTG[G/T]ATGGGATCAAAAATGAATCTCATTGAACAT TCCCATTTACCTACCACAGATGAATTTTCTTTTTCTGAAAATTTATTTGGTAAGTTGTCAAGTTCATTTGAATATCAATAAGAACAAACT
	NR1H4_003	GAGAAGCTGTGAGAGGACTTTTTACACTTTTCAGTGTTTCTCCCACAGGTTTCTTCAGGAGAAGCATTACCAAAAACGCTGTGTACAAGTGTAAAAACGGGGGCAACTGTGTGATGGATA[T/C]GTACATGCGAAGAAAGTGTCAAGAG TGTCGACTAAGGAAATGCAAAGAGATGGGAATGTTGGCTGAATGTATGTATACAGGTATTCACTTCAAGCAATTACATTTCACTAAAAATCTC
	rs1030454	CCTTTTCAGCTTGCACAAAATGAAAACAGTTACCAGAACTATGTTTTCCAAAGATGCAAATCTAAGCTGATTATCAGTTCAGATACTGTCTTTGAGAGGGGTTAGAAATGTAAAGAAACAAATGATAGACT[A/G]TTTATAAGAAGGGTA AATGCTTTAGTTAAATTTAGAAATCTTAACACCTCTACTTTTGCTTTTGAAAGTG
	rs35724	AAGACAATCAGTTGACAAAATTCTCAAAATATAGCCAAAACAATGAGTGTAATGAGTGTATGTCTAATTTTGTCAGTGTCTTTGTTTCAGTCACTACTAATAATCAATGATAATGCTCTT[C/G]AATTGCCAAAAGAATCTCAAAAACGTCT CCATTTTCCAGATGGGATTTAGTGAATCTTTTGAATATGACTATCGATAATTAAGGCAGAGATCACATCTTAAACAACTAACAAAAAGTGGT

Table 2.5: Number, sex and ages of case and control subjects

	n	Female (%)	Male (%)	Median age (range)
Controls	254	119 (46.8%)	135 (53.1%)	66.1 (55-80)
Cases	164	71 (43.4%)	93 (56.7)%	68 (30-92)

Table 2.6: Hardy-Weinberg equilibrium results

Legend: Results from control cohort only shown. Calculated using Pearson's $\chi 2$ test. P-value threshold for non-conformity to HWE set at 0.001. Abbreviations: SNP, single nucleotide polymorphism; A1, allele 1; A2, allele 2; GENO, genotype distribution; ObHet, observed heterozygosity; ExpHet, expected heterozygosity; p, p-value.

Gene	SNP	A1	A2	GENO	ObHet	ExpHet	p
ABCB11	rs497692	A	G	60/108/85	0.4269	0.4951	0.03082
	rs3755157	T	C	2/45/203	0.18	0.1768	1
	rs6709087	G	A	15/85/153	0.336	0.3512	0.4763
	rs853773	T	C	58/109/83	0.436	0.495	0.07312
	rs853772	C	A	61/112/79	0.4444	0.4974	0.09959
	rs3821120	C	G	44/115/94	0.4545	0.4805	0.4321
	rs16823014	A	G	2/34/214	0.136	0.1404	0.6397
	rs4148797	G	A	25/96/133	0.378	0.4096	0.2216
	rs16856300	C	A	44/110/97	0.4382	0.4777	0.1884
	rs4148794	C	T	43/104/106	0.4111	0.469	0.05993
	rs17267869	C	T	8/64/180	0.254	0.2671	0.4763
	rs3770585	A	G	59/127/64	0.508	0.4998	0.8993
	rs2287622	T	C	39/103/112	0.4055	0.4587	0.07478
	rs2058996	A	G	39/134/77	0.536	0.4884	0.1542
	rs3770596	A	T	51/104/89	0.4262	0.4879	0.04953
	rs13416802	C	G	4/50/199	0.1976	0.203	0.7535
	rs2287618	A	G	30/91/131	0.3611	0.4197	0.03471
	rs7605199	G	A	65/125/64	0.4921	0.5	0.8026
	rs3815676	G	A	0/15/239	0.05906	0.05731	1
	rs4148773	T	C	2/34/216	0.1349	0.1394	0.6383
	rs3814382	T	C	43/120/88	0.4781	0.4839	0.8962
	rs10930343	G	A	52/113/89	0.4449	0.4894	0.1585
	rs7577650	A	G	42/117/94	0.4625	0.4789	0.6002
ABCB4	rs31652aa	G	A	2/44/201	0.1781	0.1754	1
	rs2097937	G	A	4/71/176	0.2829	0.2652	0.472
	rs31653aa	A	G	2/38/209	0.1526	0.1544	0.6869
	rs2373593	C	A	2/38/212	0.1508	0.1528	0.6837
	rs31666aa	C	T	6/54/190	0.216	0.2292	0.4021
	rs31668aa	G	A	1/30/222	0.1186	0.1185	1
	rs8187799	G	A	3/32/217	0.127	0.1394	0.1534
	rs31676aa	T	C	15/73/164	0.2897	0.3252	0.08305
	rs1149222	G	T	15/80/156	0.3187	0.3422	0.2708
	rs4148826	G	A	17/69/168	0.2717	0.3233	0.01808
	rs2109505	A	T	17/68/168	0.2688	0.3219	0.01083
	ABCB4_indel	:-	AGAAA	0/9/243	0.03571	0.03508	1
	rs1202283	G	A	60/120/74	0.4724	0.4985	0.4498
	rs4148812	C	G	16/101/133	0.404	0.3905	0.7456
	rs2302386	G	A	9/64/179	0.254	0.2725	0.2544
ABCC2	rs717620	T	C	12/79/159	0.316	0.3271	0.5646
	rs2756109	G	T	52/113/89	0.4449	0.4894	0.1585
	rs4148391	A	T	4/61/185	0.244	0.2379	1
	rs2073337	G	A	39/124/91	0.4882	0.479	0.7948
	rs2002042	T	C	13/87/153	0.3439	0.3469	0.8572
	rs7476245	A	G	0/30/221	0.1195	0.1124	1

Gene	SNP	A1	A2	GENO	ObHet	ExpHet	p
ABCB2 **cont'd**	rs3740066	A	G	33/128/92	0.5059	0.4728	0.2902
	rs3740065	C	T	7/49/194	0.196	0.2202	0.08516
	rs3740063	C	T	47/139/67	0.5494	0.4969	0.102
ATP8B11	rs7241054	T	C	64/126/62	0.5	0.5	1
	rs1968274	C	T	11/89/154	0.3504	0.3415	0.854
	rs17685876	G	A	12/102/138	0.4048	0.375	0.2425
	rs317826	A	G	12/104/137	0.4111	0.3779	0.1866
	rs4308033	A	C	37/125/91	0.4941	0.4772	0.6924
	rs4306606	T	C	16/98/138	0.3889	0.3828	0.8705
	rs317845	C	T	11/103/140	0.4055	0.371	0.176
	rs317838	T	C	26/125/99	0.5	0.4574	0.1674
	rs317837	C	T	11/105/138	0.4134	0.375	0.1317
	rs11659313	C	T	7/67/179	0.2648	0.2689	0.8148
	rs319454	A	G	19/102/128	0.4096	0.4042	1
	rs7236365	T	C	40/125/86	0.498	0.4832	0.6957
	rs160993	A	G	5/56/189	0.224	0.2292	0.7805
	rs319448	G	A	38/118/96	0.4683	0.4735	0.8942
	rs319440	C	T	50/127/72	0.51	0.4961	0.7028
	rs17686300	A	G	15/93/142	0.372	0.371	1
	rs319457	T	C	5/44/202	0.1753	0.192	0.1797
	rs319406	G	A	33/127/90	0.508	0.474	0.2877
	rs319409	A	C	51/131/72	0.5157	0.4966	0.6134
	rs12456346	G	A	40/109/104	0.4308	0.468	0.2264
	rs9676158	T	C	0/30/217	0.1215	0.1141	1
NR1H4	rs4764980	G	A	66/120/68	0.4724	0.5	0.3814
	rs56163822	T	G	0/15/236	0.05976	0.05798	1
	rs61755050	C	T	0/2/247	0.008032	0.008	1
	rs1030454	G	A	4/64/182	0.256	0.2465	0.7965
	rs35724aa	G	C	29/114/111	0.4488	0.4479	1

Table 2.7: Results from allelic association testing, Cochran-Armitage trend testing, dominant and recessive modelling for all SNPs

Legend: This table can be found in Appendix 3

Table 2.8: Summary of SNPs nominally associated with altered risk of CC by gene

Legend: With allele frequency in each group, allelic association odds ratio and Cochran-Armitage trend test results. SNP, single nucleotide polymorphism; A1, allele 1; A2, allele 2; Freq A, frequency of allele 1 in affected group; Freq UnA, frequency of allele 1 in unaffected group; OR, odds ratio; SE, standard error; L95 and U95, lower and upper limits of OR 95% confidence interval; allelic $\chi 2$, allelic association testing result; p, p-value; cor p, Bonferroni corrected p value; FDR, false discovery rate Q value; CATT, Cochran-Armitage Trend Testing result; **bold text - p-value less than 0.05.**

ABCB11

SNP	A1	Freq A	Freq UnA	A2	OR	SE	L95	U95	Allelic χ^2	p	cor p	FDR q	CATT χ^2	p	cor p
rs3770585	A	0.4	0.49	G	0.6939	0.14	0.5192	0.9273	6.122	**0.013**	0.975	0.163	6.201	**0.013**	0.946
rs2287622	T	0.45	0.3563	C	1.478	1.1456	1.111	1.966	7.229	**0.007**	0.524	0.163	6.801	**0.009**	0.657
rs3770596	A	0.4968	0.4221	T	1.351	0.1457	1.016	1.798	4.285	**0.038**	1.000	0.163	4.096	**0.043**	1.000
rs7605199	G	0.4299	0.502	A	0.7483	0.1445	0.5638	0.9932	4.038	**0.044**	1.000	0.163	4.019	**0.045**	1.000

ABCB4

SNP	A1	Freq A	Freq UnA	A2	OR	SE	L95	U95	Allelic $\chi 2$	p	cor p	FDR q	CATT χ^2	p	cor p
rs2097937	G	0.2125	0.1574	A	1.445	0.1836	1.008	2.07	4.045	**0.044**	1.000	0.163	3.924	**0.048**	1.000

ATP8B1

SNP	A1	Freq A	Freq UnA	A2	OR	SE	L95	U95	Allelic $\chi 2$	p	cor p	FDR q	CATT χ^2	p	cor p
rs319454	A	0.3491	0.2811	G	1.371	0.1542	1.014	1.855	4.205	**0.040**	1.000	0.163	4.139	**0.042**	1.000
rs319448	G	0.4625	0.3849	A	1.375	0.1447	1.035	1.826	4.852	**0.028**	1.000	0.163	3.686	**0.032**	1.000
rs12456346	G	0.443	0.3735	A	1.334	0.1458	1.002	1.776	3.916	**0.048**	1.000	0.163	3.686	**0.055**	-

Table 2.9: Haplotype testing results

Legend: Global stat - global score statistic, df - degrees of freedom, p - p-value, cor-p - Bonferroni corrected p-value, **bold text - p-value less than 0.05**

Gene/protein	Global stat	df	p	cor p
ABCB11/BSEP	14	9	0.120	1.000
ABCB4/MDR3	9.9	8	0.270	1.000
ABCC2/MRP2	3.9	9	0.920	1.000
ATP8B1/FIC1	28	13	**0.009**	**0.047**
NR1H4/FXR	8.8	8	0.360	1.000

Table 2.10: Haplotypes in *ATP8B1* associated with altered susceptibility to CC

Legend: Hap Ref - allocated haplotype reference, Hap-Score - score statistic for association of haplotype with the binary trait, p-val - p-value for the haplotype score statistic (based on a chi-square distribution with 1 degree of freedom), pool hf - estimated haplotype frequency for cases and controls pooled together, control hf - estimated haplotype frequency for control group subjects, case hf - estimated haplotype frequency for case group subjects, glm.eff - the haplo.glm function modeled haplotype effects as: baseline (Base) or additive haplotype effect (Eff), OR. lower - lower limit of the Odds Ratio 95% Confidence Interval, OR - Odds Ratio based on haplo.glm model estimated coefficient for the haplotype, OR upper - Upper limit of the 95% odds ratio confidence interval. Individual SNPs nominally associated with altered susceptibility to CC are shown in bold.

Hap ref	Genotyped alleles contributing to haplotype																				Hap score	p-val	Pool hf	Control hf	Case hf	glm.eff	OR lower	OR	OR upper
	rs7241054	rs1968274	rs17685876	rs317826	rs4308033	rs4306606	rs317845	rs317838	rs317837	rs11659313	**rs319454**	rs7236365	rs160993	**rs319448**	rs319440	rs17686300	rs319457	rs319406	rs319409	**rs1245634**	rs9676158									
76	T	C	A	G	A	C	T	C	T	T	**G**	T	G	**A**	T	G	C	G	C	**A**	C	-3.22	0.001	0.075	0.101	0.0324	Eff	0.220	0.490	1.1
56	C	T	G	G	C	C	T	C	C	T	**G**	C	G	**A**	C	G	C	A	A	**A**	C	-1.75	0.079	0.128	0.150	0.1073	Base	NA	1	NA
119	T	T	A	G	C	C	T	T	T	C	**G**	C	A	**G**	T	G	T	A	A	**A**	C	-1.61	0.107	0.059	0.068	0.0347	Eff	0.340	0.76	1.7
23	C	T	A	A	C	C	C	T	T	T	**G**	C	G	**A**	C	G	C	G	C	**G**	C	-1.31	0.189	0.047	0.056	0.0356	Eff	0.340	0.85	2.1
81	T	C	A	G	C	C	C	T	T	T	**G**	C	G	**A**	C	G	C	G	C	**A**	C	-0.42	0.671	0.032	0.033	0.0262	Eff	0.390	1.05	2.8
101	T	T	A	G	A	T	T	C	T	T	**A**	T	G	**G**	T	A	C	A	C	**G**	C	-0.39	0.695	0.059	0.061	0.0652	Eff	0.510	1.08	2.3
85	T	C	A	G	C	C	T	C	C	C	**A**	C	A	**G**	T	G	C	A	A	**G**	C	-0.34	0.734	0.020	0.024	0.0183	Eff	0.460	1.41	4.3
50	C	T	G	G	C	C	T	C	C	T	**A**	T	G	**G**	T	A	C	A	A	**A**	T	-0.15	0.878	0.023	0.02	0.0217	Eff	0.440	1.36	4.2
18	C	T	A	A	C	C	C	T	T	T	**G**	C	G	**A**	C	G	C	A	A	**A**	C	0.37	0.709	0.022	0.02	0.0255	Eff	0.480	1.57	5.2
78	T	C	A	G	C	C	C	T	T	T	**G**	C	G	**A**	C	G	C	A	A	**A**	C	0.41	0.683	0.028	0.027	0.0262	Eff	0.620	1.65	4.4
113	T	T	A	G	C	C	C	T	T	T	**G**	C	G	**A**	C	G	C	A	A	**A**	C	0.53	0.599	0.044	0.038	0.0509	Eff	0.690	1.62	3.8
10	C	T	A	A	A	T	T	C	T	T	**A**	T	G	**G**	T	A	C	A	C	**G**	C	0.98	0.325	0.035	0.029	0.0399	Eff	0.830	2.19	5.8
12	C	T	A	A	A	T	T	C	T	T	**A**	T	G	**G**	T	A	C	G	C	**G**	C	1.08	0.279	0.073	0.064	0.0821	Eff	0.940	1.83	3.6

Table 2.11: Genotyped and associated SNPs, with linked SNPs and functional classification

Legend: LD - linkage disequilibrium, R2 - correlation coefficient of two SNPs in LD

Gene	Genotyped SNP, associated with CC	SNPs in close LD (R>0.8)	LD (%)	Function
ABCB11	rs3770585	-	-	Intronic
	rs2287622	-	-	Missense
		rs2287623	100%	Intronic
		rs2389605	100%	Intronic
		rs2241340	100%	Intronic
		rs4140386	100%	Intronic
		rs2287621	95%	Intronic
	rs3770596	-	-	Intronic
		rs3770594	84%	Intronic
		rs4148782	87%	Intronic
		rs3770591	100%	Intronic
		rs3770590	87%	Intronic
		rs3770589	84%	Intronic
		rs12692889	81%	Intronic
		rs2287620	81%	Intronic
		rs13384680	81%	Intronic
		rs6761690	81%	Intronic
		rs13430236	93%	Intronic
	rs7605199	-	-	Intronic
ABCB4	rs2097937	-	-	3' Intronic
		rs6465112	100%	3' Intronic
		rs1526090	100%	Intronic
		rs12154399	100%	Intronic
		rs4148830	100%	Intronic
ATP8B1	rs319454	-	-	Intronic
	rs319448	-	-	Intronic
	rs12456346	-	-	Intronic

Figure 2.1: Hepatobiliary transport systems in hepatocytes and cholangiocytes.

Abbreviations: ABCG5/G8, two-half transporter for cholesterol; AE2, chloride–bicarbonate anion exchanger isoform 2; ASBT, apical sodium-dependent bile-salt transporter; BCRP, breast cancer resistance protein; BS^-, bile salts; BSEP, canalicular bile-salt export pump; CFTR, cystic fibrosis transmembrane regulator; Cl^-, chloride ions; GSH, glutathione, HCO_3^-, bicarbonate ions; MDR1, multidrug export pump; MDR3, phospholipid export pump; MRPs, multidrug-resistance-associated proteins; MRP2, canalicular conjugate export pump; Na^+, sodium ions; NTCP, basolateral sodium taurocholate cotransporter; OA^-, anionic anions or conjugates; OATPs, organic anion transporting proteins; OC^+, cationic drugs; OST, organic solute transporter; PC, phosphatidylcholine. (Adapted from Geier *et al*, 2006[295])

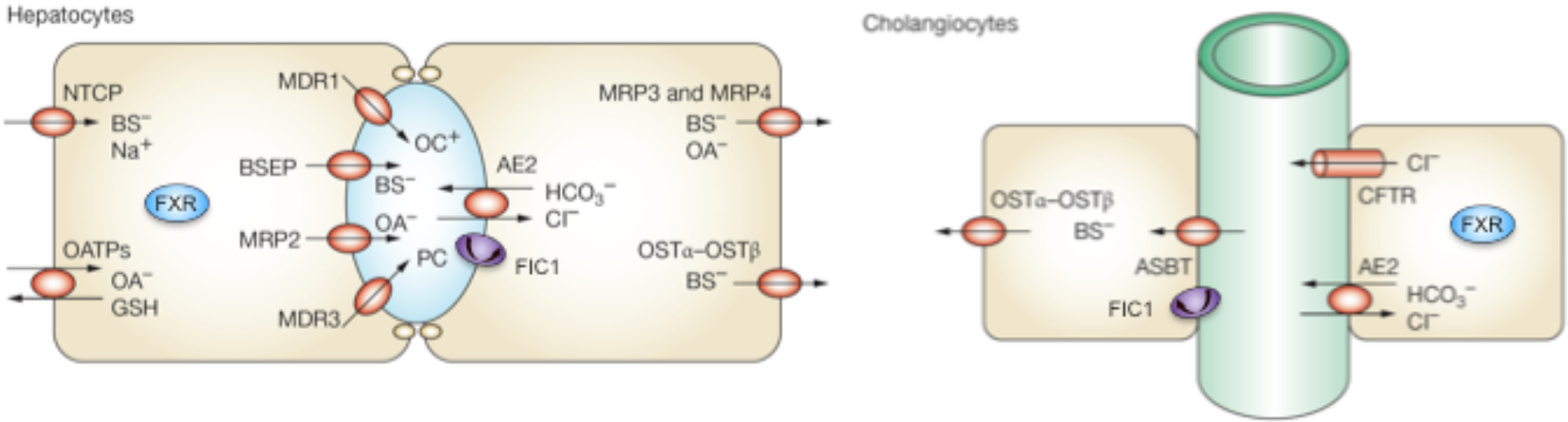


Figure 2.2 Specimen fragment of PLINK-DATA matrix for analysis of *ABCB11* results

Legend: Showing 10 subject IDs in column 1, disease status in column 2, allelic genotyping calls in 4 SNPs in columns 4-10.

10101D9	0	G	G	C	C	G	A	C	C
10188D9	0	G	G	C	C	G	A	C	C
10218D9	0	G	A	C	C	G	A	T	C
10278D9	0	A	A	C	C	A	A	T	T
10346D9	0	G	A	C	C	A	A	T	C
10380D9	0	G	G	C	C	G	A	C	C
10398D11	0	G	G	C	C	G	G	C	C
10477D9	0	G	A	C	C	G	A	T	C
10544D9	0	G	A	C	C	G	A	T	C
10577D9	0	G	G	C	C	A	A	C	C

Figure 2.3 Specimen PLINK-MAP matrix (from *ABCB11*)

Legend: Column 1 - chromosome; column 2 - SNP reference; column 4 - base pair location on chromosome.

2	rs497692	0	169497262
2	rs3755157	0	169500417
2	rs6709087	0	169507256
2	rs853773	0	169522593
2	rs853772	0	169522901
2	rs3821120	0	169525182
2	rs16823014	0	169525959
2	rs4148797	0	169526281
2	rs16856300	0	169526548
2	rs4148794	0	169529894
2	rs17267869	0	169531654
2	rs3770585	0	169532113
2	rs2287622	0	169538574
2	rs2058996	0	169542195
2	rs3770596	0	169548388
2	rs13416802	0	169550584
2	rs2287618	0	169551055
2	rs7605199	0	169564700
2	rs3815676	0	169578625
2	rs4148773	0	169582067
2	rs3814382	0	169597234
2	rs10930343	0	169598031
2	rs7577650	0	169599456

Figure 2.4 Specimen data matrix for Haplo.stats

Subject	Disease	Control	rs497692.a1	rs497692.a2	rs3755157.a1	rs3755157.a2	rs6709087.a1	rs6709087.a2	rs853773.a1	rs853773.a2
10101D9	0	1	G	G	C	C	G	A	C	C
10188D9	0	1	G	G	C	C	G	A	C	C
10218D9	0	1	G	A	C	C	G	A	T	C
10278D9	0	1	A	A	C	C	A	A	T	T
10346D9	0	1	G	A	C	C	A	A	T	C
10380D9	0	1	G	G	C	C	G	A	C	C
10398D11	0	1	G	G	C	C	G	G	C	C
10477D9	0	1	G	A	C	C	G	A	T	C
10544D9	0	1	G	A	C	C	G	A	T	C
10577D9	0	1	G	G	C	C	A	A	C	C

Figure 2.5 Specimen programming script for Haplo.stats

```
library(haplo.stats)
data(haploabcb11dataccafricanandpscexclusion)
names(haploabcb11dataccafricanandpscexclusion)
attach(haploabcb11dataccafricanandpscexclusion)
geno <- haploabcb11dataccafricanandpscexclusion[,c(6:51)]
label <-c ("rs497692", "rs3755157", "rs6709087", "rs853773", "rs853772", "rs3821120", "rs16823014",
"rs4148797", "rs16856300", "rs4148794", "rs17267869", "rs3770585", "rs2287622", "rs2058996", "rs3770596",
"rs13416802", "rs2287618", "rs7605199", "rs3815676", "rs4148773", "rs3814382", "rs10930343", "rs7577650")
geno.desc <- summaryGeno(geno, miss.val=c(0,NA))
print(geno.desc)
table(geno.desc[,3])
miss.all <- which(geno.desc[,3]>10)
haploabcb11dataccafricanandpscexclusion.updated <- haploabcb11dataccafricanandpscexclusion[-miss.all,]
library(haplo.stats)
data(haploabcb11dataccafricanandpscexclusion.updated)
names(haploabcb11dataccafricanandpscexclusion.updated)
attach(haploabcb11dataccafricanandpscexclusion.updated)
geno <- haploabcb11dataccafricanandpscexclusion.updated[,c(6:51)]
label <-c ("rs497692", "rs3755157", "rs6709087", "rs853773", "rs853772", "rs3821120", "rs16823014",
"rs4148797", "rs16856300", "rs4148794", "rs17267869", "rs3770585", "rs2287622", "rs2058996", "rs3770596",
"rs13416802", "rs2287618", "rs7605199", "rs3815676", "rs4148773", "rs3814382", "rs10930343", "rs7577650")
#haplo.cc#
y.bin <- 1 * (haploabcb11dataccafricanandpscexclusion.updated$Disease == "1")
cc.haploabcb11dataccafricanandpscexclusion <- haplo.cc(y = y.bin, geno = geno, locus.label = label, control
= haplo.glm.control(haplo.freq.min = 0.02)) print(cc.haploabcb11dataccafricanandpscexclusion, nlines = 25,
digits = 2) names(cc.haploabcb11dataccafricanandpscexclusion)
```

3. Polymorphisms in Natural Killer cell receptor protein 2D (NKG2D) as a risk factor for cholangiocarcinoma

3.1 Background

3.1.1 Natural Killer (NK) cells

Natural Killer cells are cytotoxic lymphocytes that form a critical component of the innate immune system. One important role is in the direct recognition and killing of virally infected, or malignantly transformed, cells.[332] NK cells are present in the circulation, reticuloendothelial system and other tissues of the human body. They patrol, forming synapses with individual target cells and surveying that cell's surface for activating and inhibitory ligands. Receptors that suppress NK cell activation include those activated by ligands of MHC class I. The expression of MHC class I ligands is decreased in stressed or transformed cells. Activating receptors respond to ligands expressed only by stressed cells, including RAET1 and MICA. In addition, NK cell activation can be stimulated by a number of interferons and cytokines, including IL-15, although simultaneous activation by other pathways is also required.[332, 333]

After the integration of these various signals, the target cell is either released by the NK cell undamaged or is killed via the release of perforin granules and proteases into the synapse. The apposed abnormal cell is perforated and penetrated by the released granzymes, triggering apoptosis or osmotic cell death. Activated NK cells also release cytokines that contribute to the general inflammatory process in infection and tumour killing.[334-336]

3.1.1.1 Normal function of NKG2D receptor

NKG2D is a major *activating* receptor expressed on the surface of CD8+, γδs T cells and NK cells. It is a C type lectin-like receptor formed of a type II transmembrane anchored glycoprotein. It is encoded by a single gene, *NKG2D,* which is located on chromosome 12 and which shows relatively little polymorphism. Even murine and human forms of NKG2D show 70% homology.[337]

NKG2D is activated by an unusually large and diverse range of ligands. These include MIC (A and B), ULBP (1, 2, 3 & 4), RAET1G and RAET1L. These ligands are expressed on the surfaces of stressed cells that are subject to intracellular infection, malignant transformation and heat shock. In human cells exposed to ionizing radiation or genotoxic chemicals, the subsequent DNA damaging pathways have been shown to upregulate the expression of many NKG2D ligands.[338] In NK cells, inhibitory signals generally dominate over activating signals. However, activating signals transmitted through NKG2D cannot be overcome by inhibitory signals, allowing NKG2D to function as a master switch in determining activation of NK cells.[339]

Activated NKG2D associates with the DAP10 adaptor molecule in the cell membrane. In NK cells, this induces cytolytic granule release and target cell death. Activation of the NKG2D pathway also leads to cytokine release by the NK cell. Human NK cells treated with high concentrations of NKG2D ligands secrete cytokines including interferon γ, CSF and MIP-1β. These contribute to further activation of inflammatory and immune pathways, including the recruitment of further NK cells.[340]

There is strong evidence of NK cell importance in the immunosurveillance of cancer cells. However, tumours eventually evade NK cell action and proliferate. This is thought to be due, in part, to very high levels of cell bound NKG2D ligands eventually leading to down regulation of NKG2D expression. Indeed, in advanced tumours, very large concentrations of soluble NKG2D ligands have been demonstrated – further attenuating NK function. Therefore, it is thought that NKG2D activity plays an important role in *early* tumour detection and control but with a diminishing role as the tumour progresses.[341-343]

3.1.1.2 Dysfunction and carcinogenesis

Mouse models of carcinogenesis have demonstrated reduced tumour surveillance and increased tumour progression in NKG2D receptor knock out mice, as illustrated in Figure 3.1.[344, 345] In an attempt to quantify the importance of cytotoxic immunity in

tumour surveillance, a prospective Japanese cohort study was started in 1986. This recruited normal subjects, with no known immunological defect. Circulating cytotoxic lymphocyte activity was quantified and the cohort divided in to low, medium and high activity tertiles. At an 11-year follow up, subjects with low cytotoxic immunity had increased risk of cancer compared to those with medium or high cytotoxic immunity. A subsequent study of the same Japanese cohort considered whether the differences in cytotoxicity could be attributed to lifestyle – smoking, exercise and similar factors. This concluded that only 30% of the difference between the groups could be explained by such factors. Therefore, the same investigators went on to explore genetic susceptibility factors in this cohort. They genotyped a 270kb region of natural killer complex gene region on chromosome 12, which includes *CD94* and *NKG2D* genes, and identified *NKG2D* haplotypes that were associated with lower cytotoxic lymphocyte activity. Finally, a case-control study was performed to study those in the original cohort who had developed a malignancy (n=102) with matched control subjects with high (n=204), medium (n=204) and low (n=204) cytotoxic activity. They genotyped 20 SNPs in the NKC region, covering *CD9*4, *NKG2D*, *NKG2F*, *NKG2E*, *NKG2A* and *Ly49* genes. Eight of the 20 SNPs were associated with altered susceptibility to cancer and seven of these were found to lie within *NKG2D*.[346]

3.1.1.3 NKG2D and CC in PSC

This finding led to a study of NKG2D polymorphisms as a risk factor for cholangiocarcinoma. Melum and colleagues selected 7 SNPs across *NKG2D* and compared the genotype frequencies of 46 subjects with PSC and CC with 319 control subjects with PSC and no CC. Two of these SNPs were associated with increased risk of CC, and this finding persisted after correction for multiple testing. The two SNPs implicated were rs11053781 with an OR of 1.95 (CI 1.23-3.07) and rs2617167 with an OR 2.20 (1.40-3.44). The authors concluded that common genetic variation in the NKG2D receptor may be associated with increased susceptibility to CC in PSC patients. They suggest that, in patients with PSC, good NK cell function may be critical

in the detection and abolition of premalignant biliary cells.[72] They did not genotype any non-PSC related, sporadic CC cases and so no conclusions can be drawn on a potential role in this group. The majority of CC is sporadic, and it is therefore important to seek to validate the Norwegian findings in this group.

3.1.2 Rationale for this study

Genetic variation in the NKG2D receptor has been associated with reduced receptor function and impaired NK cell activation. Such genetic variation has been associated with increased risk of a number of malignancies, including PSC related CC. The same genetic variation may reduce tumour immunosurveillance in non-PSC patients, permitting survival and proliferation of transformed cholangiocytes and so progression to advanced malignancy.

3.2 Hypothesis and study aim

3.2.1 Hypothesis

I hypothesised that variation in the gene encoding natural killer cell protein G2D (NKG2D) is associated with altered susceptibility to sporadic CC.

3.2.2 Study aim

I aimed to investigate the relationship between common polymorphisms in *NKG2D* and sporadic CC.

3.3 Materials and methods

3.3.1 Ethics

The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki. Approval from the local Research Ethics Committee and NHS R&D Department was sought and obtained for this study (Ref 09/H0712/82).

3.3.2 Power calculation

I undertook an *a priori* power calculation using statpages (Statpages) using the same method detailed in 2.3.4.[311, 347] I set a desired α (p-value) of ≤0.05, power of 80% and

case to control ratio of 1:2. The allele frequencies of the SNPs positively associated with CC in the Norwegian study, rs11053781 (case group MAF 0.66, control group MAF 0.50) and rs2617167 (case group MAF 0.39, control group MAF 0.22), were used as a surrogate for pilot data. This generated required group sizes of 112 cases and 223 controls for rs11053781 and 95 cases and 191 controls for rs2617167.

This confirmed that my proposed study had adequate power to detect a difference of the magnitude found in the Norwegian study. Indeed, with the cohort sizes available to me, it had power to detect the predicted difference in rs11053781 and rs2617167 of 88.4% and 95.1% respectively.

3.3.3 DNA resource

I used the same DNA resource that I collated for my biliary transporter study (sections 2.3.2-2.3.5) for this study of *NKG2D* polymorphisms. Therefore, the methods of sample collection, DNA extraction and DNA quality control will not be reiterated here.

3.3.4 SNP selection

I used HaploView (V 4.2, Broad Institute) to search HapMap (V3 Build R2, NCBI) data from genomic regions of interest within, and 5KB up and down stream of, *NKG2D*. The polymorphisms selected were relatively common with a minimum mean allele frequency (MAF) of >5%. Markers with a MAF of less than 5% were excluded. I then used the application tagger within Haploview to select SNPs that captured the maximum genetic variation in NKG2D, with the minimum number of SNPs. The two SNPs identified to be of interest in the Norwegian study were force included. Pair-wise comparisons only were used with an R^2 cut-off of >0.8, a measure of linkage disequilibrium (LD) between two SNPs. This resulted in a total of 7 SNPs to be genotyped in *NKG2D*. These SNPs are listed in Table 3.1. Due to LD, the SNPs selected represent far more variation around the candidate gene than the absolute number of single nucleotide polymorphisms genotyped.

3.3.5 Primer design and genotyping

I performed primer design by collating the corresponding DNA sequence from the NCBI dbSNP database for each SNP shortlisted. These are listed in Table 3.2. I reverse checked all DNA primer sequences by searching the NCBI basic local alignment search tool (BLAST). I then input these sequences into 'PrimerPicker' (KBioscience). This is proprietary, web-based software that generates designs for the oligonucleotides necessary for the KASPar genotyping process.

KBioscience performed the genotyping on the same KASPar system used for my biliary transporter study, discussed in section 2.3.7.

Raw data from the genotyping process were returned to me in comma separated values (csv) format.

3.3.6 Statistical analysis

I managed and manipulated the raw genotyping data with MS Excel (Microsoft). Differences were considered significant if $p < 0.05$.

3.3.6.1 Hardy-Weiberg Equilibrium (HWE)

I confirmed HWE in all 7 genotyped SNPs using Pearson's χ^2 test in PLINK (V1.07). I used a p-value threshold of 0.001, in line with standard practice and the HWE p-value criteria set in the tagger algorithm during SNP selection. I determined that any SNPs that breached this HWE threshold in the control cohort would be excluded from further analysis.

I prepared a PLINK-DATA matrix and a PLINK-MAP matrix of the *NKG2D* data, using the technique described in 2.3.8.1.1. I produced a programming script to run the HWE analysis, which is reproduced here:

```
>PLINK --PED PLINKNKG2DDATACCAFRICANANDPSCEXCLUSION.TXT --1 --
NO-FID --NO-PARENTS --NO-SEX --ALLOW-NO-SEX --MAP
PLINKNKG2DMAPCCAFRICANANDPSCEXCLUSION.TXT --HARDY
```

3.3.6.2 Allelic association testing (χ2)

I used the following script to execute χ^2 allelic association testing and calculate associated odds ratios, with 95% confidence intervals:

```
>PLINK --PED PLINKNKG2DDATACCAFRICANANDPSCEXCLUSION.TXT --1 --
NO-FID --NO-PARENTS --NO-SEX --ALLOW-NO-SEX --MAP
PLINKNKG2DMAPCCAFRICANANDPSCEXCLUSION.TXT --ASSOC --CI 0.95
```

3.3.6.3 Cochran-Armitage Trend Testing (CATT)

I used the following script to execute CATT on the *NKG2D* data:

```
> PLINK --PED PLINKNKG2DDATACCAFRICANANDPSCEXCLUSION.TXT --1 -
-NO-FID --NO-PARENTS --NO-SEX --ALLOW-NO-SEX --MAP
PLINKNKG2DMAPCCAFRICANANDPSCEXCLUSION.TXT --MODEL
```

3.3.6.4 Haplotype analysis

I used haplo.stats (R V2.10.1, Haplo.stats V1.4.0 [319]) to perform a haplotype analysis on all 7 SNPs genotyped in *NKG2D*.

3.3.7 Correction for multiple testing

Any p-values of <0.05 were to be corrected for multiple testing. In this case, correction was required for 7 independent tests in the SNP association tests. No correction would be required for a positive haplotype test as only one haplotype test was to be performed.

I planned to apply Bonferroni correction to all calculated p-values of <0.05 using MS Excel (Microsoft). I would also perform permutation testing (PLINK V1.07) and a predicted false discovery rate calculation (FDR, R V2.10.1, QVALUE V1.0[321]) as alternative, less conservative corrections for multiple testing.

3.3.8 HapMap and NCBI dbSNP interrogation for associated SNPs

I planned to use HapMap to identify all other, non-genotyped, SNPs known to lie in strong LD (R^2>0.8) with any genotyped SNPs found to be associated with altered susceptibility to CC. Each of the genotyped and associated individual SNPs, and those

found to be in strong LD, would then examined in dbSNP[322] (NCBI, Bethesda, USA, Build 132) to explore potential functional consequences.

3.4 Results

3.4.1 DNA extraction

For this experiment, I used the same DNA resource that I prepared for my biliary transporter study. Details of the DNA extraction and QC are as previously described in section 2.4.1.

3.4.2 Demographics of case and control groups

All samples were successfully genotyped. Data from 164 CC subjects and 257 controls were included in the analysis. All cases included in the analysis were from Caucasian patients without PSC. The two groups were well matched in terms of sex and age (Table 3.3).

3.4.3 Harvey-Weinberg equilibrium (HWE)

HWE was confirmed in all genotyped SNPs in case, control and combined groups. HWE results from the control group, for each SNP genotyped are presented in Table 3.4. As none of these SNPs breached the defined p-value threshold of <0.001, all genotyped SNPs were included in subsequent analyses.

3.4.4 Alellic and Cochran-Armitage trend testing

Allele frequency and Cochran-Armitage trend testing results for each SNP are listed in Table 3.5. None of the SNPs genotyped were associated with altered susceptibility to CC. Dominant and recessive models were also tested, with no significant difference between groups.

3.4.5 Haplotype analysis

I undertook a haplotype analysis, to detect association between different combinations of SNPs in *NKG2D* and altered susceptibility to CC. The haplotype analysis result is

summarised in Table 3.6. There was no evidence to suggest any haplotype differs in frequency between cases and controls ($p > 0.1$).

3.4.6 Correction for multiple testing

As none of the statistical tests demonstrated a significant difference, correction of p-values was not required.

3.4.7 HapMap and NCBI dbSNP interrogation for associated SNPs

As I found no SNPs to be associated with altered susceptibility to CC, this was not performed.

3.5 Discussion

In the recent Norwegian study by Melum and colleagues, polymorphisms in the gene encoding natural killer cell receptor G2D (*NKG2D*) identified two SNPs that were associated with altered susceptibility to CC in patients with PSC.[72] The same genetic variation may reduce tumour immunosurveillance in non-PSC patients, permitting survival and proliferation of transformed cholangiocytes and so progression to advanced malignancy.

My study is the first to examine *NKG2D* polymorphisms in *sporadic* CC. The study had adequate *a priori* power. I found no relationship between common genetic variation in *NKG2D* and susceptibility to CC. This is in contrast to the prior finding of the study by Melum and colleagues of an association between rs11053781 and rs2617167 and CC, in their study. The SNPs tested in my study, and those associated in the Norwegian study are illustrated in the LD plot in Figure 3.2. Although this proved to be a clear negative study, the findings are of importance none the less.

The failure to reproduce the association may reflect an important difference between the pathogenesis of sporadic CC and that of PSC-related CC. PSC, unlike risk factors such as cholelithiasis and hepatitis C, is a *strong* risk factor for CC - with a lifetime incidence of CC of around 15% in PSC patients. PSC is an autoimmune disease that remains poorly understood but is associated with other autoimmune diseases. There

are clear genetic associations between PSC and variation in the HLA genetic region.[65] PSC-related CC has significant clinical differences to sporadic CC, including a much earlier age of onset, frequent multifocal high-grade dysplasia and a particularly poor prognosis.[348-351] It is therefore conceivable that NK cell killing plays a more important role in PSC than its does in CC patients with otherwise normal bile ducts.

The populations of the Norwegian study were recruited from Norway and Sweden. My cohort and controls were Caucasians residing in the UK. It is possible that a genetic influence in the Scandinavian population may not be present in the UK.

Alternatively, my study may present a false negative. Although it was well powered to detect differences of the magnitude observed in the Norwegian study, I cannot exclude the possibility of smaller effects in non-PSC-related CC. Confidence intervals from my study suggest any such effects must have OR <1.5 and considerably larger studies would be needed to detect, or exclude, effects of this magnitude. Finally, although executed with statistical rigour and with strong positive results, the Norwegian study may have reported a false positive in PSC-related CC.

I can conclude that common genetic variation in NKG2D does not contribute substantially to *sporadic* cholangiocarcinoma risk. My findings cannot refute those of Melum and colleagues, as my study excluded patients with PSC-related CC. This could be elucidated in an additional candidate-gene validation study in further cohorts of patients with sporadic CC and PSC-related CC, along with appropriate control groups. However, as genetic susceptibility to CC is likely to be highly complex and involve many genes, a genome wide association study (GWAS) would offer the advantage of being an unbiased screen for associated genes. With increasing availability and affordability, a GWAS may also prove a more cost-effective method for further exploring such genetic factors. CC is a relatively rare disease and such a study would require a multi-centre, international collaboration to collate adequate numbers of well-characterised cases and controls.

3.6 Chapter 3 Tables and Figures

Table 3.1: SNPs in *NKG2D* selected for genotyping

Legend: By gene, RS number and location on chromosome. SNPs force added as associated in Norwegian PSC/CC study in bold.

Ref	RS number	Chromosome	BP location
1	rs7397310	12	10412260
2	rs10772271	12	10415387
3	rs1049172	12	10417007
4	**rs11053781**	**12**	**10428536**
5	rs12819494	12	10442808
6	rs2617165	12	10445197
7	**rs2617167**	**12**	**10450498**

Table 3.2: SNP DNA sequences in *NKG2D* for oligonucleotide primer design

Legend: SNP - single nucleotide polymorphism, [x/z] - variable allele of SNP in preserved DNA sequence

SNP	Sequence
rs7397310	AAAATTTAATTTAACCAACTGTTGCCTAGAAATCGATGATAATGGTTGAGCTGTCATGGAAATCACCGCTAGAATGCACACATTGGCCCATATTCCAGTTCAGACTTGGTCTGGGTGGTCCCCGGATTCCTTGTTTGGAGGATG GTTCTCAGCCTTTAGAGGATTCAAAAGCCTCATTGGTGGGTTCTTGGTTATTCTTGGCATCTGCCTCATCCTCCCTTGCCTTTTACCCCTGTTTATTAGAAGTATTCAGTCAACTATGGAGGCAATAGTGGCCTGATACACTAC CATGCAGATGATGGCATTAACCAGATATCAGTCGCTGCTGGTAGAAGCAGCAGCTCAGCTCCACGAAGACGTGGCAAATAGTGGTGCTTTCTGTTAACACATTTGTTATAAAAAGCACCAAAGGGGGGAATGGAACAGGAATTA AAAGAAATTAAAGAATGTGTAAGCAAAAACTCAGTTGTATGTAAGAAAACCCAATTCCCCCTGAGGAAGAGAAAGAACTGGAGTCCTTTAAAAATTAACTGCCTGTTTTTTCCGTGGCTAGTGAGCCTTATCTCTCCCTTTCCCA GGCATTGTGAAGACCCTGTTTCTCTAGCTGTGCAGCTGCAAGGTCACTAGACAGACAAACTCAAGTCATAAAACATGCTTTTCCTTGAAAAGTAAGAAACAATGTAATGTATGTCTCAATTGAATAACTGTCTTTGTTTCTTGC TTCTGTAATATGCTTCCCCCTGCACAAATCTTCCCCCACAGACAAAATACTTAAAAGATAACTTGACTCTTTGTTCAGGGCTCAGTCATTTGGATGTTAATCCAACTGGGCCAGTGCACCTAAATAATAAAAATCCTCCTCAAC CCCTTGGTGTTCCTGA[C/T]TCCTTATCAATCCTGCTACAGGTGGAGAAAAAAGAAAGACCAGCAGAAAAGAAAAAAACTAGGAGAGGAAGTCAGAGGGAAAGAAAGCAAAAACAGCAAGCATAGAGCCAAACAGCCAGGCAG GCACGCCAAGGGGTAAGTCCCTCTCCCCAAAGAGCTTGGGGACTGGTGCCGGGAGGAGAGCAGAGGAGGTGAAAGGGCATAATTCTTGCAATTTGCGGCAGGCATCTGCCAATCCTCTGGGCCGGTGACTGCCTGGGGCCCAGA CTA
rs10772271	TGGAGTGTATGGTGGTTCTTTTTCAAAATCCCACTGCTTGGGATAATCCATCATACTGAGCGCTCTCTGCTATTTCCTGAAATTCAACACCCTGCGGGTCAGCCTCCGAGGGCAATCCAGCTTCCACCTCTAAATGCCAAGTTT ACTTTGTGCCTCTCATGGCAGGAGGAAAATTTGGCGTTCCTTGAAAGCATAAAAGGATACAAGAAGCCCAAGCCTTTCCAAGAACTTGCCCATCAGTCCATGCTTAGCCATCCCCAAGCAAATGTATAGTGGTACTGTGGAAGA CCTTTACTGGACACTCGGCCAAATAACTGGAGCAGTACTTGTGCTCTAATCCAATTGGCTATCCGCTTCACGCTGGCATTTCGTCAATCTAAAAAGGTAAAAACAAAACACCGCAGGCCAAAGGAAACTCCTTGAGTCCTTTAA TCCTCAGGTTTACGTAAATGTTATTAAGGTCCGTGAGAAGTGCCAAATAAGTTTAAAGCACAAAAATCAAATAGCTGCAAAATTTAAATCCATGTTGGTCTGGTAGGTAACCATAAATTAAAAAAAAATTTAACTTCAATTCAA TTCTCACCTCGTTT[A/G]CCCTTACTATTGGTGTACTTATTCTTGTAGACTGTTATATTATGCCCTGCCTTCAAGGTTTTATGCAAAAACTTGTCTCTGCTGCTCTTACAGAGTTAACTCCTAACTATCCTCCACCCTATTGA AAAAAAATTACCTCTCTTAGAAGGACAAAGAATTAAGTCAAAACATATTAAACAAGACTGAGGAGGAATTATGAAATAAAAAAAGGGGGAATATTGTAAAAAGTAAAGTAGAGGTTCCTCTTCAAAGACTTTCCTCCCCATCTA ATTAGAAATAAACAGTAACTTCTCTTAGAAGCAAAATTTATTCAAAGACCTGTGCTCCTAAATATTTGCCCTGGCATGCTTATACTGGTCCAAGCAAGCGTTAGGTCATAGCCTGTTCCTCTTCCTTATTTAAAAGTGTTTTTA TCTTTCTCAGCATTCCACAAGTTACTTCCTCCTTCCTTTGTTCTCCTCTACCTTTGCCTCTTTTAAAAAGTTCTAAGTTGCTAGCCAATCGGGACAAATACAGAATGTGAGGTCCCGTTCCAGCCAGTGGAAACCGGACACAGC AGTAGTGTGGACATGTCAGGTTATAAATGACCCTGTCTCCTTTGTTCGGTGTACTCTCATGGCAGAACTGCTAGTGAGTG
rs1049172	TCCTAAAATGGTATTAAGTTCCTGTATTTGAAGTTCAATAGGTATGATTCATTTTTATTTTCCCGCGTTAATAAAATCATGTATGTAATTTGTAATTGTCTCTCCTCTAGACTAACAATAATTGAAATGCAGAAGGGAGACTGT GCACTCTATGCCTCGAGCTTTAAAGGCTATATAGAAAACTGTTCAACTCCAAATAC[A/G]TACATCTGCATGCAAAGGACTGTGTAAAGATGATCAACCATCTCAATAAAAGCCAGGAACAGAGAAGAGATTACACCAGCGGT AACACTGCCAACTGAGACTAAAGGAAACAAACAAAAACAGGACAAAATGACCAAAGACTGTCAGATTTCTTAGACTCCACAGGACCAAACCATAGAACAATTTCACTGCAAACATGCATGATTCTCCAAGACAAAAGAAGAGAG ATCCTAAAGGCAATTCAGATATCCCCAAGGCTGCCTCTCCCACCACAAGCCCAGAGTGGATGGGCTGGGGGAGGGGTGCTGTTTTAATTTCTAAAGGTAGGACCAACACCCAGGGGATCAGTGAAGGAAGAGAAGGCCAGCAGA TCACTGAGAGTGCAACCCCACCCTCCACAGGAAATTGCCTCATGGGCAGGGCCACAGCAGAGAGACACAGCATGGGCAGTGCCTTCCCTGCCTGTGGGGGTCATGCTGCCACTTTTAATGGGTCCTCCACCCAACGGGGTCAGG GAGGTGGTGCTGCCCCAGTGGGCCATGATTATCTTAAAGGCATTATTCTCCAGCCTTAAGTAAGATCTTAGGACGTTTCCTTTGCTATGATTTGTACTTGCTTGAGTCCCATGACTGTTTCTCTTCCTCTCTTTCTTCCTTTTG GAATAGTAATATCCATCCTATGTTTGTCCCACTATTGTATTTTGGAAGCACATAACTTGT
rs11053781	TCATATGACTCAAAGCAATCTACAGATTCAGTGCAATTCCTAACACGATACCAGTAACATTCTCTACAGAAATAGAAAAATATCTTAAAATTTATATGGTACCACAAAAGACTGAATAGCCAAAGCAATCTTGAACAAAAATGA AGAAAACTGGTGGCATCACGTTACCTGACTTCAATATATACGACAAAACTACAACAACCAAAACAGTATGGCATTGATATAAAAATAGACACCTAGACCAATAAAACAGAATAGACAACCCAGGAATAAATCCATGTATTATAG CCAACTGATTTTTCCAGAAAGTCACCAATAACATACACTGGGAAAGGAGAGCCTCTTCAATAAATGAAGCTGAAAAAACTGGATATCCACATAGAGGAATGAAATTAGACCCTCATCTCACACCATATACAAAAATCAACTCAA AATGGATTAATGACTTAAATATAAGAACTGAATCTATAACACTACTAGAAGAGATAAGAGAAACACCA[C/T]AGGACATTAGTCTAGGCAAAGATTTTATGGGTAAGAATTCAAAAGTACAGGCAACAAAAGCAAAAAGAAAC AAATGAGATTGTATCAAACTAACACACTTCTCTACTGCATAGGAAACAATTGACAGAGTGAAGAGACAACTTGTAGAATAAGAGAAAATACTCGCAAACTATTCCTCTAACAAGAGACCAATATCCAGAATATACAAGGAATTC AAACAACTCAACAGCAACAACAAAAACTGATTTGAAAATGAGCAAAGGATCTGAATAGATATTTCTGAGAAGATATACAAATAGCCACAGATATATTTTTTGAAATTCAACATCATTCATCATCAAGAAGATGAAAATCAAAAC CACAATGAATATCATCTCACCCTAGTTAGAATAGCTATTATTTAAAAAGACAAAAAATAACAAATGACAGTGAGGATGCAGGGAAAGGCTAATGTTTATACAATTTTGGTGGGAATGTAAATTAGTATAGCCATTATGGAA
rs12819494	AAATAATATATTGGCTGACAACCAAGCTATGCAATTCACAAGGGCAATCCCTAGGAAATGGGGCTAAAAAATAAAAATAAAAAATTAGAATTTGAAACAAAAACATCAGCAAATGCATTTTAAAGGAGAAAGATTCTGCAGTAA AATTTTATGTAAGTTACCCCCCCAAAATCTTTATAAACCCCCCAAATTTATTTGTATGATACTATATTATAAAAAATGTCTATTATAAGAAATTAGAAAACATGCAGAAACTGGAAAATGTCTTCCATACTTGGAGGGAAAGAC AATCAGTCAATAAAATCTGACTTTTTCTGGGACAAGATGGTAGATTTATCAGAAAATATATTTTAAAGTAACTATTATAAATGTACTTAAATTATACAGAAAAGTATGTTCAATTAATTAAATGAAAATATGATGGCAATGAGA AAACAAGTAGGAAATTTTGTCAGAGAAATGAAAACCAAGAAATAAACCAAATTCTGAGTTGAGAAGTA[C/T]ATTACTTGAAACGAAAAATTCACTAAATTGGCTAGAAAAAAAATTGAGATTCAAAAGGAAGAATTCATCAA CTTTGGGATATATTAATAGAAATTATTCATCCCAGAGAAAAAGAGAATACAGAAAGATGTAGAAAAATAATATTAAGACAAAACTGACTTAAAAACTTTTCAAATTTGCTTTTTTAAAAATGTATACATCCAAGATGCTAAGCA AACTAATGACTAATAGAATCAGCATGTAGAAATCCCCATCTCAAAACATCATAATCAAATGATTGAAGAGGATTTTAAGACAAAGACCCTGAGATAATCTTGTGGAACAGGTAGGACGGATTATGAGATGACTTGAGTCTGTGT GAATGCCAGGGAATGGTTGAGGTATCCCCCACCAGCAGGAGTATTTTGACCACATTACCTTCCAGAGAACCCACATTTGCCCTGAATACTATTCTAGATTGGAAGGAGGCAACATCAACAAGATAGAGGAATAGCATACCC

rs2617165	TTGTAAAATGTGAATATAGTTAACACTACTGACCTGTTCACTTAAAAATGGTTGAGATATATGATATGGTTAAGATGTAATGGCTCAGATGTAACATACTGTATGAGTTCAACTATATGACATTCTGGAAAAGTAAACAACTAT GGAGACAGTACAAAAATCGATGCTTGCCATGTGTTTGAAGGAGGAAAGGATGAATG[A/G]GCAGAGCACAGAAGATTTTCAGGGCAATGCAATTATTCTGTATGATCTATGATGATAAAGACATGTCATTATCATTATATATT TGGCCAGACCCATAAAGTGTAAAACACTGAGTGAAACTTAACATAAGCTATGGTCTTTCAATAATAATGATGTTTCAATTGTAACAAATATACCATGCTGATGGGGGATATTGATCA
rs2617167	TTCTCCAGATATTCTGATTAAATAAATGTATTTTTCCACTATGTGTCTCATAGTGGAACTTTTCCTCCTTGAAGTTTAGAAGGAAATTGAAAGGTTGCCATTGAGGAAATAAATGACACAGCAGCCAAAAATTGGCTTTTCTGG GCACCACACCACTCTCTAACTAGATAAGCCCTCACTTTTATCTGATTTGGTTTTGGTGGAATAACAATGGTCTTGACCAGGTACATGGACTCCTTTTTCTCTGCTACTCCCTAATAAATACAACTCTAAACTCTGCAAGTAATG AAAGACACACAG[A/G]AAAACTCTGAAAGGTGGAAAGAGGAAGGCAAAATAGTTTGGGACCCCAGAATCAGGGATCAACAGAGTGTCCGGGCATCTTCAGTCCCCCACTCAGCAGAAGAAAGTGATCTAGTCCTCTGTTTCTC AGAACCAAATCTGGTAACAGGAGGTGGGTCAGCTAAGCCCATTCCCACTCTGTATTGAAAGGGAGTCCTGCCAGCAACATTAGGCAAGCCTGGTACCACTGGCAAGGATGACCTACCAGGAGGCCTGCTACCAGTTAGGAGCTA GAGGAAGTGCTTCTCCTCCCAAACCAAAA

Table 3.3: Demographics of case and control groups

Legend: n – number in group

	n	Female (%)	Male (%)	Median age (range)
Controls	257	121 (47%)	136 (53%)	66.1 (55-80)
Cases	164	71 (43.4%)	93 (56.7)%	68 (30-92)

Table 3.4: Hardy-Weinberg equilibrium results for SNPs tested in *NKG2D*

Legend: Using Pearson's $\chi 2$ test. P-value threshold for non-conformity to HWE set at 0.001. Abbreviations: SNP, single nucleotide polymorphism; A1, allele 1; A2, allele 2; GENO, genotype distribution; ObHet, observed heterozygosity; ExpHet, expected heterozygosity; p, p-value. Results from control cohort only shown.

SNP	A1	A2	GENO	ObHet	ExpHet	P
rs7397310	T	C	12/71/167	0.284	0.3078	0.2191
rs10772271	G	A	33/129/85	0.5223	0.4778	0.1826
rs1049172	G	A	22/99/129	0.396	0.4084	0.6428
rs11053781	T	C	53/119/73	0.4857	0.4967	0.7968
rs12819494	T	C	2/60/190	0.2381	0.2217	0.3909
rs2617165	A	G	6/63/175	0.2582	0.2601	0.8091
rs2617167	A	G	19/92/140	0.3665	0.3838	0.5102

Table 3.5: Results from allelic association testing, Cochran-Armitage trend testing, dominant and recessive modelling for all SNPs genotyped in *NKG2D*

Legend: Abbreviations: SNP - single nucleotide polymorphism; A1 - allele 1; A2 - allele 2; Freq A - frequency of allele 1 in affected group; Freq UnA - frequency of allele 1 in unaffected group; OR - odds ratio; SE - standard error; L95 and U95 - lower and upper limits of OR 95% confidence interval; allelic $\chi 2$ - allelic association testing result; p - p-value.

SNP	A1	Freq A	Freq UnA	A2	OR	L95 OR	U95 OR	χ^2 TEST	Freq A	Freq UnA	χ^2	DF	P
rs7397310	T	0.1925	0.19	C	1.017	0.7122	1.451	ALLELIC	62/260	95/405	0.00822	1	0.9278
								TREND	62/260	95/405	0.00762	1	0.9304
								GENO	8/46/107	12/71/167	0.008457	2	0.9958
								DOM	54/107	83/167	0.005106	1	0.943
								REC	8/153	12/238	0.006038	1	0.9381
rs10772271	G	0.4317	0.3947	A	1.165	0.8759	1.549	ALLELIC	139/183	195/299	1.1	1	0.2942
								TREND	139/183	195/299	1.127	1	0.2884
								GENO	33/73/55	33/129/85	4.004	2	0.1351
								DOM	106/55	162/85	0.002734	1	0.9583
								REC	33/128	33/214	3.661	1	0.0557
rs1049172	G	0.2625	0.286	A	0.8886	0.6481	1.218	ALLELIC	84/236	143/357	0.5383	1	0.4632
								TREND	84/236	143/357	0.5219	1	0.4701
								GENO	12/60/88	22/99/129	0.5229	2	0.77
								DOM	72/88	121/129	0.4527	1	0.5011
								REC	12/148	22/228	0.2168	1	0.6415
rs11053781	T	0.4938	0.4592	C	1.149	0.8669	1.523	ALLELIC	159/163	225/265	0.9335	1	0.334
								TREND	159/163	225/265	0.9504	1	0.3296
								GENO	36/87/38	53/119/73	1.959	2	0.3756
								DOM	123/38	172/73	1.876	1	0.1708
								REC	36/125	53/192	0.03005	1	0.8624
rs12819494	T	0.1415	0.127	C	1.133	0.752	1.708	ALLELIC	45/273	64/440	0.3577	1	0.5498
SNP	**A1**	**Freq A**	**Freq**	**A2**	**OR**	**L95**	**U95**	**χ^2 TEST**	**Freq A**	**Freq UnA**	**χ^2**	**DF**	**P**

			UnA			**OR**	**OR**						
Cont'd								TREND	45/273	64/440	0.3447	1	0.5571
								GENO	7/31/121	2/60/190	NA	NA	NA
								DOM	38/121	62/190	NA	NA	NA
								REC	7/152	2/250	NA	NA	NA
rs2617165	A	0.1398	0.1537	G	0.8946	0.5998	1.334	ALLELIC	45/277	75/413	0.2986	1	0.5848
								TREND	45/277	75/413	0.2867	1	0.5923
								GENO	5/35/121	6/63/175	0.9733	2	0.6147
								DOM	40/121	69/175	0.5815	1	0.4457
								REC	5/156	6/238	0.1535	1	0.6952
rs2617167	A	0.2112	0.259	G	0.7661	0.5486	1.07	ALLELIC	68/254	130/372	2.454	1	0.1173
								TREND	68/254	130/372	2.353	1	0.125
								GENO	8/52/101	19/92/140	2.356	2	0.3079
								DOM	60/101	111/140	1.955	1	0.1621
								REC	8/153	19/232	1.083	1	0.2979

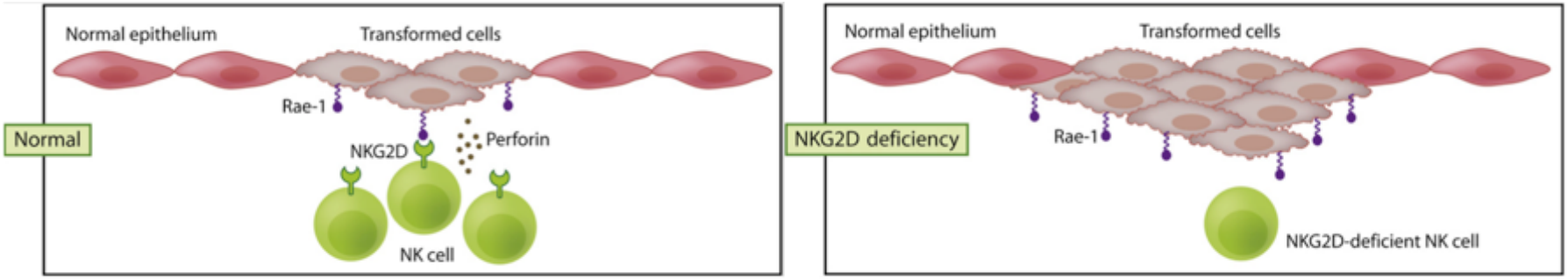
Table 3.6: Summary haplotype results in NKG2D

Legend: Hap Ref - allocated haplotype reference, Hap-Score - score statistic for association of haplotype with the binary trait, p-val - p-value for the haplotype score statistic (based on a chi-square distribution with 1 degree of freedom), control hf - estimated haplotype frequency for control group subjects, case hf - estimated haplotype frequency for case group subjects, glm.eff - the haplo.glm function modeled haplotype effects as: baseline (Base) or additive haplotype effect (Eff), OR. lower - lower limit of the Odds Ratio 95% Confidence Interval, OR - Odds Ratio based on haplo.glm model estimated coefficient for the haplotype, OR upper - Upper limit of the 95% odds ratio confidence interval.

Hap ref	Genotyped alleles contributing to haplotype							Hap score	p-val	Control hf	Case hf	glm.eff	OR lower	OR	OR upper
	rs7397310	rs10772271	rs1049172	rs11053781	rs12819494	rs2617165	rs2617167								
17	C	G	G	**C**	T	G	**G**	-1.15	0.25	0.0981	0.0706	Eff	0.43	0.76	1.3
2	C	A	A	**C**	C	G	**A**	-1.05	0.29	0.0651	0.0507	Eff	0.43	0.84	1.7
1	C	A	A	**C**	C	A	**A**	-0.93	0.35	0.1399	0.1215	Eff	0.53	0.85	1.4
10	C	G	A	**C**	C	G	**A**	-0.83	0.41	0.0349	0.0226	Eff	0.23	0.63	1.7
20	T	G	G	**C**	C	G	**G**	-0.37	0.71	0.1823	0.1739	Eff	0.64	0.96	1.4
5	C	A	A	**T**	C	G	**G**	-0.18	0.86	0.3533	0.3434	Base	NA	1	NA
7	C	A	A	**T**	T	G	**G**	1.4	0.16	0.0238	0.0386	Eff	0.81	1.87	4.3
13	C	G	A	**T**	C	G	**G**	1.61	0.11	0.0595	0.0901	Eff	0.85	1.6	3
3	C	A	A	**C**	C	G	**G**	NA	NA	0.0055	0.0034	R	1.09	1.98	3.6
4	C	A	A	**T**	C	A	**A**	NA	NA	0.012	0.0142	R	1.09	1.98	3.6

Figure 3.1: Postulated mechanism for increased tumour progression in NKG2D receptor knock out mice

Legend: NK cells lacking expression of the activating receptor NKG2D fail to recognise transformed epithelial cells expressing Rae-1 proteins. Transformation of normal epithelium and other cell types often leads to the expression of ligands for the activation receptor NKG2D. In mice, such ligands belong to a family of several Rae1 as well as H60 and MULT1 proteins. This figure illustrates the destruction of tumor cells expressing NKG2D ligands by NK cells. In the absence of the NKG2D receptor, these cells cannot be rejected by NK cells or other NKG2D-expressing immune cells. The result is aggressive tumour development[346, 352] *Adapted from Ljunggren et al, Immunity, 2008*



Overleaf:

Figure 3.2: LD Plot, *NKG2D*

Legend: SNPs genotyped in current study (blue), and prior Norwegian PSC/CC study (green) are circled (HaploView V 4.1, HapMap data V3 build R2, CEU, Chr12, 10410KB-10460KB.)

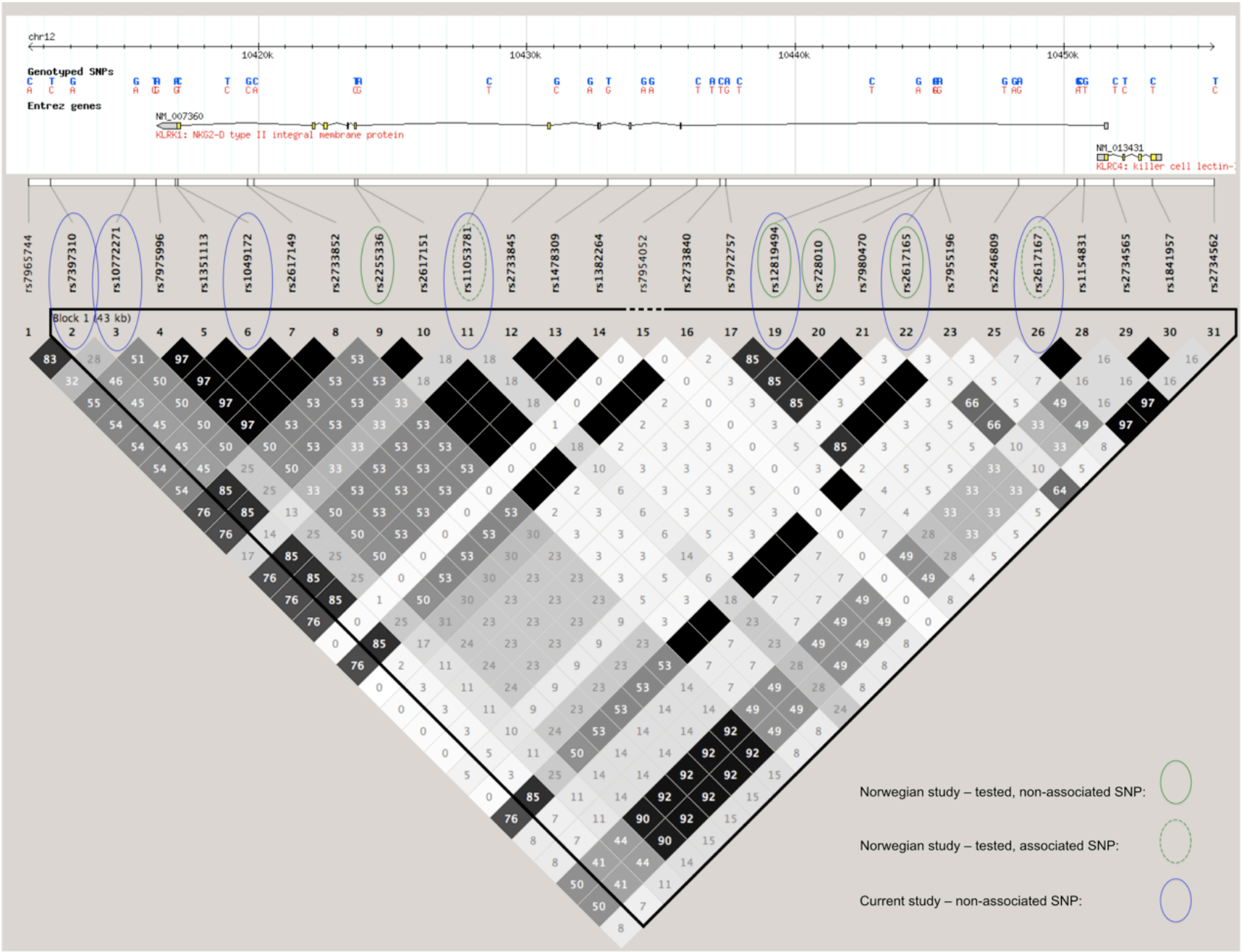

chr12
10420k
10430k
10440k
10450k
Genotyped SNPs
Entrez genes
NM_007360
KLRK1: NKG2-D type II integral membrane protein
NM_013431
KLRC4: killer cell lectin-
rs7965744
rs7397310
rs10772271
rs7975996
rs1351113
rs1049172
rs2617149
rs2733852
rs2255336
rs2617151
rs11053781
rs2733845
rs1478309
rs1382264
rs7954052
rs2733840
rs7972757
rs12819494
rs728010
rs7980470
rs2617165
rs7955196
rs2246809
rs2617167
rs1154831
rs2734565
rs1841957
rs2734562
Block 1 (43 kb)
Norwegian study – tested, non-associated SNP:
Norwegian study – tested, associated SNP:
Current study – non-associated SNP:

Figure 3.3: Haplo.stats script for NKG2D haplotype analysis

```
>library(haplo.stats)
data(haplodatankd2ccallafricanandpscexclusion)
names(haplodatankd2ccallafricanandpscexclusion)
attach(haplodatankd2ccallafricanandpscexclusion)
geno <- haplodatankd2ccallafricanandpscexclusion[,c(6:19)]
label <-c ("rs7397310", "rs10772271", "rs1049172", "rs11053781", "rs12819494", "rs2617165", "rs2617167")
geno.desc <- summaryGeno(geno, miss.val=c(0,NA))
print(geno.desc)
table(geno.desc[,3])
miss.all <- which(geno.desc[,3]==7)
haplodatankd2ccallafricanandpscexclusion.updated <- haplodatankd2ccallafricanandpscexclusion [-miss.all,]
library(haplo.stats)
data(haplodatankd2ccallafricanandpscexclusion.updated)
names(haplodatankd2ccallafricanandpscexclusion.updated)
attach(haplodatankd2ccallafricanandpscexclusion.updated)
geno <- haplodatankd2ccallafricanandpscexclusion.updated[,c(6:19)]
label <-c ("rs7397310", "rs10772271", "rs1049172", "rs11053781", "rs12819494", "rs2617165", "rs2617167")
#haplo.cc#
y.bin <- 1 * (haplodatankd2ccallafricanandpscexclusion.updated$Disease == "1")
cc.haplodatankd2ccallafricanandpscexclusion <- haplo.cc(y = y.bin, geno = geno, locus.label = label, control =
haplo.glm.control(haplo.freq.min = 0.02))
print(cc.haplodatankd2ccallafricanandpscexclusion, nlines = 25, digits = 2)
names(cc.haplodatankd2ccallafricanandpscexclusion)
```

4. Proteomic profiling of blood plasma in CC

4.1 Background

The inadequate performance of current CC tumour markers has led to investigation of alternatives. Biomarkers have been sought in CC tissue, bile and the blood. Biomarker discovery in tissue and bile may inform our knowledge of CC pathogenesis, and the search for down-stream biomarkers in the plasma or serum. However, circulating markers of disease offer the greatest clinical, diagnostic potential as the blood may be accessed safely, easily and cheaply.

4.1.1 Protein biomarkers for CC

4.1.1.1 Tissue markers

Studies of CC tissue have mainly utilised immunohistochemistry techniques and have focused on the histopathological diagnosis of CC and its differentiation from other types of malignancy. Combinations of cytokeratin stains have shown some promise in differentiating CC from metastatic cancer,[353] and in predicting poor prognosis in CC.[354] Mucin staining also has some discriminatory power, MUC1 being expressed by 70%, and MUC5AC by 50%, of CCs but not by HCC.[355] Other markers that have shown promise in differentiating CC from other malignancies include the glucose transporter GLUT-1 and p63. Those that have been associated with variation in prognosis include the adhesion molecules CD24 and CEACAM6, the cell-surface transmembrane heparin sulphate proteoglycan syndecan 1, connective tissue growth factor, and the cyclin-dependent kinase inhibitor P27Kip1.[356] Transcriptomic studies of CC tissue have demonstrated upregulation of many genes. Increased expression of survivin, p-cadherin, DNA replication complex GINS protein PSF2 and insulin like growth factor-binding proteins 1 & 2 has been shown.[357, 358]

Proteomic studies have been performed on CC tissue and cells. In one study, 2D-PAGE and mass spectrometry was used to analyse protein from a CC cell line

(HuCCA-1) in comparison with breast cancer and HCC lines. Cytokeratin 7 and 19 were only expressed by the CC cell line. Calgizzarin, exrin, moesin, radixin and hippocalin-like protein 1 were also only expressed in the HuCCA-1 cell line.[359] In another study, surface enhanced laser desorption/ionisation time of flight mass spectrometry (SELDI-TOF MS) was performed on 22 samples of CC and adjacent normal duct tissue. With this technique, mass spectra (a plot of signal intensity against mass-to-charge ratio (*m/z*)), are produced. Differentiating *m/z* peaks were identified and a panel of three of these offered sensitivity and specificity of 93% and 92% and a ROC of 0.96. The proteins represented by these peaks were not identified.[360]

4.1.1.2 Bile markers

Bile is an attractive target for biomarker discovery as all CCs are in intimate contact with the biliary tree and bile flow. However, the high lipid, bile salt and inorganic ion content of bile make analysis difficult. Collection of bile also requires ERCP, an invasive procedure with significant risks. In an early bile study, Mac-2-binding protein (Mac-2BP), in combination with CA19-9, offered increased accuracy in diagnosis of CC.[361] Biliary fibronectin is elevated in the bile of CC patients, but also in gallstone disease and PSC. A pilot study found that biliary levels of insulin-like growth factor were highly accurate in discriminating extrahepatic CC from pancreatic cancer and benign biliary abnormalities.[163] Proteomic studies have been undertaken on bile samples from CC patients. A variety of pre-fractionation strategies have been employed, including delipidation and immunoglobulin depletion of the most abundant proteins. In the first, one-dimensional sodium dodecyl sulphate polyacrylamide gel electrophoresis (1D SDS-PAGE) and liquid chromatography followed by tandem mass spectrometry (LC-MS/MS) was performed. Although only one CC sample was analysed, the detection of CEACAM1, CA125, Mac-2BP and lipocalin 2 was achieved.[361, 362] A similar, more recent study of bile from 3 patients with CC

demonstrated over expression of CEACAM6 and MUC1.[363] Recent work from the Imperial College London cholangiocarcinoma group, to which I contributed, utilised a label-free proteomic approach (GeLC-MS/MS) on bile from patients with CC. We demonstrated and then validated increased levels of neutrophil gelatinase-associated lipocalin (NGAL) in bile from patients will pancreatobiliary malignancy, including CC.[364] This is addressed in more detail in Chapter 5.

4.1.1.3 Circulating markers

Experimental circulating biomarkers have been evaluated alone and in combination with CA19-9. Other carbohydrate antigens (CA 50, 95, 242, 195, 125) individually or in panels, contribute a minimal amount to diagnostic accuracy.[161] Receptor-binding cancer antigen on SiSo cells (RCAS1), cytokeratin fragment 19 (CYFRA 21-1) and MUC5AC all show more promise, although replicative studies are required.[365-367] Interleukin 6 may improve sensitivity in combination with CA19-9, but is also elevated in benign biliary disease and HCC, reducing specificity further.[368] The same group that performed a SELDI-TOF MS proteomic study of CC tissue compared serum from 20 patients with CC with that from 20 patients with benign biliary disease and 25 healthy volunteers. Four of 31 peaks with *m/z* lower than 20kDa were differentially expressed. The most differentiating peak, of *m/z* 4462Da, was up regulated in CC.[360] The proteins were not identified and only a single type of interactive chip was used. The sample size was small and further, independent validation of these findings is outstanding.

4.1.2 Plasma proteomic profiling

The proteome is the functional protein output of the genome, including any post-translational modifications. The total protein complement of a specified fluid or sub-cellular compartment can be considered its proteome.[369] The human proteome is considered to be a rich source of potential disease biomarkers. Proteomic profiling of plasma has been used to study biomarkers in various malignancies, including CC.[163,

356] In common with other types of rare cancer, our limited knowledge of the pathogenesis of CC limits systematic investigation of candidate protein markers in specific pathways. However, proteomic techniques permit the rapid analysis of large sections of the proteome in multiple samples. A range of analytical techniques are utilised, but the same overall strategy is applied. Samples are selected to address the research question, proteins are separated to focus on a proportion of the proteome, results are analysed to identify differentially expressed proteins, an assay is developed, the potential biomarker is confirmed in further samples and, finally, validatory clinical trials are undertaken.[356] The huge number of proteins in plasma, and the massive relative abundance of a few of these, can hamper the resolution of proteomic techniques for low abundance proteins. In biomarker discovery, the target proteins are likely to be in low abundance and therefore depletion of the most abundant proteins (such as albumin) is often attempted first. Frequently, multiple protein separation techniques need to be performed serially.

4.1.2.1 GeLC-MS/MS

LC-MS/MS combines the physical separation capabilities of liquid chromatography with the accuracy of mass spectrometry and has been used widely in proteomic research.[370] Despite the excellent peptide separation capability of LC-MS/MS, a further, pre-LC, fractionation of the proteome is desirable. Simple 1D SDS-PAGE can be used to achieve this in complex protein samples. The strategy of SDS-PAGE followed by LC-MS/MS has been termed GeLC-MS/MS. A typical workflow for GeLC-MS/MS is shown in Figure 4.1. SDS-PAGE first separates proteins according to their size-dependent electrophoretic mobility (a function of the length of polypeptide chain or the molecular weight). The resulting gels are cut into 20-30 bands, which are analysed separately by LC-MS/MS. After trypsin digestion of the fractionated protein samples, the resulting peptides are separated by reverse phase liquid chromatography before undergoing tandem mass spectrometry (Illustrated in Figures

4.2 and 4.3). Once the peptide MS/MS spectra are obtained, protein identification is achieved by searching for each peptide against a reference database. To permit quantitative results, this technique sometimes utilises heavy isotope labelling of the protein samples before analysis. However, label-free approaches, avoiding this need for expensive heavy isotopes, have been developed. Label free GeLC-MS/MS has been used successfully in the detection and identification of new biomarkers in other diseases.[370, 371] The advantages of GeLC-MS/MS include the contemporaneous identification and quantification of the proteins detected, as well as relatively high sensitivity for high molecular weight proteins. However, it is a relatively labour intensive technique, with manual production of gel fragments required for tryptic digestion and LC-MS/MS. Although automated, LC-MS/MS itself, and the analysis of the data generated, is a time consuming process.

4.1.2.2 Surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS)

In this technique, the protein mixture is applied directly to a chip, the surface of which is modified with a chemical functionality. Some proteins in the sample bind to the surface, while the others are removed by washing. This equates to a highly effective initial fractionation step, as those hugely abundant proteins that are not bound are almost entirely excluded from subsequent analytical steps. The subset of proteins that bind to the specific surface are easier to analyse and resolve in the spectra generated. Common chip surfaces include CM10 (weak-positive ion exchange), H50 (hydrophobic surface, similar to C6-C12 reverse phase chromatography), IMAC30 (metal-binding surface), and Q10 (strong anion exchanger). The range of different chip surface properties is illustrated in Figure 4.4. The commonest surfaces used to most success in biomarker discovery studies are CM10 and H50, the biochemical function of which is illustrated in Figure 4.5. An energy absorbing matrix (EAM) solution is applied to the chip surface, which releases the bound protein component and co-crystallizes with it as it dries. Proteins retained within the matrix on the SELDI

chip surface are analysed using time-of-flight mass spectrometry. A laser desorbs and ionizes the proteins from the crystals of the sample/matrix mixture. The ions are accelerated through an electric potential and fly down a field-free flight tube. A detector measures ions as they reach the end of the tube (Figure 4.6). The mass-to-charge ratio of each ion can be determined from i) the kinetic energy given to ions by the electric field, ii) the length of the tube and iii) the time taken to travel the length of the tube. SELDI-TOF MS has been used to successfully identify malignant protein patterns in plasma from ovarian, hepatocellular, pancreatic, prostate and breast carcinomas.[371, 372] SELDI-TOF MS does not require initial time-consuming manual separation stages. It is relatively sensitive in the low molecular weight range and the instrument itself processes samples quickly and is highly automated. There is no significant obstacle to analysing large numbers of samples on this system. SELDI-TOF MS does not identify proteins from the peaks generated, and is less sensitive for higher mass peptides.

4.2 Hypothesis and study aim

4.2.1 Hypothesis

Plasma proteomic profiles can differentiate patients with cholangiocarcinoma from those with benign biliary disease, and from healthy control subjects.

4.2.2 Aims

1) To compare the plasma proteomic profile of patients with CC to that of patients with benign disease and healthy controls utilising:

i) GeLC-MS/MS to compare plasma protein profiles of patients with CC to those with gallstone disease of the bile duct

ii) SELDI-TOF MS to compare plasma protein profiles of patients with CC to those of PSC patients and those of healthy controls

2) To identify specific protein peaks for further investigation as potential clinical biomarkers of CC

4.3 Materials and Methods

4.3.1 Ethics

The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki. Approval from the local Research Ethics Committee and NHS R&D Department was sought and obtained for these studies (Ref 09/H0712/82). All participants provided written, informed consent.

4.3.2 Subjects

4.3.2.1 CC patients

I recruited subjects with CC from the gastroenterology clinics and ERCP lists at Imperial College London Healthcare NHS Trust (ICH) and University College Hospitals NHS Trust (UCH). CC patients were considered for recruitment if their diagnosis was secured on the basis of a) pre- or post-operative histology or b) multidisciplinary team consensus on the basis of ≥2 imaging modalities, clinical course and plasma markers.

4.3.2.2 PSC and gallstone disease patients

I recruited PSC and gallstone disease patients in a similar fashion, from specialist hepatology clinics at ICH and UCH. The diagnosis of gallstone disease was made on the basis of clinical course, imaging findings and liver function tests. The diagnosis of PSC was determined on the basis of clinical course, prior imaging, consistent autoantibody titres and, where available, biopsy results. Patients with advanced cirrhosis (Child-Pugh B or C, according to standard criteria, Table 4.1) were excluded, as were patients who had undergone previous liver transplant. Patients currently being investigated for suspected cholangiocarcinoma were excluded. However, PSC patients with dominant strictures that had been fully investigated and characterised were included. To reach an adequate cohort size, I approached

collaborators at the Norwegian PSC Research Center (NoPSC), Rikshospitalet, Oslo, Norway. They were able to provide plasma samples from well-characterised PSC patients, with appropriate on-going consent in place for use in studies of liver disease.

4.3.2.3 Healthy controls

Healthy volunteers were recruited at ICH and UCH. After completing a short, confidential health questionnaire, individuals with no current inflammatory or malignant disease, or liver disease of any sort, were recruited.

4.3.3 Sample collection and processing

The plasma proteome is vulnerable to modification during sample collection and pre-processing. For these proteomic studies, a total of 8mL of blood was collected into EDTA tubes, placed onto ice immediately and centrifuged within one hour. 200µL aliquots were placed into Eppendorf tubes (Eppendorph, Histon, UK) and frozen at -80^{o}C. Where samples were collected prospectively, I used my sample collection standard operating procedure (SOP) (Appendix 1).

I collected demographic data, including age, sex, ethnicity, and short medical history onto my case record form (CRF, Appendix 2). Where available, the most recent routine clinical laboratory indices were also documented, including white cell count, C-reactive protein (CRP), CA19-9, alanine transferase (ALT), albumin, alkaline phosphatase (ALP), bilirubin, urea and creatinine. As healthy control volunteers were not undergoing routine laboratory blood testing, these parameters could not be collected for those subjects. Where available for CC cases, I collated imaging information to determine the anatomical location of the CC.

Plasma samples from the NoPSC archive had been collected and processed according to a strict protocol that did not diverge from my SOP significantly.

4.3.4 Method, GeLC-MS/MS study

All reagents used were from Sigma-Aldrich (Gillingham, Dorset, UK), unless stated otherwise.

4.3.4.1 Subjects included

Plasma samples from eight CC patients and seven gallstone disease patients were included. All 15 of these samples were collected at ICH.

4.3.4.2 Protein depletion

I first carried out depletion of the high abundance plasma proteins in each sample, using a Plasma IgY partitioning column (Mixed Seppro, Genway Biotech, US). This process depletes the 12 most abundant plasma proteins, including fibrinogen, IgG, transferrin and albumin. The workflow undertaken is detailed here and illustrated in Figure 4.7.

i) 490 µL of dilution buffer (100mM ammonium formate, pH 7.4) and 10 µL of plasma were applied to the Ig-Y spin column

ii) The column was incubated on a rotor for 15 min at 4°C

iii) The column was centrifuged at 400g for 30s and the eluate collected and retained

iv) Steps (i) to (iii) were repeated with a further 500µl of dilution buffer

v) 500µl of stripping buffer (1M glycine, pH 2.5) was applied to the column and incubated with mixing on a rotor for 3 min

vi) The column was centrifuged at 400g for 30s and this eluate (containing the unwanted, high abundance 12 proteins) was discarded

vii) The IgY column was immediately neutralised with 600µL of 1x neutralizing buffer (1M tris-HCl, pH 8.0)

viii) Steps i-vii were repeated for each sample

ix) The retained eluate from steps iii and iv was fully dehydrated in a vacuum centrifuge and frozen at -20^{o}C pending SDS-PAGE

4.3.4.3 1D SDS-PAGE

1D SDS-PAGE was the second protein separation technique that I applied to each plasma sample.

4.3.4.3.1 1D SDS PAGE methodology

i) Each of the depleted, dehydrated protein samples were rehydrated in 13µL of H_2O

ii) Sample proteins were then denatured by addition of 5µL of 4x lithium dodecyl sulphate (LDS) buffer (Invitrogen, Paisley, UK)

iii) The samples were then reduced by adding 2µL of 0.5M DTT and heating at 100^{o}C for 2 min

iv) Samples were cooled and further diluted with 8µL H_20 and 52µL of 4x LDS buffer

v) Samples were then alkylated with 5µL of 1M iodoacetamide (IAA) in the dark for 30 min, resulting in a total volume of 85µL

vi) 20µl of each sample was then applied to Novex 10% bis-tris gels (Invitrogen) in 2-(*N*-morpholino)ethanesulfonic acid (MES) buffer with 500µL NuPAGE antioxidant (Invitrogen) in the inner chamber

vii) PageRuler pre-stained protein ladder marker mix (Fermentas, St. Leon-Rot, Germany) was placed in lanes 1 & 10 of each gel, plasma samples in lanes 2-9

viii) Gels were subject to electrophoresis at 200V for 30min

ix) Gels were stained with Coomassie brilliant blue (CBB) (InstantBlue, Novexin, UK) overnight

4.3.4.3.2 Confirmation of protein depletion

To confirm efficacity of protein depletion, I initially performed 1D SDS PAGE on samples from 4 subjects: benign samples 1 & 2 and CC samples 1 & 2. A pre-depletion 'raw' plasma sample and a post-depletion sample from each of these 4 subjects were included on a single gel. The depleted, dehydrated sample was reconstituted and prepared to a total volume of 85µL using the steps detailed in 4.3.4.3.1, above. Samples of equivalent volume and concentration of 'raw' plasma were achieved by adding only 3µL of H_2O to 10µL of raw plasma at step (i). Corresponding raw and depleted samples from each of the 4 subjects were run on a gel in lanes 2-9, according to the parameters in 4.3.4.3.1, above.

4.3.4.3.3 Preparation of gels for LC-MS/MS study

Once protein depletion had been confirmed, I prepared two gels with the 8 CC samples and 7 benign samples.

i) Samples were prepared as per 4.3.4.3.1, above, and were applied to the two gels according to the template shown in Figure 4.8. The gels were then run according to the parameters outlined in 4.3.4.3.1

ii) After completion of 1D SDS-PAGE, all gel lanes were sliced into equivalent regions, illustrated in Figure 4.9. Gel regions were defined by the protein standard included in lanes 1 & 10, with each gel slice having the same mass

iii) Gel slices were kept in dry conditions in a multiwell plate overnight

iv) Slices were destained with 1mL 100mM ammonium bicarbonate in 50% ACN overnight

v) Each slice was washed with 1mL of water and dehydrated in 1mL ACN overnight

vi) Gel slices were transferred to low-binding Eppendorf tubes and kept chilled pending tryptic digestion

vii) The proteins in each slice were digested in 0.3mL 50mM ammonium bicarbonate containing 1ng/µl trypsin at 37^{o}C overnight

viii) After digestion, trypic peptides were extracted with 0.3mL 0.1% formic acid in 2% ACN

ix) 400µl of the supernatant was dried and then dissolved in 30µl of water containing 0.1% TFA

4.3.4.4 Liquid chromatography MS/MS

Samples extracted from equivalent gels regions from all 15 samples were batched for LC-MS/MS and each batch required *circa* 20h to process through the LC-MS/MS system. An Agilent LC series system was used (Agilent Technologies UK Ltd, Berkshire, UK). 8µL of each peptide solution was applied to ProteCol traps (SGE Ltd, Milton Keynes, UK) and separated by nano-LC using 75um ID x 15 cm C18 PepMap columns (Dionex, Leeds, UK) at a flow rate of 250nL/min over a linear gradient of 2–45% ACN (LC-MS grade, Sigma) containing 0.1% formic acid (Sigma) for 30 min. MS analyses of eluted peptides was performed on-line using an LTQ linear ion trap MS (Thermo) connected to the nano-LC effluent through an electrospray interface. Dynamic activation time was set at 30s with scans every 2µs. Data-dependent MS/MS spectra were obtained.

4.3.4.5 Protein identification

Progenesis software (Nonlinear Dynamics Ltd, Newcastle, UK) was used to match LC and *m/z* data between equivalent gels regions from each subject. SEQUEST software was used to match peaks from each tandem mass spectrum to the NCBI human RefSeq peptide database. Peptides with acceptable tandem mass spectra quality were included in further analysis (Xcorr values >1.5 for single charged, >2.0 for double charged and >2.5 for triple charged peptides). Proteins that were identified

by ≥2 unique peptides were included in further analysis. Where proteins were identified in more than one gel region, the region in which the most unique associated peptides were identified was used. Finally, Progenesis software (Nonlinear Dynamics) was used to compare normalised intensities of component peptide ions of interest and thereby quantify the protein concentration in each gel region.

4.3.4.6 Statistical analysis, LC-MS/MS study

Statistical analyses were performed using Excel (Microsoft), Prism Version 5 (Graphpad Software, San Diego, USA) and SYSTAT (Chicago, US). In calculating a normalised abundance for each protein in each gel region, Progenesis software log transforms the abundance of all component peptides before summing. Student's t-test (on arcsin transformed data) and linear regression analysis were used to identify differentially expressed proteins. Proteins were considered to be differentially expressed when abundance was ≥1.5 fold different between groups and Student's t-test p-value was <0.05. ROC-AUC analysis was then performed on any protein that was differentially expressed between the two groups. Finally, a general discriminant function analysis was undertaken to assess the discriminatory utility of combinations of the putative protein markers.

4.3.4.7 Further investigation of identified proteins

Any candidate proteins identified were investigated further by searching in the Universal Protein Resource Unit Knowledge Base (UnitProtKB) database (v61, European Bioinformatics Institute, Hinxton, UK) and the PubMed (National Center for Biotechnology Information, U.S. National Library of Medicine, Bethesda, USA) protein database.

4.3.5 Method, SELDI-TOF MS studies

All reagents were sourced from Sigma-Aldrich, unless stated otherwise. This study comprised three phases:

i. Preliminary optimisation studies – identifying the conditions under which the best resolution and magnitude of protein *m/z* peaks could be obtained

ii. Discovery study – performing SELDI-TOF MS under a shortlist of chip conditions to identify any discriminatory *m/z* peaks of interest and determine the best conditions under which to conduct a validation study

iii. Validation study – performing SELDI-TOF MS on further, larger cohorts of CC, PSC and healthy controls, but under a reduced number of conditions, to validate the *m/z* peak findings from the discovery SELDI-TOF MS study

4.3.5.1 Subjects included

In total, plasma samples were included from 99 CC patients, 64 PSC patients and 107 healthy controls. In the preliminary optimisation studies, a single healthy control sample, 'N', was used. In the initial discovery SELDI-TOF MS experiment 18 CC, 10 PSC and 18 healthy control samples were used. In the final, validation, experiment 81 CC, 54 PSC and 90 healthy control samples were used.

4.3.5.2 Selection of chip surface

As discussed in Section 4.1, a wide range of SELDI chip surface properties are available. However, the chips are single use and are expensive to purchase. Given my budget and the number of samples I intended to investigate, it was necessary to narrow the number of different surfaces to two. Previous work by our group, and work published by others, has found that H50 and CM10 chips have yielded the best number, and resolution, of peaks when investigating crude plasma with a hypothesis free, non-biased approach. I therefore elected to use CM10 and H50 chips in my experiments.

4.3.5.3 Preparation of samples

Each plasma sample was denatured and diluted using the following protocol:

i) 10µL of 500mM DTT was added to 1mL of 9M urea/2% CHAPS to make a 9M urea/2%CHAPS/5mM DDT solution

ii) 20µL of each plasma sample was added to 180µL of the solution from step (i) to achieve a 1:10 dilution

iii) Samples were left at room temperature for 30 min to denature

4.3.5.4 Preparation of buffers and binding solutions

A range of buffers and binding solutions was required for initial condition optimisation, my discovery SELDI-TOF MS study and validation SELDI-TOF MS study. I made up adequate volumes of each buffer and binding solution to complete all the anticipated experiments, with a margin for error. Based on previous successful work within my group, and that reported by other groups, I prepared the following buffers and binding solutions.

4.3.5.4.1 pH buffers for CM10 chip conditioning

All buffers were made up to 100mM. I prepared and experimented with the following buffer solutions on CM10 chips:

i. pH3: 0.1M glycine and hydrochloric acid

ii. pH4: 0.1M sodium acetate solution and 100% acetic acid

0.1M ammonium acetate and 100% acetic acid

iii. pH6: sodium phosphate buffer

0.1M citric acid and sodium hydroxide

iv. pH7: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)

v. pH8: Tris (hydroxylmethyl) aminomethane and hydrochloric acid

4.3.5.4.2 Binding solutions for H50 chip conditioning

I prepared and experimented with the following acetonitrile (ACN) binding solutions on H50 chips:

i. 10% ACN with 0.1% TFA

ii. 20% ACN with 0.1% TFA

iii. 20% ACN with 0.1% TFA plus 100mM NaCl

iv. 50% ACN with 0.1% TFA

4.3.5.5 Preparation of energy absorbing matrices (EAMs)

4.3.5.5.1 Sinapinic acid (SPA)

A saturated solution of SPA in 50% acetonitrile, 49.50% HPLC grade H_2O and 0.50% trifluoroacetic acid (TFA) was made up. The saturated solution was centrifuged briefly, to remove any non-dissolved SPA particles, and the supernatant extracted for use.

4.3.5.5.2 Alpha-cyano-4-hydroxycinnamic acid (CHCA)

A saturated solution of CHCA was made up in 50% acetonitrile, 49.75% HPLC grade H_2O and 0.25% TFA. The saturated solution was centrifuged briefly, to remove any non-dissolved CHCA particles, and the supernatant extracted for use.

4.3.5.6 Application of samples to SELDI chips

Each SELDI chip array can accommodate up to 12 chips (1-12) with 8 spots (A-H) per chip giving a total of 96 spots per array.

4.3.5.6.1 CM10

i) An appropriate number of CM10 chips were assembled within the SELDI bioprocessor, forming a well over each spot

ii) 200µL of appropriate buffer was applied to each well/spot and left on the shaker for 5 min

iii) The buffer was discarded from each well

iv) Steps (ii)-(iii) were repeated, to complete conditioning of spots

v) A further 135µL of buffer was applied to each well, with addition of 15µL of sample, prepared in 4.3.4.5.1, above

vi) The samples were covered with an adhesive acetate sheet and incubated on the shaker at room temperature for 60 min

vii) The buffer/sample mix was discarded from each well

viii) 200µL of buffer was applied to each well, left on the shaker for 5 min, before discarding. Repeated 2 more times

ix) 200µL deionised water was applied to each well, left on the shaker for 5 min, before discarding. Repeated 1 more time

x) The bioprocessor was disassembled and the spots on the array allowed to air-dry for 10 min

xi) 0.5µL of saturated EAM (SPA or CHCA) was applied to each spot and allowed to dry. Repeated 1 more time

xii) The samples were then processed through the SELDI instrument

4.3.5.6.2 H50 chips

i) An appropriate number of H50 chips were assembled within the SELDI bioprocessor, forming a well over each spot

ii) To pre-condition the chips, 100µL of 50% acetonitrile (ACN) solution was applied to each spot and left for 5 min. This step was repeated once

iii) 200µL of appropriate ACN binding solution was applied to each well/spot and left on the shaker for 5 min

iv) The buffer was discarded from each well

v) Steps (iii)-(iv) were repeated, to complete conditioning of spots

vi) A further 135µL of ACN binding solution was applied to each well, with addition of 15µL of sample, as prepared in 4.3.4.5.1, above

vii) The samples were covered with an adhesive acetate sheet and incubated on the shaker at room temperature for 60mins

viii) The buffer/sample mix was discarded from each well

ix) 200µL of ACN binding solution was applied to each well and left on the shaker for 5 min, before discarding. Repeated 2 more times

x) 200µL deionised water was applied to each well, left on the shaker for 5 min, before discarding. Repeated 1 more time

xi) The bioprocessor was disassembled and the spots on the array allowed to air-dry for 10 min

xii) 0.5µL of saturated EAM (SPA or CHCA) was applied to each spot and allowed to dry. Repeated 1 more time

xiii) The samples were then processed through the SELDI instrument

4.3.5.7 Protein and peptide standard

I made up a protein and peptide standard solution that was included on two spots within each array of chips, to permit calibration of the SELDI-TOF MS instrument. The constituent proteins and peptides of the protein/peptide standard are listed in Table 4.2. In addition, a standard healthy control sample ('N') was applied to 2 spots on each array to confirm inter-assay consistency.

4.3.5.8 SELDI-TOF MS

All chips were analysed by SELDI-TOF MS using a Protein Biology System IIc Reader (Ciphergen Biosystems Inc., Freemont, CA, USA). Chips were processed through the SELDI-TOF MS device under the same settings (Table 4.3).

4.3.5.9 Preliminary study to select optimum chip conditions

It was not considered necessary, or good practice, to run all samples under all possible conditions. Initially, I investigated a single, standard sample ('N') under numerous conditions. I selected these initial chip/condition combinations based upon previous experience of successful SELDI-TOF MS experiments on plasma in my group. As discussed in Section 4.1, above, the chemical and physical conditions at the chip surface can vary the milieu of bound protein dramatically. The selection of

conditions under which the samples are applied to the SELDI chip is therefore crucial, as this fundamentally affects the results produced. The CM10 chip surface includes negatively charged carboxylate chemistry and therefore acts as a weak cationic exchanger and binds proteins that are positively charged at the prevailing pH (Figure 4.5A). To modulate the sensitivity and range of proteins binding, the pH of the buffer may can be increased or decreased. With decreased pH, a greater number of proteins within the sample become positively charged, increasing protein binding. With increased pH, a greater number of proteins become negatively charged, resulting in less binding (and therefore greater specificity). In the case of H50, the concentration of ACN applied can be modulated. Only the most hydrophobic proteins will be retained with wash buffers containing a high organic solvent and so increasing the concentration of organic solvent in the binding solution will increase the selectivity of the surface (Figure 4.5B). The SELDI chip manufacturer also suggests that to increase the salt concentration, and thereby the hydrophobic interactions on H50 chips, additional sodium chloride can be included in the binding solution. Finally, it is possible to modify the energy-absorbing matrix (EAM) that co-crystallises with the proteins applied to each spot. Varying the EAM used on the SELDI chip modifies the spectra achieved. Generally, CHCA is considered best for small proteins and peptides (1-30kDa). SPA is considered to have broader utility and permits detection of larger proteins as well as many peptides (10-150kDa).

4.3.5.9.1 Selection of buffer solution and EAM for CM10 chips

The standard control sample ‘N’ was applied to multiple spots on CM10 chips according to the protocol detailed in 4.3.5.6.1. The seven buffers listed in 4.3.5.4.1 were each used in the preparation of four spots, two of which were treated with SPA and two with CHCA as the EAM (i.e. 7x4=28 spots). The chip, buffer and EAM combinations trialled are detailed in Table 4.7. Chips were subjected to SELDI-TOF MS as per the protocols and settings detailed in 4.3.5.8. Initial data management and

analysis was with the Biomarker Wizard function that is integrated into the SELDI-TOF MS control software (Ciphergen). The Biomarker Wizard generates the spectra, identifies and matches peaks between spectra and assigns intensity values to each peak. Spectra were inspected manually for number of peaks, amplitude of peaks, SNR and resolution of peaks.

4.3.5.9.2 Selection of binding solution and EAM for H50 chips

Standard sample 'N' was applied to multiple spots on H50 chips according to the protocols detailed in 4.3.5.6.2. The four different binding solutions detailed in 4.3.5.4.2 were each used in the preparation of four spots, two of which were treated with SPA and two with CHCA as the EAM (i.e. 4x4=16 spots). The chip, binding solution and EAM combinations trialled are detailed in Table 4.7. Chips were subjected to SELDI-TOF MS as per the protocols and settings detailed in 4.3.5.8, above. The spectra generated were inspected manually and analysed within the Biomarker Wizard software for number of peaks, amplitude of peaks, SNR and resolution of peaks.

4.3.5.10 Discovery SELDI-TOF MS study

4.3.5.10.1 Subjects included, discovery study

Three cohorts were analysed and compared. Plasma from 18 CC subjects, 10 PSC subjects and 18 healthy control subjects were included in the discovery study.

4.3.5.10.2 Chips and conditions used, discovery study

The selection of the chip conditions used was informed by the results of the preliminary studies, 4.3.5.9, above. The following 4 chip and condition combinations were used:

i. CM10 chip, pH 4 sodium acetate buffer, SPA EAM

ii. CM10 chip, pH 6 sodium phosphate buffer, SPA EAM

iii. CM10 chip, pH8 Tris buffer, SPA EAM

iv. H50 chip, 10% ACN binding solution, CHCA EAM

4.3.5.10.3 SELDI-TOF MS, discovery study

All samples were prepared from the raw plasma aliquots in one batch, according to the protocol in 4.3.5.3. Samples, the protein standard solution and the 'N' standard plasma samples were all applied to two spots each on a 96 spot chip array, according to the protocol in 4.3.5.6. The various combinations of chip and condition used are detailed in 4.3.10.2. SELDI-TOF MS was then undertaken on each sample, using the protocol and settings in 4.3.5.8.

4.3.5.10.4 Statistical analyses, discovery study

The Biomarker Wizard software was used to auto-detect all peaks within the specified m/z range and with a signal to noise ratio (SNR) of ≥3. Peaks were aligned within 0.3% of mass error. Data were further manipulated in MS Excel and statistical analyses were undertaken in STATISTICA (StatSoft Inc., Tulsa, USA). Peak intensities at each *m/z* point were averaged between the paired spots for each sample. Where any spot failed to generate a spectrum, the spectrum from its duplicate was used in isolation. Where neither paired spots generated spectra, the subject was excluded from further analysis. After testing whether peak intensities were normally distributed, values were subjected to either Mann-Whitney U test (not normally distributed) or Student's t-test (normally distributed) to compare groups. Initially, CC subjects were compared to PSC subjects; CC subjects were then compared to healthy control subjects. P-values of <0.05 were considered significant. Those peaks identified as significantly different were examined manually in the Biomarker Wizard (Ciphergen). Those peaks that were found to be misaligned, those shouldered on a dominant adjacent peak and those that had been misclassified as a peak rather than noise were manually excluded from further analysis. The two combinations of chip and conditions that produced the most discriminatory peaks in

this discovery SELDI-TOF MS study were determined. These were then used in the subsequent, validation SELDI-TOF MS study.

4.3.5.11 Validation study, SELDI-TOF MS

4.3.5.11.1 Subjects included, validation study

Three cohorts were investigated, to validate the peaks of interest identified in the preceding discovery study. Samples from 81 CC subjects, 54 PSC subjects and 90 healthy control samples were subjected to SELDI-TOF MS analysis.

In parallel, I subjected fresh samples from the original discovery cohorts (18 CC, 10 PSC, 17 healthy controls) to repeat SELDI-TOF MS analysis. This was to confirm reproducibility of the findings.

4.3.5.11.2 Chip conditions used, validation study

The selection of chip conditions for the validation study was informed by the results of the discovery study. Two chip/condition combinations were selected:

i. CM10 chips, Tris buffer - pH 8, SPA as EAM

ii. H50 chips, ACN 10% solution binding solution, CHCA as EAM

4.3.5.11.3 SELDI-TOF MS

All samples were freshly sourced and processed from the primary raw plasma aliquots according to the protocol in 4.3.5.3. Samples, the protein standard solution and the 'N' standard plasma samples were all applied to two spots each on a 96 spot chip array, according to the protocol in 4.3.5.6. The two combinations of chip and condition used are detailed in 4.3.5.11.2, above. Samples from all cohorts were applied to randomly allocated spots across all arrays. The two spots used for each subject were always applied to separate chips. SELDI-TOF MS was then undertaken on each sample, using the protocol and settings in 4.3.5.8.

4.3.5.11.4 Statistical analyses, validation study

4.3.5.11.4.1 Initial analysis, validation samples only

Initial analysis, using the Biomarker Wizard, was as in section 4.3.5.10.4. *M/z* peaks corresponding to the seven identified as of interest in the discovery study were matched according to *m/z* (within 0.15% of each other) and peak characteristics. Once the seven peaks had been identified, the groups were compared. P-values of <0.05 were considered significant. Initially, the new validation cohorts were studied, seeking to validate the findings from the original discovery study. CC subjects were first compared to PSC subjects, CC subjects were then compared to healthy controls and PSC subjects were then compared to healthy controls. Where possible, subgroup analyses of subjects recruited at ICH and UCH were undertaken, to investigate reproducibility of the findings on both sites. Peak intensity values were compared between the UK PSC cohort and Norwegian PSC cohort, to identify any geographical variation.

4.3.5.11.4.2 Repeat analysis of samples from discovery study

Subsequently, the new dataset on the original discovery cohort was analysed, seeking to demonstrate reproducibility of the original significant peaks in the same subjects; CC subjects were first compared to PSC subjects; CC subjects were then compared to healthy control subjects.

4.3.5.11.4.3 Combination of cohorts

Only those peaks identified in the discovery study and confirmed as significantly different in the validation study were subject to further analysis. To maximise cohort size, this was performed on the pooled cohorts, i.e. 99 CC samples, 64 PSC and 107 healthy controls. Only data from the validation study and the parallel *repeat* SELDI-TOF MS analysis of the discovery samples were used. Demographic data on the combined cohort were collated and descriptive statistics were generated. Sex, age and ethnicity were compared between groups using χ^2 testing or Kruskal-Wallace ANOVA as appropriate.

4.3.5.11.4.4 Anatomical subtype of CC

The CC cohort was further characterised by collating details of anatomical subtype of the primary malignancy (i.e. intrahepatic, hilar or distal CC). Peak intensities were compared between these subtypes using Kruskal-Wallace ANOVA (GraphPad Prism).

4.3.5.11.4.5 Routine lab parameters

Where available, routine laboratory test results of CC and PSC patients were collated for the combined cohort. These parameters were not available for healthy controls. Descriptive statistics on each parameter were generated for each cohort. CC and PSC cohorts were compared using Mann-Whitney U test (GraphPad Prism).

4.3.5.11.4.6 ROC-AUC analysis

ROC-AUC analysis was performed for each of the validated peaks (GraphPad Prism). Diagnostic utility in discriminating CC from PSC, and then CC from healthy controls, was tested.

4.3.5.11.4.7 Correlation matrix

A correlation matrix of all validated peaks was generated, to identify any potential correlation between each of them (STATISTICA).

4.3.5.11.4.8 Correlation of discriminatory peaks with routine lab parameters

Where they were known, routine laboratory parameters from all subjects were correlated with the intensity of the seven peaks of interest. Pearson's correlation was used (STATISTICA).

4.3.5.11.4.9 Multiple linear regression (general discriminant analysis)

Multiple linear regression (MLR) was used to perform a general discriminant analysis to identify the peaks generated under both chip conditions that maximally discriminated the study groups with f to include set at 3.5, and f to exclude to 3.4 and maximum steps to 100 (STATISTICA). CC and PSC groups were analysed, then CC and healthy controls. ROC-AUC analysis was then used to assess the performance of these models.

4.3.5.11.4.10 Combination of peaks with CA19-9

MLR was then used to determine whether the addition of CA199 would enhance the model or displace any of the previously included peaks. Again, ROC-AUC analysis was performed to assess performance.

4.3.5.11.4.11 Potential confounding factors analysis

MLR was repeated, offering only the seven peaks for inclusion but sequentially weighting for each of the following potential confounding factors: age, sex, ethnicity and site of collection. The following laboratory indices were also corrected for: ALT, albumin, CRP and CA19-9. Again, ROC-AUC analysis was performed to assess performance.

*4.3.5.11.4.12 Post hoc analysis of PSC **and** CC group*

Although none of the 99 CC patients in the discovery and validation cohorts had PSC **and** CC, I subjected samples from six such subjects to SELDI-TOF MS in parallel to the validation cohort. Mann-Whitney U test was used to simply analyse any peaks that showed utility in discriminating CC from PSC in the larger study, comparing these six subjects with the 64 PSC subjects.

4.3.5.11.5 Database search for m/z peak matches

All *m/z* peaks confirmed to be of interest in the validation study were researched further. The online TagIdent tool (ExPASy, Swiss Institute of Bioinformatics, Geneva Switzerland) was used to search the UniProtKB database v61 (NCBI). The search was restricted by species of origin *homo sapiens.* All previously registered proteins within 0.1% of the *m/z* of the validated peaks of interest were sought. Each of these proteins was then explored manually in the UniProtKB database.

4.4 Results

4.4.1 Results of GeLC-MS/MS study

4.4.1.1 Demographics of subjects

Eight patients were included in the CC group, five were male (63%) and the median age was 60y. Four members of the CC cohort were Caucasian (50%), with four of other ethnicity. Seven patients were included in the control group, two were male (39%) and the median age was 55y. All benign control subjects were Caucasian. All benign control cases had uncomplicated gallstone disease. There was no significant difference in age, sex or ethnicity between the two groups. Further details are included in Tables 4.4 and 4.5.

4.4.1.2 Routine laboratory markers

Routine laboratory indices were collected for each patient. Individual and median values are detailed in Table 4.6. Tumour marker values were generally not available for gall stone disease cases. No blood test results were available for one of the benign control cases. There were no significant differences between CC and benign control groups in any of the laboratory indices collected (p-values 0.3-0.91).

4.4.1.3 Confirmation of depletion

The gel prepared and run to confirm efficacity of the IgY column protein depletion step (Figure 4.10) demonstrated clearly that the most abundant proteins (e.g. albumin, arrowed) was partially depleted from the sample. However, it was also clear that a substantial fraction of such proteins remained after the depletion step.

4.4.1.4 Proteins identified and differentially expressed

128 different non-redundant proteins were identified from the peptides detected in the samples analysed. These are listed in Table 4.8.

Student's t-test of the logarithmically transformed normalised abundances identified six proteins that were differentially expressed between CC and the gallstone disease groups. These are highlighted in bold text in Table 4.8 and summarised in Table 4.9.

One of the six was found to be in reduced abundance in CC, five were at increased abundance. ROC-AUC analysis demonstrated good diagnostic performance in all six of these individual proteins, with ROC-AUCs of between 0.87 and 0.98 (Table 4.10). After three steps, multiple linear regression included two of these proteins in the optimum model generated. No benefit was achieved by including the other putative protein markers. This model achieved a ROC-AUC of 0.998 (95% CI 0.99-1.0).

4.4.1.5 Further investigation of identified proteins

Each of the six candidate proteins was investigated further by searching in the UniProtKB database (v61) and PubMed search. Where known, functions and disease associations of the identified proteins are summarised in Table 4.11.

4.4.2 Results of SELDI-TOF MS studies

4.4.2.1 Results of preliminary optimisation studies

This investigation of chip, buffer and EAM combinations resulted in a short list of 4 chip/condition combinations for the subsequent discovery SELDI-TOF MS experiment.

4.4.2.1.1 Selection of CM10 chip buffer and EAM

Several buffers produced no or attenuated spectra and were therefore not utilised in subsequent experiments. Glycine and citric acid based buffers (at pH3 and pH6) produced no spectral peaks al all. HEPES (pH7) and ammonium acetate (pH 4) buffers produced some spectral signal, but peaks were markedly less abundant and intense than the three best performing buffers. Three buffers, sodium acetate (pH 4), sodium phosphate (pH 6) and Tris (pH8) generated the best spectra (Figure 4.11). At low to mid mass range, SPA produced the highest resolution of peaks (Figure 4.11A). High mass range detection was non-existent with CHCA (spectra not presented). With SPA, multiple peaks were detectable in the high mass range at all three pH conditions (illustrative spectra can be found in Figure 4.11B). Therefore, the

following three CM10 chip/condition combinations were selected for the discovery SELDI-TOF MS experiment:

i. CM10 chip, pH 4 buffer sodium acetate buffer, SPA EAM

ii. CM10 chip, pH 6 sodium phosphate buffer, SPA EAM

iii. CM10 chip, pH8 Tris buffer, SPA EAM

4.4.2.1.2 Selection of H50 chip binding solution and EAM

10% ACN with 0.1% TFA resulted in the most peaks with the greatest amplitude (illustrative spectra can be found in Figure 4.12). Increasing the concentration of ACN reduced spectral resolution and peak intensity. Those spots that were treated with the NaCl supplemented ACN binding solution generated no spectra whatsoever. At low mass ranges, CHCA EAM outperformed SPA (Figure 4.12A). High mass range peak detection was almost non-existent with both EAMs (Figure 4.12B). The fourth and final chip/condition combination selected for inclusion in the discovery experiment was therefore:

iv. H50 chip, 10% ACN binding solution, CHCA EAM

4.4.2.2 Results of discovery SELDI-TOF MS study

I performed this experiment on three cohorts: samples from 18 CC subjects, 10 PSC subjects and 18 healthy control samples.

4.4.2.2.1 Demographics of subjects, discovery study

Of the 18 subjects with CC, nine were collected at ICH and nine from UCH. All ten of the samples from PSC patients were from NoPSC Biobank, Norway. Of the 17 healthy controls analysed, 9 were from ICH and 8 from UCH. There were significant differences in the male to female ratio between the three groups. There was a significant difference between the age of the CC cohort (median age 68y) and the age of the PSC (39y) and HC (42y) cohorts. The majority of subjects in all three groups were Caucasian, however there were significantly more non-Caucasians in

the CC (17%) and healthy control groups (16%) than in the PSC groups (0%). More detailed demographic data is presented in Table 4.12.

4.4.2.2.2 CM10 chips, discovery study

The SP060 sample failed on both spots under all three conditions and was excluded from further analysis. It was subsequently reanalysed successfully in the validation cohort. In a further seven cases a single spectrum was generated and analysed. In all other cases peak magnitude was calculated as a mean value from two adequate spectra.

4.4.2.2.2.1 pH4

A total of 51 peaks were detected and assigned an *m/z* value by the Biomarker Wizard. No peaks were found to be significantly different between CC and PSC or between CC and healthy control groups.

4.4.2.2.2.2 pH6

39 peaks were detected and assigned an *m/z* value. One peak was found to be significantly different between CC and PSC or CC and healthy control groups. *M/z* 4631 was differentially expressed in both comparisons (p=0.03 and 0.007 respectively).

4.4.2.2.2.3 pH8

61 peaks were detected and assigned an *m/z* value. One peak was found to be significantly different between CC and PSC or CC and healthy control groups. *M/z* 4630 was differentially expressed in both comparisons (p=0.036 and 0.001 respectively).

4.4.2.2.3 H50 chips, discovery study

A total of 28 peaks were detected and assigned an *m/z* value. Six peaks were found to be significantly different between CC and PSC or CC and healthy control groups; *m/z* 4297 (CC vs. PSC), *m/z* 5768 (CC vs. healthy controls), *m/z* 8572 (CC vs. PSC

and CC vs. healthy controls), *m/z* 8701 (CC vs. PSC and CC vs. healthy controls), *m/z* 11520 (CC vs. healthy controls) and *m/z* 17313 (CC vs. PSC).

Results from of all seven significantly different peaks are summarised as 'Primary analysis discovery cohort' in the first rows of Table 4.13.

4.4.2.2.4 Decision on chip conditions for validation study

CM10/pH4 was excluded as no significantly different peaks were identified in the discovery study and pH4 was expected to be the least selective condition (i.e. most proteins able to bind), thereby reducing resolution. pH6 was excluded as the only significantly different peak, *m/z* 4630, was also detected at pH8 (as *m/z* 4631). The greatest absolute number of peaks was detected at pH8 on the CM10 chip. The significantly different peak, *m/z* 4631, correlated with the positive finding at pH6. pH8 was the most selective condition used on the CM10 chip yet had yielded the only peak of interest. The H50 chip had yielded a substantial number of significantly different peaks (six) with good spectral amplitude and SNR. The use of two different surfaces was considered complimentary, as they are highly likely to bind different proteins. Therefore, CM10/pH8 and H50/ACN 10% were selected as the chip/condition combinations for the validation study.

4.4.2.3 Results of validation SELDI-TOF MS study

Spectra were successfully generated from all subjects under all conditions. 98% of spectra were averages between 2 spots, 2% were from an isolated spot (due to paired spot failure).

4.4.2.3.1 Demographics of subjects, validation study

Samples from 81 patients with CC (48 from ICH, 33 from UCH), 54 patients with PSC (1 ICH, 12 UCH, 41 Norway) and 90 healthy controls (65 ICH, 25 UCH) were included. The characteristics of these patients are summarised in Table 4.14. There were significant differences in age between the CC cohort (median age 66y) and the PSC (42y) and healthy control (42y) cohorts. There were significant differences in

ethnicity between the PSC group, which was 89% Caucasian, and the CC (65% Caucasian) and healthy control (57% Caucasian) cohorts. There were also significant differences in the male to female ratio between the PSC cohort (76% male) and the CC (57%) and healthy control (53%) cohorts. Samples from all subjects from the discovery study also underwent repeat SELDI-TOF MS analysis, but were subject to separate statistical analysis initially. The demographic details of the discovery cohort can be found in Table 4.12.

4.4.2.3.2 Matching of peaks from discovery and validation studies

The *m/z* values of each of the seven peaks from the discovery study, and the *m/z* value of matched peaks in the validation study are summarised in Table 4.15. All seven peaks were matched within the 0.15% range.

4.4.2.3.2.1 CM10 chips, validation study

A total of 41 peaks were detected and assigned an *m/z* value by the Biomarker Wizard. A single peak, *m/z* 4630, was sought and identified within 0.1% of *m/z* 4631. This peak was examined within the Biomarker software and had similar appearances to the *m/z* 4631 peak identified as of interest in the discovery study.

4.4.2.3.2.2 H50 chip, validation study

A total of 34 peaks were detected and assigned an *m/z* value by the Biomarker Wizard. Peaks corresponding to the six identified of interest in the H50 discovery study were identified within 0.15% of their *m/z* value. These peaks were examined manually and validated within the Biomarker software.

4.4.2.3.3 Statistical analysis of validation study cohorts

Results of analysis of all seven peaks of interest are summarised in Table 4.13. Each of the seven peaks was initially analysed in isolation:

4.4.2.3.3.1 New CC cohort vs. new PSC cohort

Three peaks had strongly significant ($p<0.0001$) differences between CC and PSC cohorts: m/z 4631, *m/z* 5768 and *m/z* 11526, all of which were more intense in CC

subjects that PSC subjects. A significant difference was detected in two other peaks: *m/z* 8699 (p=0.03) and *m/z* 17287 (p=0.02), both of which were relatively less intense in CC than PSC groups. The two remaining peaks showed no difference in peak intensity.

4.4.2.3.3.2 New CC cohort vs. new healthy controls

Significant differences were detected in all seven peaks. Four peaks were significantly (p<0.0001) less intense in CC: *m/z* 4291, *m/z* 8701, *m/z* 17287 and *m/z* 8577. Three peaks were significantly higher intensity in CC: *m/z* 11526 and *m/z* 4630 (p<0.0001) and *m/z* 5768 (p=0.001).

4.4.2.3.3.3 New PSC cohort vs. new healthy controls

Significantly lower intensity was found in four peaks: *m/z* 4291, *m/z* 8572, *m/z* 8699 and *m/z* 17287 (p<0.05). Two peaks were increased in PSC: *m/z* 4631 *m/z* 11526.

4.4.2.3.3.4 New Norwegian PSC vs. UK PSC cohort

No peaks were found to vary between the Norwegian PSC subjects and the UK PSC subjects (p=0.30 to 0.75).

4.4.2.3.3.5 New ICH CC vs. new ICH healthy controls

The significant differences in all seven peaks detected in the full cohorts were all sustained when only the ICH subgroups were analysed ($p=2.0x10^{-7}$ to 0.007).

4.4.2.3.3.6 New UCH CC vs. new UCH healthy controls

The significant differences in six of seven peaks detected in the full cohorts were sustained when only the UCH subgroups were analysed ($p=1.2x10^{-7}$ to 0.028). There was no difference found in *m/z* 5765 (p=0.35).

4.4.2.3.4 Analysis of repeat SEDLI-TOF MS in discovery cohorts

The new results for the original discovery cohorts were then subject to analysis, comparing CC to PSC and then CC to healthy control cohorts. There was little consistency in the CC to PSC comparison; only one of five peaks originally found to

be different remained so (*m/z* 4631, p=0.02). There was much better consistency in the CC to healthy control comparisons where five of the six peaks originally found to be different remained so in the repeat study (p=0.016 to $3.0x10^{-4}$).

4.4.2.3.5 Combined cohort

The results from the repeated SELDI-TOF MS analysis of the discovery cohort were then merged with the validation study results to form a combined cohort. All subsequent analyses were performed on this combined cohort.

4.4.2.3.5.1 Demographics of combined cohort

There were samples from 99 CC subjects (58% from UCH, 42% from UCH), 64 PSC subjects (31% UCH, 69% Norway) and 107 healthy control subjects (69% ICH, 31% UCH) in the combined cohort. A summary of demographic data for this combined cohort can be found in Table 4.16. Similar differences persisted. The median age was higher in the CC cohort (median 68y) than in the PSC (42y) and healthy control groups (42y). There were more men in the PSC cohort (78%) than in CC (59%) and healthy control cohorts (52%).

4.4.2.3.5.2 Anatomical subtypes within combined CC cohort

The anatomical subtype was confirmed in 89 of the 99 CC subjects. The majority were hilar (66%) or distal (23%). 11% were intrahepatic. A breakdown is shown in Table 4.18. There was no significant difference in any of the seven peak intensities between different anatomical subtypes.

4.4.2.3.5.3 Laboratory blood results of combined cohort

Routine blood parameters were collated for the combined cohorts and can be found in Table 4.17. Laboratory data were not available for healthy controls. Mann-Whitney U test confirmed that there were no significant differences in bilirubin, ALP, white cell count and AFP between the CC and PSC cohorts. There were significant differences (<0.0001) in ALT, albumin, CRP, Cea and CA19-9. Creatinine and urea could not be compared.

4.4.2.3.6 ROC-AUC analysis of each peak

Each peak was subject to ROC-AUC analysis. Four individual peaks had some utility in discriminating CC from PSC (ROC-AUC of 0.62 to 0.75). The best peak was *m/z* 4631. When CC and healthy control groups were compared, all seven peaks had some utility in discriminating CC and healthy control groups (ROC-AUC of 0.58-0.82). The joint best peaks were *m/z* 8572 and *m/z* 11526. The ROC-AUC results, with derived optimum sensitivities and specificities, are summarised in Table 4.19.

4.4.2.3.7 Correlation matrix, between significant peaks

This matrix can be found in Table 4.20. There was strong correlation between two pairs of peaks: *m/z* 8572 with *m/z* 4191 (R^2 0.89) and *m/z* 5765 with *m/z* 11526 (R^2=0.70). There was moderate correlation between two other pairs of peaks *m/z* 4291 with *m/z* 8699 (R^2=0.51) and *m/z* 8572 with *m/z* 8699 (R^2=0.52). There were no other correlations seen.

4.4.2.3.8 Correlation matrix, between peaks and routine laboratory indices

This matrix can be found in Table 4.21. There was no significant correlation between any of the peaks and any of the routine laboratory indices.

4.4.2.3.9 Multiple linear regression (general discriminant analysis)

4.4.2.3.9.1 CC vs PSC

After four steps, three peaks were included in the optimum model: *m/z* 11526, *m/z* 17287, and *m/z* 4631. β-values, f-to include and f-to exclude values for each of these are included in Table 4.22. When this model was applied to the combined cohort the ROC-AUC reached 0.76, with sensitivity of 75% and specificity of 64%.

4.4.2.3.9.2 CC vs HC

After six steps, the optimum model reached included five peaks: *m/z* 8572, *m/z* 4291, *m/z* 11526, *m/z* 5765, *m/z* 4631. β-values, F-to include and f-to exclude values for each of these are included in Table 4.22. When this model was applied to the

combined cohort the ROC-AUC reached 0.90, with sensitivity of 95% and specificity of 74%.

4.4.2.3.9.3 Combination of peaks with addition of CA19-9

Making CA19-9 values available to the MLR did not displace any of the three protein peaks from the CC vs. PSC model. CA19-9 was included and there was marginal improvement in the ROC from 0.76 to 0.77. Sensitivity was 75% and specificity 69%. No CA19-9 values were available for the healthy control group and so this analysis could not be repeated on the CC vs healthy control model.

4.4.2.3.9.4 Weighting for potential confounders using MLR

Weighting for age, sex, ethnicity and site of collection made no difference to the protein peaks selected for inclusion in the MLR, and no significant changes to the β values occurred.

4.4.2.3.10 Post hoc analysis - patients with PSC and CC

In parallel with the validation cohort, I also subjected 6 samples from patients with PSC **and** CC to SELDI-TOF MS. These were compared only to the PSC cohort (n=64) and only using the three peaks found to be of most utility in discriminating CC from PSC in the main study. MW-U testing identified a significant difference in m/z 4631 (p=0.01) but not in *m/z* 11526 or *m/z* 17287 (Table 4.23).

4.4.2.3.11 UniProtKB database search

A total of 46 proteins fell within the search parameters (Table 4.24). Between 3 and 13 *m/z* matches were found within 0.1% range each of the seven peaks of interest. After consideration of known subcellular location of expression (principally secreted proteins), known function (tumour suppressor or biliary functions) and known malignant disease associations (functional or association as a plasma biomarker), matched proteins that were empirically considered to be of logical interest as CC biomarkers are highlighted in bold in Table 4.24.

4.5 Discussion

I have employed two fundamentally different, but complimentary, proteomic techniques to undertake a unbiased search for potential new protein biomarkers of CC. GeLC-MS/MS offers optimal performance at high mass ranges and offers direct identification of proteins. SELDI-TOF MS offers optimal performance at lower mass ranges, is high throughput and does not require multiple protein separation steps.

In the following sections, I will summarise the work I have undertaken with these techniques. Each of the identified proteins, and unidentified protein peaks of interest, will be considered in turn. In the case of unidentified *m/z* peaks, this will include exploration of the potential identity of the protein in the current literature. I will then consider the relative merits and weaknesses of the experimental approaches that I took. Finally, I will consider the further work that might be undertaken to expand upon the work presented in this chapter.

4.5.1 GeLC-MS/MS

Plasma samples from eight patients with confirmed CC and samples from seven patients with benign, gallstone related biliary disease have been subjected to GeLC-MS/MS. Six proteins that significantly differentiated CC from gallstone disease patients were identified. In combination, two of these six proteins had maximal utility in discriminating CC from gallstone disease, with a ROC-AUC approaching 1.0.

4.5.1.1 Proteins identified

4.5.1.1.1 Carboxypeptidase N subunit 2 (CPN2)

This protein was identified on the basis of four tryptic peptides specific to CPN2. It was expressed at lower levels in the plasma of CC patients (0.77 fold). It is an 83kDa protein and is a regulatory subunit of Carboxypeptidase N (CPN), binding the catalytic subunit and keeping it in circulation. CPN is a plasma zinc metalloprotease that cleaves carboxy-terminal arginines and lysines from peptides found in the bloodstream such as complement induced anaphylatoxins, kinins, and creatine

kinase MM. CPN is constitutively expressed in the liver and secreted into the bloodstream.[373] Very low constitutional levels of CPN have been rarely associated with cases of hereditary angioedema.[374] There is no published evidence of reduced plasma levels of CPN or CPN2 in other diseases, or utility as a cancer biomarker.

4.5.1.1.2 C-reactive protein (CRP)

CRP was identified on the basis of seven tryptic peptides specific to CRP. It was three-fold higher in the CC cohort than the benign group. In the study cohort it had a favourable ROC-AUC of 0.91. CRP is a 25kDa acute phase protein. It is synthesised in the liver in response to factors released by macrophages and secreted into the plasma. CRP is thought to have physiological function, binding to the phosphocholine expressed on the surface of dying cells, and some types of bacteria, thereby activating the complement system through the C1Q complex.[375] There is limited evidence from one study that modest elevation of CRP predicts poorer prognosis in patients undergoing curative or palliative therapy for hilar CC.[376] There is more compelling evidence of elevation of CRP in asymptomatic subjects who subsequently develop colorectal carcinoma.[377] Whether the increased level of CRP noted in my cohort is due to low-grade biliary infection, a tumour related inflammatory response or other inflammatory cause is not clear. It is interesting to note that the median CRP level in the routine laboratory analyses was actually *higher* (15 mg/L) in the benign cohort than in the CC cohort (7.5 mg/L). Although this was not a significant difference, the trend is in the opposite direction of that seen in the LC-MS/MS analysis. Direct pair-wise correlation of the laboratory CRP and the result from LC-MS/MS showed no correlation between the results.

4.5.1.1.3 Glutathione peroxidase 3 (GPX3) precursor

This protein was identified from four tryptic peptides specific to GPX3 and was two-fold higher in the CC group. The ROC-AUC of 0.98, and inclusion in the ideal panel of proteins in the multiple linear regression model, suggests very good diagnostic

function in these small groups of subjects. GPX3 is a secreted protein that protects cells and enzymes from oxidative damage, by catalysing the reduction of hydrogen peroxide, lipid peroxides and organic hydroperoxide by glutathione.[375] There is limited evidence of disease associations with GPX3. Relatively old studies have shown *low* plasma levels in patients with breast, GI and various childhood cancers.[378, 379] Furthermore, *low* levels of GPX3 have recently been described in patients with ovarian cancer.[380] There is experimental evidence, in pancreatic cancer cell lines, that glutathione peroxidase may have a tumour suppressant effect.[381] The fact that my study has shown *increased* levels of the precursor molecule is interesting and, perhaps, may point to increased production due to an upstream consumptive process.

4.5.1.1.4 Leucine-rich α-2-glycoprotein 1 (LRG1)

LRG1 was identified from two tryptic peptides specific to LRG1 and found to be 2.2-fold higher in the CC group. It had good individual diagnostic performance with a ROC-AUC of 0.93. LRG1 is a secreted protein and involvement in protein-protein interaction, signal transduction, and cell adhesion and development has been demonstrated.[375] A Japanese proteomic study (utilising multi-dimensional liquid chromatography and two-dimensional difference gel electrophoresis) of pooled plasma from 5 pancreatic cancer patients and five control subjects showed upregulation of LRG1. This was confirmed by Western blotting in validation cohorts.[382] A UK group, using a two dimensional difference gel electrophoresis and tandem mass spectrometry proteomic approach, has recently reported increased plasma LRG1 in biliary tract cancer patients. 37 biliary tract cancer patient, 11 PSC patients and 7 autoimmune pancreatitis patients were studied. They verified their finding in the same subjects using ELISA. In combination with plasma CA19-9 and IL-6, a ROC-AUC of 0.98 was attained.[383] LRG1 is also elevated in various

autoimmune conditions, ovarian cancer, appendicitis and ulcerative colitis, perhaps raising questions about its specificity as a pancreatobiliary tumour biomarker.[384-388]

4.5.1.1.5 Serum amyloid A4 (SA4)

SA4 was identified from two tryptic peptides specific to SA4 and was 1.7-fold higher in the CC cohort than the benign groups. Its ROC-AUC 0.93 suggests good performance in this cohort of patients. SA4 is an apolipoprotein of the HDL complex and an acute phase protein. It is expressed by the liver and secreted into the plasma. A number of small LC-MS/MS plasma proteomic studies have identified SA4, at elevated levels, in breast cancer, hepatocellular carcinoma and pancreatic cancer.[389-391] It has also been identified in proteomic studies of a host of other disparate diseases, including traumatic brain injury and infective exacerbation of COPD.[392, 393] Despite its frequent elevation as known an acute phase protein, a number of experimental studies have indicated a directly tumour related increase of SA4.[394]

4.5.1.1.6 Zinc finger protein 844 (ZFP844)

This zinc finger protein was identified by two tryptic peptides specific to ZFP844 and was two-fold higher in the CC group. It had an excellent ROC-AUC of 0.98 and was one of the two proteins included in the multiple linear regression optimum model. ZFP844 is expressed in the nucleus and is likely to be a transcription factor.[375] I was not able to identify any previous associations between this protein and disease, benign or malignant.

4.5.1.1.7 Combinations of identified proteins

The multiple linear regression (MLR) discriminant function analysis identified a panel of two of these proteins, GPX3 and ZFP844, which had a 100% sensitivity and specificity in *these cohorts*. Given the small cohort sizes, and the very high ROC-AUC (up to 0.98) of two of the individual markers, this does not serve as any

indicator of which of these six proteins is a priority of further investigation and which should be discounted.

4.5.1.2 Experimental technique

A major advantage of GeLC-MS/MS is the integral identification and quantification of the proteins detected, and this was achieved in my study. The technique was expected to be somewhat laborious, limiting the number of samples that can be processed in a reasonable timeframe, and this also proved to be the case. The first stage of the protocol, using an IgY column to deplete the most abundant plasma proteins, proved technically challenging. It required repetition after several optimisation experiments to achieve success. Initially, the column became progressively less effective with each sample depleted, with comparative overabundance of albumin in later samples, representing a clear potential confounding factor. I ran gels comparing the stripped and eluted fraction of each specimen, and concluded that the dilution buffer was attenuating the acid conditions required to strip the bound protein fraction. A second stripping cycle, with reapplication of an acidic buffer, solved this issue. The performance of the protein depletion was confirmed, and appeared to meet the manufacturer's claim of 95% efficacity. However, the difference between the most abundant proteins and the least abundant proteins currently detectable in plasma is of about nine to 12 orders of magnitude. This has been described as 'equivalent to the difference between the height of an ant and the distance from the Earth to the moon'.[386, 395] It is clear that the partial, even 95%, depletion of albumin and other high abundance proteins will only mitigate the masking effects on moderately abundant proteins. Furthermore, in removal of these proteins I may also have removed other proteins of potential interest – including proteins that bind to albumin, or proteins that undergo unintended binding to the IgYs in the depletion column due to antibody cross-reaction.

Gel electrophoresis, the LC-MS/MS and statistical analyses proceeded without undue complication. The design of the study completed, with a single (discovery) phase testing two small groups of samples, is vulnerable to Type I error due to multiple testing. Therefore, the results can only be regarded as exploratory. The verification of findings in the current cohorts using a different technique, and validation in new samples, would have addressed this. The results of ROC-AUC analyses on these small groups must also be treated with caution and simply give an indication of the degree of separation in the cohorts that are under investigation. The results cannot be extrapolated beyond this.

4.5.1.3 Cohorts recruited

This study was conducted at the beginning of my research period, when I had collected a small number of samples. The cohort sizes of eight and seven are comparable with most recently published proteomic studies that have utilised a gel and tandem MS based approach. Many studies still employ the strategy of pooling all the case and all the control samples and then passing just two samples through the analytical technique. This has obvious advantages if time and resources are short but, with an *n* of one in each group, results can only be regarded as exploratory at best. Perhaps more significantly in my study, the limited samples available meant that the demographics of the case and control groups were not as consistent as I would have liked for such small cohorts. Although age was well matched between the groups, the case and control groups were significantly different in terms of ethnicity and sex. These factors may impinge on any conclusions that can be drawn about the differentially expressed proteins. If more samples had been available I would not have necessarily included more in each group, but would certainly have endeavoured to match the subjects included more closely for sex and ethnicity. As all differentially expressed proteins are rapidly identified by this technique, any protein

differences that were obviously attributable to female sex, for example, could have been be promptly excluded from further investigation.

There were no significant differences in the known standard laboratory indices between the two groups. In one benign case, all of the standard laboratory indices were unavailable. No tumour marker results were available for all but one of the benign control group. It is likely that tumour markers, especially CA19-9, would have been different between the groups. Having the values would have permitted me to explore combinations of traditional CC biomarkers with the six putative protein markers identified with MLR. However, the fact that two of the putative markers had individual ROC-AUCs of 0.98 means that such a MLR analysis would be very unlikely to offer further insight.

The CC samples in my study were from patients with symptomatic, fully diagnosed disease. They therefore have, by definition, relatively late CC. Extrapolation of my findings to early, asymptomatic disease is not possible. A control disease group were chosen, over a healthy control group, to mitigate the potential confounders of cholestasis and biliary epithelial inflammation on the proteome. The similarity of the normal plasma markers of cholestasis and inflammation in the two groups suggest that this was at least somewhat successful. However, gallstone disease is not a close corollary of PSC, a group in which accurate diagnosis of CC is of particular importance and difficulty. I cannot extrapolate my findings to PSC related CC without further work in appropriate new cohorts.

4.5.1.4 Further work warranted

The number of peptides on which the identification is made is an indicator of the certainty of that identification. The veracity of the protein identifications in my study is robust, with at least two specific peptides identifying each, but internal verification within the current cohort is required. Secondly, the limited cohort sizes in the study mandate that any verified protein differences be validated in further patients. Current

standard best practice in proteomic studies would be to confirm the presence and concentration of each protein identified in the current samples using Western blotting. If the identification and differentiation of the proteins is verified, validation of the findings in a further, larger cohort of patients is indicated. GPX3 and ZFP844 were the two strongest biomarkers in this discovery study. There is relatively little known about any disease associations of GPX3, and nothing about associations with ZFP844. These two candidate proteins warrant formal verification in the current cohorts with Western blotting, before validation in a large new cohort with Western blot, or preferably ELISA. At first sight, SA4 seems rather a non-specific candidate to warrant consideration as a CC biomarker, but it has also been found to be elevated in the plasma in HCC and pancreatic cancer studies. Confirmation in a further cohort, alongside the other potential markers could therefore be warranted. CRP seems highly unlikely to represent a specific marker of CC but would be amenable to simple further investigation as it is available in all clinical labs. It could easily and cheaply be performed in a study of patients with indeterminate biliary stricture, or in a longitudinal study of patients with PSC. A quick, cheap retrospective analysis might also be feasible. CPN2 is the inactive, stabilising subunit of an enzyme and has no known disease associations. It was reduced in CC patients, but only moderately so (0.77-fold). It therefore seems a less attractive potential marker for further investigation. However, it was firmly identified with four specific peptides and could simply be investigated in a further cohort, alongside GPX3 and ZFP844. LRG1 has now been shown to be elevated in bile duct cancer by two separate research groups using different techniques - at University College London using two-dimensional difference gel electrophoresis and tandem mass spectrometry and then verified in the same cohorts with ELISA, and now in my study using SELDI-TOF MS. It seems that study of this protein in a larger, clinical trial may now be warranted.

4.5.2 SELDI-TOF MS

SELDI-TOF MS has been applied to plasma samples from a total of 99 CC patients, 64 PSC patients and 107 healthy control subjects, initially divided into a smaller discovery cohort and a larger validation cohort. After identifying the optimum conditions under which to analyse the samples, a discovery cohort was investigated and seven individual *m/z* peaks, each representing an as yet unknown protein, were detected that were differentially expressed between cohorts. All seven of these were confirmed in the subsequent validation study. All seven peaks exhibited some diagnostic utility in discriminating CC subjects from PSC patients, or CC subjects from healthy controls. The discriminatory function of combinations of these peaks was assessed for two key comparisons: CC subjects compared to PSC subjects and CC compared to healthy control subjects. A combined panel of three of these seven peaks demonstrated substantial utility in discriminating CC from PSC, with a ROC-AUC of 0.76, sensitivity of 75% and specificity of 64%. In discriminating CC from healthy controls, a combined panel of five peaks demonstrated very good diagnostic function, with a ROC-AUC of 0.90, sensitivity of 95% and specificity of 74%.

4.5.2.1 Peaks identified as of interest

All peaks identified as of interest were singly resolved, of good (>20%) relative intensity and had no interference from neighbouring peaks.

4.5.2.1.1 *m/z* 4291

This peak was significantly less intense in CC and PSC groups than in healthy controls. This peak had no function in discriminating PSC from CC but was useful when CC and healthy controls were compared (ROC-AUC 0.76), where it was included in the calculated optimum peak panel. Peak *m/z* 4291 was closely correlated (R^2 0.90) with peak *m/z* 8572, with which it also shared the same pattern of discriminatory characteristics. The expected *m/z* of a doubly charged ion of *m/z* 8572 is 4286 and it is therefore highly likely that these peaks represent the same protein. Peak *m/z* 4291 did not correlate with any routine laboratory indices.

4.5.2.1.2 *m/z* 4631

This peak was the only one of the seven peaks that was identified using the CM10 chip surface. It was found to be at increased intensity in both PSC and CC patients. It was the most effective individual peak at discriminating both CC from PSC subjects (ROC-AUC 0.74), and CC from healthy controls (ROC-AUC 0.82), and was included in the optimum panel of proteins calculated for both comparisons. This peak was also elevated in a small additional group of six patients with CC **and** PSC that I was able to analyse and compare to the PSC only group as a *post hoc* analysis. *M/z* 4631 was not correlated with any of the other peaks, or with any routine laboratory results.

4.5.2.1.3 *m/z* 5765

This peak was significantly more intense in CC patients than in PSC subjects and healthy controls. It had moderate function as a potential diagnostic test in identifying CC from PSC subjects (0.63) and CC from healthy controls (0.59). It was included in the optimum panel of peaks to discriminate CC from HC cohorts. *M/z* 5765 was strongly correlated with another peak, *m/z* 11526 (R^2 0.70), with which it shared a similar pattern of discriminatory characteristics. The expected *m/z* of double charged version of 11526 is 5763. It is therefore highly likely that these two peaks represent the same protein. *M/z* 5765 was minimally correlated with CRP (R^2 0.15).

4.5.2.1.4 *m/z* 8572

This peak was less intense in both CC and PSC cohorts. It had no significant utility in discriminating CC from PSC subjects, but good performance in discriminating CC from healthy controls (ROC-AUC 0.82), where it was included in the optimum model. As discussed above, it is likely to represent a singly charged version of the same protein detected as peak *m/z* 4191. *M/z* 8572 was also moderately correlated with *m/z* 8699 (R^2 0.51), but not with any laboratory parameters.

4.5.2.1.5 *m/z* 8699

This peak was significantly less intense in CC and PSC groups. Differentiation of CC and PSC (ROC-AUC 0.69), and CC and healthy controls (ROC 0.74), was good. However, it was not included in the optimum discriminatory panel for either of these key comparisons. *M/z* 8699 was moderately correlated with both *m/z* 4291 and *m/z* 8572, with which it shares the same pattern of variation in intensity. It is not clear why this is. It did not correlate with any lab parameters.

4.5.2.1.6 *m/z* 11526

This peak was more intense in both CC and PSC groups. Discrimination of CC from PSC (ROC 0.73) and CC from HC (ROC 0.83) were both good. This peak was included in the optimum panel of peaks for both key group comparisons. As discussed above, it was correlated with *m/z* 5765 and these are likely to represent differently charged versions of the same protein. There was very modest correlation between this peak and plasma CRP (R^2 0.20).

4.5.2.1.7 *m/z* 17287

This peak was of lower intensity in CC and PSC groups. Discrimination of CC subjects from those with PSC was modest (ROC 0.61) and somewhat better when comparing CC with healthy controls (ROC 0.72). In multiple linear regression this peak was included in the optimum panel of proteins that discriminated CC from PSC. There was no correlation between this peak and any others, or the laboratory indices.

4.5.2.1.8 Combinations of peaks

Multiple linear regression analyses allowed me to explore combinations of peaks. Doing so achieved improved diagnostic performance. The model differentiating CC and PSC included three peaks and achieved a ROC of 0.76. This is somewhat disappointing as the best individual peak had a ROC-AUC of 0.74. The panel differentiating CC from healthy controls included five peaks and reached a ROC-AUC of 0.90. When CA19-9 was offered for inclusion in the CC vs. PSC, all three peaks

were retained in the model. Addition of CA19-9 enhanced the CC vs. PSC panel slightly, achieving a ROC of 0.77, with sensitivity of 75% and specificity of 69%. No CA19-9 values were available for healthy controls and so addition of CA19-9 to that panel could not be undertaken.

4.5.2.1.9 Potential identity of *m/z* peaks

4.5.2.1.9.1 Published SELDI-TOF MS studies

I reviewed published work that has previously utilised SELDI-TOF MS on plasma from patients with CC, pancreatic cancer or hepatocellular carcinoma. The peaks of interest from these nine studies, and any subsequent protein identifications, are summarised in Table 4.25. Of the 43 peaks identified of interest in the nine studies considered, four fall within 0.15% of the seven peaks that I have identified. Three of these matched peaks were reported in the Scarlett paper, the first to report application of SELDI-TOF to plasma from CC patients.[360] The differences shown by Scarlett between CC and non-CC subjects in each of these three matching peaks are contiguous with the findings of my SELDI-TOF MS experiment. One of the matching Scarlett peaks, *m/z* 11535, was the one found to be of most value in discriminating CC from healthy controls (ROC-AUC 0.87) in that study. The fourth matching peak, *m/z* 8560, is from a very recent Chinese study of pancreatic cancer. In this case, the direction of difference conflicts with my findings, with increased peak intensity in CC compared to controls. The matching peaks are summarised in Table 4.26. Neither investigating groups identified any of the four matched peaks, nor have they published any subsequent work on these peaks. The correlation between the findings in Scarlett and colleagues and my study are of interest, and certainly support the need to identify the proteins that these *m/z* peaks represent.

4.5.2.1.9.2 Search for m/z values in protein TagIdent database

Of the 46 protein matches on the TagIDent database, several bear some scrutiny as potential biomarkers. However, this is entirely exploratory and many of the promising

matches are actually expressed at counter-intuitively increased or decreased levels. One of the hits, MHC class I polypeptide-related sequence, has previously been shown to be shed from epithelial cell tumours, including HCC, into the blood plasma and could represent the peak at *m/z* 4291. However, *m/z* 4291 is *reduced* in CC in my study, which would make little sense. As a further example, the serum amyloid A1 (SAA1) could represent the *m/z* peak at 11526, which I have shown to be increased in CC. This would also correlate with my finding of increase in another serum amyloid, SAA4, in the GeLC-MS/MS study. However, SAA1 is also in the short list of proteins at *m/z* 8572, which is *reduced* in CC. Speculative examination of this database was worthwhile, but I am doubtful that any reasonable conclusions can be drawn.

4.5.2.2 Experimental techniques

The sample processing and SELDI-TOF MS process itself were relatively straightforward and reliable, and processed large numbers of specimens quickly, when it was up and running. The choice of two chip surfaces did prove worthwhile as, despite it being the only significant chip identified on CM10, *m/z* 4631 proved to be the strongest individual peak found, and did not correlate at all with any of the six peaks of interest found on H50 chips. Several of the buffers that I used in the optimisation experiments failed to generate any spectra and I did not reach an explanation for this. One sample failed to generate spectra, in both replicate spots. A sample from the same subject (prepared as a fresh sample from the raw plasma aliquot) was included in the validation cohort and generated normal, adequate spectra. In the validation cohort, spectra were achieved from all samples included. The significant differences seen in the discovery cohort were mitigated, and in some cases abrogated, when the same samples were subjected to repeat SELDI-TOF MS. However, the candidate peaks were confirmed in the validation cohorts. The design of the study, with discovery and validation phases, adds strength to the findings and

minimises the risk of type I error due to multiple testing. Once differences had been confirmed in the validation cohort, I elected to combine the results from the validation cohort results with the repeated SELDI-TOF results from the reanalysed discovery samples. This permitted analysis of the maximum sized cohorts in the subsequent ROC-AUC and MLR analyses, with consequent strengthening of the findings.

4.5.2.3 Cohorts recruited

The size of cohorts that I was able to assemble for this study compare very favourably with previously published SELDI-TOF MS biomarker discovery studies. My study included substantially more CC subjects than any of the nine summarised in Table 4.25. I elected to study three groups – with PSC as a disease control group, excluding those with advanced cirrhosis or previous liver transplantation. Although this complicated recruitment, simply comparing patients with cancer to completely healthy controls would not be adequate. Potential confounding factors such as biliary obstruction could not be investigated if the only control group all, by definition, had normal liver function tests. Diagnosing PSC related CC poses a very specific, frequent and very important clinical dilemma. It was therefore appropriate to use PSC as the primary control group, with secondary comparisons to healthy controls. However, this decision was not without ramification. PSC is an uncommon disease and neither of the UK centres at which I was recruiting patients have large PSC practices. Although I was able to identify over 60 potential recruits, successfully calling them in to consider participation and give blood proved difficult. Non-cirrhotic, non-transplanted PSC patients are generally young, and frequently have full time jobs and young families. They are not ideal targets for the research fellow's telephone calls and, as generally well, are not frequent attenders in outpatient clinics. I therefore had to enter into collaboration with colleagues at the Oslo based NoPSC Biobank. They were in a position to help, have a large resource of well-characterised PSC patient's samples and their sample processing is in line with our SOP.

Even with access to the NoPSC Biobank, there was an unavoidable preponderance of young, male, Caucasian patients in the PSC control group. This cohort was not a good match for the older, more gender balanced, ethnically diverse CC group. I endeavoured to collate a healthy control group that was better matched to the CC group and this was generally achieved in terms of sex and ethnicity. However, an age difference persisted. A final important factor was my decision to collect samples at two UK hospitals, ICH and UCK along with the Norwegian site. This afforded quicker recruitment of minimum cohort sizes and eventually bigger cohorts for comparison, but could become a confounding factor. All three institutions have long established, high quality clinical and research governance standards. All three collaborating groups are very experienced in clinical research and have undertaken proteomic studies, so understand the importance of consistent, careful sample treatment. Compatibility of sample collection and processing SOPs was ensured.

It was very likely that there would be significant differences in other factors between the three cohorts. Some degree of biliary inflammation, cholestasis and cholangitis is a frequent, if not universal, feature of CC and PSC. In patients with malignancy, and a lesser extent PSC, nutritional status and renal function are also likely to be impaired. However, the severity of these features is highly variable and therefore and the impact on the proteome is likely to also vary. The routine laboratory indices that I was able to collect offer insight into some of these factors. Importantly, the principal markers of cholestasis, ALP and bilirubin, were not significantly different between CC and PSC groups. ALT and albumin were significantly higher in the PSC group. CRP and CA19-9 were significantly higher in CC. There were no differences in white cell count or AFP between the two groups. Renal function could not be compared, as these numbers were not available for the PSC cohort. It is worth noting that none of the CC or PSC patients recruited were known to have renal failure. It was not possible to make the same comparisons between CC and healthy control groups as I

did not have laboratory data for the healthy controls, although it seems reasonable to assume that routine lab indices in this groups would fall in a normal distribution within, and just out of, the laboratory defined normal range.

The subgroup analyses and other statistical investigations that I was able to perform also offer some reassurance about potential confounding factors. The pattern of significantly increased or decreased intensity in CC groups for all but one of the seven peaks was the same when UCH and ICH samples were analysed in isolation. Each of the seven peak intensities in the Norwegian PSC cohort was not different to the UK PSC cohort. All laboratory indices were correlated with all seven peaks of interest. Apart from minor correlation between CRP and peaks *m/z* 5765 and *m/z* 11526, no correlation was found. Multiple linear regression (MLR) is able to offer some correction for potential confounding factors. I repeated the principal MLR analysis with correction for age, sex, ethnicity (Caucasian or non-Caucasian) and site of collection, in case any of these factors were affecting my findings. In all cases, such correction had no more than a negligible positive or negative effect on the diagnostic performance. Finally, the MLR model discriminating CC vs. PSC model was repeated with correction for the various differential laboratory blood indices: ALT, albumin, CA19-9 and CRP. Such analyses are reassuring but are not a substitute for better-matched groups. Certainly my study would have been strengthened if CC and healthy control samples had been available from the Norwegian group.

The CC group recruited had heterogeneity in the extent and primary location of their cancer. I investigated this with analysis of the three anatomical subgroups of CC and found no difference in the peak intensities between them. All patients had fully diagnosed symptomatic CC and so results cannot be extrapolated to those with early, asymptomatic disease. None of the discovery or validation CC groups had PSC. However, I had a small additional cohort of six CC patients who did have

underlying PSC. A *post hoc* comparison of *m/z* 11526, 17287 and 4631 between this small group and the 64 PSC patients confirmed a significant difference in 4631 (p=0.01).

4.5.2.4 Further work warranted

Given the large sample sizes and strong validation of my discovery findings, I believe that it is appropriate to pursue identification of the seven peaks of interest. This will be pursued by other members of our group. For each of the seven peaks, we have identified four samples that generate particularly intense signal and pooled them. Protein purification, using sequential ion exchange, reverse-phase chromatography and SDS-PAGE, will be undertaken. Subsequent protein identification will be achieved by nanoflow liquid chromatography tandem MS analysis of trypsin-digested peptides.[371, 396]

4.5.3 Conclusion

In the course of this work, I have come to understand some of the challenges that proteomic studies all face. Clearly appropriately matched control groups are of critical importance. However, the heterogeneity of HPB diseases, and CC in particular, makes matching the groups particularly challenging. In clinical studies, some degree of compromise is necessary but careful thought must be given if reasonable comparisons and conclusions are to be drawn. Proteomic studies are particularly vulnerable to patient, sample collection and sample processing variation. The huge relative abundance of proteins that are of no interest, and the miniscule concentration of those that are, makes biomarker discovery a technically demanding challenge, one that is perhaps not readily appreciated. The volume of data that modern analytical devices can generate on single specimens is huge, and is compounded by the number of specimens that can be processed rapidly. This offers great opportunity but could become overwhelming if adequate bioinformatics software and support are not applied. I was not able to personally complete

validation of the proteins identified in the GeLC-MS/MS study or identification of the SELDI-TOF MS peaks during my PhD programme, and this limits some of the conclusions that I can draw.

The findings of the GeLC-MS/MS study that I undertook have identified a number of proteins that, when considered in the context of other published work, do look of interest and worthy of further investigation. The contemporaneous, provisional identification of the proteins detected is a huge advantage of this technique. However, the statistical significance of these findings in such small cohorts requires cautious interpretation. The proteins identified require careful validation before any conclusions can be drawn. Since my work was completed, other members of my group have sought to validate one of the markers, Carboxypeptidase N, in a fresh cohort of CC patients.

The SELDI work that I have undertaken has delivered strong positive results with large validation cohorts studied. Combination of three of these peaks discriminated CC from PSC patients with sensitivity of 75% and specificity of 64%. Combination of five of these peaks discriminated CC from healthy control patients with sensitivity of 95% and specificity of 74%. Again, it was pleasing to find some supportive, corresponding results from a different group along with some unique new peaks to report. The collaboration with the NoPSC Biobank, Oslo proved fruitful and highlights the importance of well curated, disease specific Biobanks. I believe that all seven of the *m/z* peaks have been subjected to robust testing and all warrant further study. It is likely that these seven peaks represent five unique proteins. Peaks corresponding closely with three have these have been identified in another SELDI-TOF study of CC, supporting my findings. My strongest peak, *m/z* 4631, appears to be a novel finding, never previously identified in SELDI-TOF MS studies of HPB cancer.

4.6 Chapter 4 Tables and Figures

Table 4.1: Child-Pugh classification of cirrhosis

Legend: Child-Pugh (C-P) classification is a scale of severity of cirrhosis that predicts survival. A score is calculated based on a total of 5 laboratory and clinical features. Using the second part of the table, a categorization and predicted 1 and 2 year survival are calculated. No PSC patients with C-P B or C disease were included in my PSC cohort.[397, 398] INR – international normalized ratio of prothrombin time.

Measure	1 point	2 points	3 points
Total bilirubin, umol/l	<34	34-50	>50
Plasma albumin, g/l	>35	28-35	<28
INR	<1.7	1.71-2.30	>2.30
Ascites	None	Mild	Moderate to Severe
Hepatic encephalopathy	None	Grade I-II	Grade III-IV

Points	C-P classification	1 year survival	2 year survival
5-6	A	100%	85%
7-9	B	81%	51%
10-15	C	45%	35%

Table 4.2: Constituents of protein standard, SELDI-TOF MS study

Legend: A standard calibrant solution with an appropriate range of MW peptides and proteins was required to calibrate the SELDI-TOF MS. Two spots on every 96 spot array were reserved for this protein standard.

	Molecule	MW (kDa)	Concentration (µM)
Peptides	Angiotensin	1.2965	1
	[Glu1] fibrinopeptide B	1.570.6	1
	Dynorphin A	2.1475	1
	ACTH	2.9335	1
	B-endorphin	3.465	1
	Insulin, bovine	5.7336	1
	Ubiquitin, bovine	8.5648	1
Proteins	Cytochrome C	12.360	1
	Lysozyme	14.307	1
	Carbonic anhydrase	29.095	1
	Ovalbumin	44.287	1
	Bovine serum albumin	66.393	1

Table 4.3: SELDI-TOF MS instrument settings

Legend: The same instrument settings were used for all studies. Standard settings differed according to which energy absorbing matrix (EAM) was in use. SPA – sinapinic acid, CHCA - α-cyano-4-hydroxycinnamic acid. Unless otherwise specified, units are arbitrary to the instrument.

Setting	**Value set**	
	SPA as EAM	*CHCA as EAM*
High mass (kDa)	100	20
Optimum range (kDa)	10,000-100,000	1,500-20,000
Laser intensity	240	180
Detector sensitivity	8	8
Lag time (Da)	7.5	20
Deflector	1500	3000
Transients per position (n)	20	20
Start position	20	20
End position	80	80
Moving	5	5
Warm up shots (n)	2	2
Laser intensity for warm up	190	250
Include warm up shots?	No	No

Table 4.4: Demographic data, GeLC-MS/MS study

Legend: Details of sex, age and ethnicity of cases in the malignant (CC) group and benign gallstone disease (GSD) group

Ref	Sex	Age	Ethnicity	Diagnosis
CC1	M	78	Caucasian	CC
CC2	M	60	Caucasian	CC
CC4	M	59	Asian	CC
CC5	F	62	Caucasian	CC
CC6	M	66	Afro Caribbean	CC
CC8	F	59	Asian	CC
CC9	M	59	Caucasian	CC
CC10	F	48	Asian	CC
BB1	M	78	Caucasian	GSD
BB2	F	86	Caucasian	GSD
BB3	F	51	Caucasian	GSD
BB4	F	55	Caucasian	GSD
BB5	F	51	Caucasian	GSD
BB6	M	78	Caucasian	GSD
BB7	F	50	Caucasian	GSD

Table 4.5: Descriptive statistics of demographic factors, GeLC-MS/MS study

Legend: CC and benign gallstone disease cohorts

	CC	Benign
n	8	7
Male n (%)	5 (63%)	2 (29%)
Median age (IQ range)	60 (59-66)	55 (51-80)
Ethnicity		
Caucasian (%)	4 (50%)	7 (100%)
Afro-Caribbean	1 (13%)	0
Asian	3 (37%)	0

Table 4.6: Standard laboratory indices for each patient, LC-MS/MS study

Legend: Median and range for each parameter is shown for both groups. P-value shown is from M-W U test between malignant (CC) and benign (stone disease) groups. Where value unavailable, field is hyphenated.

Ref	CRP	Alb	WCC	Bili	ALT	ALP	CeA	CA19-9
Malignant								
CC1	-	20	10.6	102	86	359	1	54
CC2	7	35	25	5	17	152	2	131
CC4	35	29	19	15	68	299	3	42
CC5	8	35	14	9	23	108	4	24
CC6	147	29	1.1	7	81	126	10	885
CC8	8	34	6.2	15	15	109	1	92
CC9	-	29	3.6	44	49	454	1	266
CC10	21.7	31	8.5	11	15	102	-	9
Median (range)	**7.5 (6-147)**	**30 (20-35)**	**9.6 (1-25)**	**13 (5-102)**	**36 (15-86)**	**139 (108-454)**	**2 (1-10)**	**73 (9-885)**
Benign								
BB1	24	27	11.9	10	8	173	-	-
BB2	5	42	7	11	27	131	-	-
BB3	-	-	-	-	-	-	-	
BB4	26	36	7.6	4	15	102	-	-
BB5	6	37	-	17	35	362	14	132
BB6	5	29	53	26	131	234	-	-
BB7	9	35	7.1	9	265	111	-	-
Median (range)	**14.8 (5-26)**	**34.5 (27-37)**	**9.45 (7-53)**	**10.5 (4-26)**	**17.5 (8-265)**	**152 (102-362)**	-	-
P-value	**0.29**	**0.3**	**0.9**	**0.91**	**0.56**	**0.86**	**-**	**-**

Table 4.7: Shortlist of chip, condition and EAM combinations for preliminary optimisation study

Legend: A total of 14 different CM10, and 8 different H50, combinations were trialled.

Chip	Buffer	Matrix
CM10	Glycine and HCl, pH 3	SPA
CM10	Glycine and HCl, pH 3	CHCA
CM10	Sodium acetate, pH 4	SPA
CM10	Sodium acetate, pH 4	CHCA
CM10	Ammonium acetate, pH 4	SPA
CM10	Ammonium acetate, pH 4	CHCA
CM10	Sodium phosphate, pH 6	SPA
CM10	Sodium phosphate, pH 6	CHCA
CM10	Citric acid, pH 6	SPA
CM10	Citric acid, pH 6	CHCA
CM10	HEPES, pH 7	SPA
CM10	HEPES, pH 7	CHCA
CM10	Tris, pH 8	SPA
CM10	Tris, pH 8	CHCA
H50	ACN 10% + 0.1% TFA	SPA
H50	ACN 10% + 0.1% TFA	CHCA
H50	ACN 20% + 0.1% TFA	SPA
H50	ACN 20% + 0.1% TFA	SPA
H50	ACN 20% + 0.1% TFA + NaCl	CHCA
H50	ACN 20% + 0.1% TFA + NaCl	CHCA
H50	ACN 50% + 0.1% TFA	SPA
H50	ACN 50% + 0.1% TFA	CHCA

Table 4.8: Non-redundant proteins identified, GeLC-MS/MS study

Legend: All proteins identified from their peptide fingerprint are listed. The NCBI (National Center for Biotechnology Information, U.S. National Library of Medicine, Bethesda, USA) reference is a unique identifier used in the collation of proteomic data in international databases. Proteins found in different abundance between CC and healthy control subjects are highlighted in bold within this table. Proteins are listed met minimum degrees of certainty (Xcorr value >1.5 for single charge, >2.0 for double charge and >2.5 for triple charge. Minimum number of unique peptides: 2.)

NCBI number	Approved protein name
NP_000005.2	alpha-2-macroglobulin precursor
NP_000020.1	angiotensinogen preproprotein
NP_000030.1	apolipoprotein A-I preproprotein
NP_000031.1	apolipoprotein C-III precursor
NP_000032.1	apolipoprotein E precursor
NP_000033.2	apolipoprotein H precursor
NP_000053.2	serpin peptidase inhibitor, clade G, member 1
NP_000055.2	complement component 3 precursor
NP_000056.2	complement component 6 precursor
NP_000087.1	ceruloplasmin precursor
NP_000177.2	complement factor H isoform a precursor
NP_000195.2	complement factor I preproprotein
NP_000217.2	keratin 9
NP_000362.1	transthyretin precursor
NP_000375.2	apolipoprotein B precursor
NP_000403.1	histidine-rich glycoprotein precursor
NP_000412.3	keratin 10
NP_000414.2	keratin 2
NP_000415.2	keratin 5
NP_000437.3	paraoxonase 1 precursor
NP_000468.1	albumin preproprotein
NP_000473.2	apolipoprotein A-IV precursor
NP_000474.2	apolipoprotein C-II precursor
NP_000479.1	serpin peptidase inhibitor, clade C, member 1
NP_000497.1	coagulation factor II preproprotein
NP_000499.1	fibrinogen, alpha polypeptide isoform alpha-E prep
NP_000500.2	fibrinogen, gamma chain isoform gamma-A precursor
NP_000509.1	beta globin
NP_000510.1	delta globin
NP_000517.2	keratin 14
NP_000542.1	von Hippel-Lindau tumor suppressor isoform 1
NP_000549.1	alpha 1 globin
NP_000558.2	**C-reactive protein, pentraxin-related**
NP_000574.2	vitamin D-binding protein precursor
NP_000578.2	complement component 7 precursor
NP_000597.2	complement component 8, gamma polypeptide
NP_000598.2	orosomucoid 1 precursor
NP_000599.1	orosomucoid 2
NP_000604.1	hemopexin

NP_000629.3	vitronectin precursor
NP_000706.1	complement component 4 binding protein, alpha chain
NP_000884.1	kininogen 1 isoform 2
NP_000925.2	alpha-2-plasmin inhibitor
NP_001002029.3	complement component 4B preproprotein
NP_001007526.2	NACHT and WD repeat domain containing 1
NP_001035519.1	protocadherin 17 precursor
NP_001054.1	transferrin
NP_001073871.1	coiled-coil domain containing 61
NP_001073982.1	**carboxypeptidase N, polypeptide 2, 83kD**
NP_001076.2	serpin peptidase inhibitor, clade A, member 3 pre
NP_001092877.1	hypothetical protein LOC55184
NP_001107854.1	tuberous sclerosis 2 isoform 5
NP_001121134.1	gelsolin isoform b
NP_001121179.1	serine proteinase inhibitor, clade A, member
NP_001124.1	afamin precursor
NP_001129973.1	**zinc finger protein 844**
NP_001138422.1	hypothetical protein LOC284297
NP_001139.3	ankyrin 2 isoform 1
NP_001176.1	alpha-2-glycoprotein 1, zinc
NP_001613.2	alpha-2-HS-glycoprotein
NP_001624.1	alpha-1-microglobulin/bikunin preproprotein
NP_001630.1	serum amyloid P component precursor
NP_001634.1	apolipoprotein A-II preproprotein
NP_001638.1	apolipoprotein D precursor
NP_001701.2	complement factor B preproprotein
NP_001724.3	complement component 1, r subcomponent
NP_001726.2	complement component 5 preproprotein
NP_001728.1	complement component 9 precursor
NP_001804.2	centromere protein E
NP_001822.2	clusterin isoform 1
NP_002069.2	golgi autoantigen, golgin subfamily a, 4
NP_002075.2	**glutathione peroxidase 3 precursor**
NP_002105.2	human immunodeficiency virus type I enhancer
NP_002206.2	inter-alpha (globulin) inhibitor H1
NP_002207.2	inter-alpha globulin inhibitor H2 polypeptide
NP_002208.3	inter-alpha (globulin) inhibitor H3 preproprotei
NP_002209.2	inter-alpha (globulin) inhibitor H4
NP_002263.2	keratin 4
NP_002266.2	keratin 15
NP_002606.3	serine (or cysteine) proteinase inhibitor, clade
NP_002955.2	S100 calcium-binding protein A8
NP_002956.1	S100 calcium-binding protein A9
NP_003269.2	C-type lectin domain family 3, member B
NP_003310.3	titin isoform N2-B
NP_004039.1	beta-2-microglobulin precursor
NP_004229.1	thyroid hormone receptor interactor 12
NP_004961.1	insulin-like growth factor binding protein, acid l
NP_005132.2	fibrinogen, beta chain preproprotein

NP_005134.1	haptoglobin isoform 1 preproprotein
NP_005546.2	keratin 6B
NP_005568.2	lipoprotein Lp(a) precursor
NP_006112.3	keratin 1
NP_006253.2	peripherin
NP_006503.1	**serum amyloid A4, constitutive**
NP_006653.2	Snf2-related CBP activator protein
NP_006735.2	retinol-binding protein 4, plasma precursor
NP_009117.2	centrosomal protein 2 isoform 1
NP_009224.2	complement component 4A preproprotein
NP_036192.2	dynamin 1-like protein isoform 1
NP_036375.1	solute carrier family 35 member 3A
NP_055542.1	myeloid/lymphoid or mixed-lineage leukemia 4
NP_055704.2	metalloprotease 1
NP_055836.1	zinc finger protein 292
NP_055861.3	senataxin
NP_055907.3	PI-3-kinase-related kinase SMG-1
NP_059672.2	MAP/microtubule affinity-regulating kinase 2 isoform
NP_060719.4	CDK5 regulatory subunit associated protein 2 isoform
NP_065076.2	akt substrate AS250
NP_066275.3	haptoglobin-related protein
NP_079052.2	NKF3 kinase family member
NP_110381.2	serum amyloid A2 isoform a
NP_443204.1	**leucine-rich alpha-2-glycoprotein 1**
NP_473375.2	fibronectin 1 isoform 7 preproprotein
NP_510880.2	myosin 18A isoform a
NP_570602.2	alpha 1B-glycoprotein precursor
NP_653247.1	immunoglobulin J chain
NP_705694.2	keratin 13 isoform a
NP_835260.2	PDZ domain containing 2
NP_853512.1	keratin 25
NP_938073.1	kelch-like 17
NP_954630.1	serum amyloid A1 preproprotein
NP_958850.1	complement component 1, s subcomponent
NP_997639.1	fibronectin 1 isoform 6 preproprotein
XP_001714007.1	PREDICTED: similar to 22kDa peroxisomal membrane
XP_001719048.1	PREDICTED: similar to hCG2042722
XP_001719515.1	PREDICTED: hypothetical protein, partial
XP_001724477.1	PREDICTED: similar to kappa immunoglobulin
XP_001724484.1	PREDICTED: hypothetical protein

Table 4.9: Identified, differentially expressed proteins, GeLC-MS/MS study

Legend: Summary of the six differentially expressed proteins identified. UniProtKB reference refers to the Universal Protein Resource database and is a unique protein identifier. The number of unique peptides used in the identification of the protein, the expected MW of the protein, the gel region in which I detected the protein in maximal abundance, the fold change between in CC compared with gallstone disease cohorts are listed. The p-value relates to the t-test between CC and gallstone disease groups.

Approved protein name	UniProtKB	Unique peptides	MW (kDa)	Gel region (kDa)	Fold change (CC)	P-value
Carboxypeptidase N	CPN2_HUMAN	4	58.3	95-110	0.77	0.02
C-reactive protein	CRP_HUMAN	7	25	24-30	2.9	0.009
Glutathione peroxidase 3	GPX3_HUMAN	4	25.5	24-30	2.0	0.001
Leucine-rich α-2-glycoprotein 1	A2GL_HUMAN	2	38.2	53-60	2.2	0.002
Serum amyloid A4	SAA4_HUMAN	2	14.8	18-22	1.7	0.002
Zinc finger protein 844	ZN844_HUMAN	2	77	250-276	2.2	<0.0001

Table 4.10: Results of ROC-AUC and PCA analyses, GeLC-MS/MS study

Legend: The utility of each identified protein as a diagnostic marker was assessed. The ROC-AUC value is shown, along with 95% confidence intervals and the ROC-AUC derived p-value. In the principal components analysis (PCA) only two of the proteins were required to reach the optimum model. The β-value of each of these is included in the final column.

Protein name	ROC-AUC	95% CI	p-value	β-value
Carboxypeptidase N	0.87	0.69-1.0	0.015	-
C-reactive protein	0.91	0.76-1.0	0.008	-
Glutathione peroxidase 3	0.98	0.93-1.0	0.002	0.363
Leucine-rich alpha-2-glycoprotein 1	0.93	0.80-1.0	0.005	-
Serum amyloid A4	0.93	0.81-1.0	0.005	-
Zinc finger protein 844	0.98	0.93-1.0	0.002	0.626

Table 4.11: Results of database search for known function and disease associations

Legend: Summary of database and literature review. UniProtKB v61 was interrogated for each of the six identified proteins. PubMed was searched for any combination of each proteins name and the word 'cancer'.

Protein	Subcellular location	Known function	Known disease associations
Carboxypeptidase N	Secreted	The 83 kDa subunit binds and stabilizes the catalytic subunit at 37 degrees Celsius and keeps it in circulation. Under some circumstances it may be an allosteric modifier of the catalytic subunit. Important in the regulation of peptides like kinins and anaphylatoxins	Rare cases of hereditary angioedema
C-reactive protein	Secreted	Promotes agglutination, bacterial capsular swelling, phagocytosis and complement fixation through its calcium-dependent binding to phosphorylcholine. Can interact with DNA and histones and may scavenge nuclear material released from damaged circulating cells.	Acute phase response to tissue injury, or other inflammatory stimuli. Non specific elevation in some malignancies
Glutathione peroxidase 3 precursor	Secreted	Protects cells and enzymes from oxidative damage, by catalyzing the reduction of hydrogen peroxide, lipid peroxides and organic hydroperoxide, by glutathione.	**Low** levels have been described in a number of malignancies
Leucine-rich alpha-2-glycoprotein 1	Secreted	Has been shown to be involved in protein-protein interaction, signal transduction, and cell adhesion and development. LRG1 is expressed during granulocyte differentiation	High in acute appendicitis, several malignancies – including pancreatic Ca and CC.
Serum amyloid A4	Secreted	Apolipoprotein of the HDL complex. Expressed by the liver, secreted into plasma	Major acute phase reactant. Non specifically elevated in some cancers
Zinc finger protein 844	Nucleus	May be involved in transcriptional regulation	-

Table 4.12: Subject demographics, discovery SELDI-TOF MS study

Legend: p-values represent the result of χ^2 testing of age and ethnicity, K-W ANOVA for age.

	CC	PSC	HC	P-value
n	18	10	17	
Male n (%)	12 (67%)	9 (90%)	9 (47%)	0.14
Age (IQ range)	68 (56-74)	39 (30-59)	42 (38-64)	<0.01
Ethnicity				
Caucasian (%)	15 (83%)	10 (100%)	14 (82%)	0.37
Other	3	0	3	0.37
Asian	2	0	3	
African	1	0	0	
Unknown	0	0	0	
Institution				
ICH (%)	9 (50%)	0	9 (50%)	-
UCH (%)	9 (53%)	0	8 (47%)	-
Norway	0	10 (100%)	0	-

Table 4.13: Principal results of SELDI-TOF MS experiments.

Legend: P-values from the initial discovery (1^o) study for each significantly different m/z peak are shown first. P-values from the new cases analysed in the validation cohort are then shown. P-values from the repeated analysis of the discovery cohort (2^o) are then shown. Where possible, subgroup analyses of UCH and ICH cohorts are shown. Comparisons between UK and Norwegian PSC samples are also shown. CC – cholangiocarcinoma cohort, PSC – PSC cohort, HC – healthy control cohort, SI – signal intensity, rSI – relative signal intensity.

Cohort	Comparison	4297/4291	4630/4631	5768/5765	8577/8572	8701/8699	11520/11526	17313/17287
1^o analysis	CC vs PSC (p)	0.046	0.036	0.317	0.0176	0.049	0.209	0.022
discovery cohort	CC vs HC (p)	0.054	0.001	0.038	0.0033	0.023	0.034	0.169
Validation	Mean SI CC (rSI)	17.15	2.273	3.032	22.1	36.0	0.850	0.441
cohort	Mean SI PSC (rSI)	17.62	0.899	1.731	25.3	42.4	0.252	0.639
	Mean SI HC (rSI)	25.49	0.750	1.923	35.9	48.6	0.147	0.838
	CC vs PSC (p)	0.785	$9.7x10^{-7}$	$2.6x10^{-4}$	0.113	0.03	$2.6x10^{-5}$	0.018
	CC vs HC (p)	$1.2x10^{-7}$	$1.3x10^{-11}$	0.001	$9.4x10^{-13}$	$1.4x10^{-7}$	$4.2x10^{-7}$	$9.8x10^{-8}$
	PSC vs HC (p)	$4.5x10^{-6}$	0.024	0.25	$1.3x10^{-7}$	0.0168	0.029	0.03
	Nor PSC vs UK PSC (p)	0.461	0.46	0.37	0.56	0.75	0.34	0.30
	ICH CC vs ICH HC (p)	$5.0x10^{-4}$	$3.6x10^{-6}$	0.007	$2.0x10^{-7}$	$4.5x10^{-4}$	$7.5x10^{-5}$	$3.6x10^{-5}$
	UCH CC vs UCH HC (p)	$7.0\ x10^{-4}$	$1.2x10^{-7}$	0.354	$1.5x10^{-4}$	0.006	0.028	0.014
2^o analysis	CC vs PSC (p)	0.1747	0.02	0.266	0.203	0.58	0.19	0.54
discovery cohort	CC vs HC (p)	$3.0x10^{-4}$	0.001	0.104	$5.9x10^{-4}$	0.016	0.044	0.129
Summary		**↓CC, ↓PSC**	**↑CC, ↑PSC**	**↑CC**	**↓CC and ↓PSC**	**↓CC, ↓PSC**	**↑CC, ↑PSC**	**↓CC and ↓PSC**

Table 4.14: Subject demographics, validation SELDI-TOF MS study

Legend: p-values represent the result of χ^2 testing of age and ethnicity, K-W ANOVA for age.

	CC	PSC	HC	p-value
n	81	54	90	
Male n (%)	46 (57%)	41 (76%)	48 (53%)	0.02
Age (IQ range)	66 (58-78)	42 (31-51)	42 (33-55)	<0.001
Ethnicity				
Caucasian (%)	53 (65%)	48 (89%)	51 (57%)	<0.001
Other	28	6	39	
Asian	*16*	*5*	*36*	
African	*8*	*0*	*3*	
Unknown	*4*	*1*	*0*	
Institution				
ICH (%)	48 (59%)	1 (2%)	65 (72%)	<0.001
UCH (%)	33 (41%)	12 (22%)	25 (28%)	<0.001
Norway (%)	0	41 (76%)	0	-

Table 4.15 Matching of discovery *m/z* peak values with validation *m/z* peak values

Legend: The calculated permissible +/-0.15% range is shown for each peak from the discovery study. Δ – actual difference between *m/z* peak identified in discovery study and its matched partner in validation study.

m/z discovery	**4297**	**4630**	**5768**	**8577**	**8701**	**11520**	**17313**
+/- 0.15% (Da)	+/-6	+/-6	+/-9	+/-13	+/- 13	+/-18	+/- 26
***m/z* validation**	4291	4631	5765	8572	8699	11526	17287
Δ (Da)	6	1	3	5	2	6	26
Δ (%)	0.14%	0.02%	0.05%	0.06%	0.02%	0.05%	0.15%

Table 4.16: Subject demographics, combined SELDI-TOF MS cohorts

Legend: p-values represent the result of χ^2 testing of age and ethnicity, K-W ANOVA for age.

	CC	PSC	HC	p-value
n	99	64	107	
Male n (%)	58 (59%)	50 (78%)	57 (53%)	0.005
Age (IQ range)	68 (55-74)	42 (31-52)	42 (34-57)	<0.001
Ethnicity				
Caucasian (%)	68 (69%)	58 (91%)	65 (61%)	<0.001
Other	31	6	42	
Asian (%)	*18 (18%)*	*5 (8%)*	*39 (36%)*	
African	*9 (9%)*	*0*	*3 (3%)*	
Unknown	*4 (4%)*	*1 (1%)*	*0*	
Institution				
ICH (%)	57 (58%)	1 (2%)	74 (69%)	<0.001
UCH (%)	42 (42%)	12 (18%)	33 (31%)	<0.001
Norway (%)	0	51 (80%)	0	

Table 4.17: Routine laboratory indices, combined SELDI-TOF MS CC cohort

Legend: Routine laboratory values were not available for healthy control subjects. Where adequate number of values permits, p-value from a M-W U test between CC and PSC groups is shown. NS – not significant.

	CC	PSC	p-value
Ur (mmol/L)	5.9 (3.9-8)	5.2 (-)	-
Cr (µmol/L)	70 (62-81)	82 (-)	-
Bili (µmol/L)	20 (10-77)	24 (12-41)	NS
ALP (IU/L)	263 (143-607)	217 (124-395)	NS
ALT (IU/L)	46 (24-80)	91 (52-131)	<0.00001
Alb (g/L)	30 (27-35)	40 (36-43)	<0.00001
CRP (mg/L)	38 (14-83)	3.9 (1.7-11)	<0.00001
WCC (x10^9/L)	6.6 (5.5-10.5)	6.3 (4.8-7.7)	NS
AFP (ng/mL)	5 (2.5-5.5)	4 (3-5)	NS
Cea (µg/L)	3 (1.9-10.5)	1.45 (1.2-2.0)	0.00200
CA19-9 (IU/mL)	216 (24-1000)	22 (17-39)	0.00040

Table 4.18: Anatomical location of tumour, combined SELDI-TOF MS CC cohort

Legend: Breakdown of the 99 CC subjects by site of primary biliary malignancy. % is of total where anatomical location known.

Site of disease	n (%)
Intrahepatic CC	10 (11%)
Hilar CC	59 (66%)
Distal CC	20 (23%)
Total known	**89**
Indeterminate location CC	10
Grand total	**99**

Table 4.19: ROC-AUC analysis of 7 peaks of interest

Legend: P-value derived from ROC-AUC analysis. Sens – sensitivity, spec - specificity

Peak	4291	4631	5765	8572	8699	11526	17287
P-value CC/PSC	0.7851	<0.001	<0.001	0.1139	<0.001	<0.001	0.018
AUC-ROC	0.55	0.76	0.63	0.61	0.69	0.73	0.61
Sens/Spec (%)	65/51	83/57	63/52	65/51	70/51	79/56	63/56
P-value CC/HC	<0.001	<0.001	<0.001	< 0.001	<0.001	<0.001	<0.001
AUC-ROC	0.76	0.82	0.59	0.82	0.74	0.83	0.72
Sens/spec (%)	88/54	89/57	65/49	91/63	89/54	83/64	75/56

Table 4.20: Correlation matrix of the seven significant peaks

Legend: Values in matrix are R^2 for the correlation between the pair of m/z peaks. Values in in the matrix in bold text represent significant correlation ($p<0.05$)

m/z	4291	4631	5765	8572	8699	11526	17287
4291	1						
4631	0.03	1					
5765	0.01	0.01	1				
8572	**0.89**	0.01	0.03	1			
8699	**0.51**	0.01	0.01	**0.52**	1		
11526	0.00	0.00	**0.70**	0.00	0.00	1	
17287	0.06	0.00	0.00	0.06	0.06	0.01	1

Table 4.21: Correlation matrix of the seven significant peaks and routine laboratory indices

Legend: Number in matrix represent R^2 values. Pairwise exclusion of missing data.

	4291	4631	5765	8572	8699	11526	17287
Ur	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Cr	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bili	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Alp	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Alt	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Alb	0.01	0.05	0.05	0.02	0.02	0.07	0.00
CRP	0.00	0.04	0.15	0.01	0.00	0.20	0.00
WCC	0.00	0.00	0.01	0.00	0.00	0.02	0.00
AFP	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cea	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CA199	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 4.22: Summary of multiple linear regression analysis of the seven significant peaks

Legend: Settings: F to include 3.5, F to exclude 3.4, max number of steps 100.

Peak (m/z)	B-value	F to exclude	No. of steps	ROC-AUC (95% CI)
CC vs. PSC			4	0.76 (0.68-0.82)
11526	.236	9.97179		
17287	-.17	5.19979		
4631	.141	3.51676		
CC vs. HC			6	0.90 (0.85-0.94)
8572	-.91	27.33		
4291	.563	10.99		
11526	.793	18.38		
5765	-.62	11.81		
4631	.119	4.46		

Table 4.23: Analysis of six subjects with CC *and* PSC

Legend: Limited analysis of six additional subjects with CC *and* PSC. Comparison is only between these six cases and 64 PSC only subjects in the three peaks of principal interest in discriminating CC from PSC in the principal study.

Peak	MW-U test p-value	ROC-AUC	ROC 95% CI	ROC p-value
4631	0.01	0.82	0.70-0.94	0.01
11526	0.10	0.70	0.50-0.94	0.10
17287	0.44	0.60	0.37-0.82	0.44

Table 4.24: Results of *m/z* database search for seven peaks of interest, SELDI-TOF MS study

Legend: Each of the seven *m/z* peak values, +/- 0.1%, was searched in turn in TagIdent (ExPASy, Swiss institute of bioinformatics). The number of protein hits for each m/z peak is give, the UniProt code, approved name and consensus molecular weight of the protein. The classification of each candidate protein, its known function and known disease associations are listed. Candidate proteins empirically considered of particular logical interest are highlighted in bold.

m/z	↑/↓ in CC	UniProt Code	Approved name	MW	Classification	Known function	Known disease association
4291	↓	EDN1_HUMAN	Big endothelin-1	4273	Secreted	Endothelium-derived vasoconstrictor peptide	-
		MICB_HUMAN	**Isoform 3 of MHC class I polypeptide-related sequence**	**4276**	**Cell membrane**	**Stress-induced s elf-antigento T cells**	**Released from epithelial tumours, inc HCC**
		ADM2_HUMAN	Intermedin-short (Potential)	4287	Secreted hormone	Regulation of gastrointestinal function	
		DSCR8_HUMAN	Isoform 4 of Down syndrome critical region protein 8	4287	Membrane	-	-
4631	↑	**CCKN_HUMAN**	**Cholecystokinin-39**	**4626**	**Secreted, hormone**	**Induces gall bladder contraction**	**-**
		LST1_HUMAN	Isoform 7 of Leukocyte-specific transcript 1 protein	4629	Cellular membrane	Modulating B cell immune response	-
		RBMS1_HUMAN	Isoform 4 of RNA-binding motif, single-stranded-interacti	4632	Nucleus	DNA replication	-
5765	↑	PRB2_HUMAN	Basic proline-rich peptide IB-7	5737	Secreted	-	-
		KTAS1_HUMAN	Putative uncharacterized protein KTN1- AS1	5753	Dubious prediction	-	-
8572	↓	SAPL1_HUMAN	Saposin B-Val-like (By similarity).	8569	Secreted	Lysosomal degradation of sphingolipids	-
		LZTS1_HUMAN	**Isoform 7 of Leucine zipper putative tumour suppressor 1**	**8571**	**Cytoplasmic**	**Tumour supressor**	**-**
		U2AF1_HUMAN	Isoform 3 of Splicing factor U2AF 35 kDa subunit	8572	Nuclear	Mediating protein-RNA interactions	-
		NARG2_HUMAN	Isoform 2 of NMDA receptor-regulated protein 2	8572	Nuclear	-	-
		BAALC_HUMAN	Isoform 4 of Brain and acute leukemia cytoplasmic protein	8573	Cytoplasmic	interacting with CAMK2A	Acute leukaemia
		IGLL1_HUMAN	Isoform 2 of Immunoglobulin lambda-like polypeptide 1	8574	Secreted	Critical for B-cell development	Agammaglobulinemia type 2 (AGM2)
		SAA1_HUMAN	**Amyloid protein A**	**8575**	**Secreted**	**Major acute phase reactant**	**Amyloidosis**
		UBB_HUMAN	Ubiquitin	8575	Cytoplasm/	Regulation of protein trafficking	In tissue of neurodegenerative disease
		ISK8_HUMAN	Serine protease inhibitor Kazal-type 8.	8579	**Secreted**	Serine protease inhibitor	**-**
		APOA2_HUMAN	Truncated apolipoprotein A-II.	8580	Secreted	HDL metabolism	-
8699	↓	PAOX_HUMAN	I12 of Peroxisomal N(1)-acetyl-spermine/spermidine	8685	Cytoplasm	Polyamine conversion	Low level in neoplastic tissues
		CC078_HUMAN	UPF0640 protein C3orf78	8695	Membrane	-	
		SPIT4_HUMAN	Kunitz-type protease inhibitor 4	8696	Secreted	-	-
11526	↑	LV101_HUMAN	Ig lambda chain V-I region VOR	11515	Immunoglobulin	-	This is a Bence-Jones protein

		COBL1_HUMAN	Isoform 5 of Cordon-bleu protein-like 1	11516	Not known	-	-
		RT16_HUMAN	28S ribosomal protein S16, mitochondrial	11516	Mitochondrial	-	COXPD2): fatal neonatal metabolic acidosis
		LV403_HUMAN	Ig lambda chain V-IV region Hil.	11517	Immunoglobulin	-	Myeloma
		MRAP_HUMAN	Isoform 2 of Melanocortin-2 receptor accessory protein	11519	Cytoplasmic	Intracellular trafficking pathways adipocytes	-
		CHCH7_HUMAN	Isoform 2 of Coiled-coil-helix-coiled-coil-helix domain	11520	Mitochondrial	-	Benign epithelial tumours of the salivary gland
		VPRE3_HUMAN	Pre-B lymphocyte protein 3	11520	Immunoglobulin	B cell maturation	-
		LV107_HUMAN	Ig lambda chain V-I region BL2.	11522	Immunoglobulin	Complement activation	-
		ATP7A_HUMAN	Isoform 6 of Copper-transporting ATPase 1	11522	Golgi membrane	-	Menkes disease (curly hair syndrome)
		SAA1_HUMAN	**Serum amyloid protein A(2-104)**	**11527**	**Secreted**	**-**	**Major acute phase reactant**
		CD28_HUMAN	Isoform 2 of T-cell-specific surface glycoprotein	11528	Membrane	T-cell activation	-
		ENY2_HUMAN	Enhancer of yellow 2 transcription factor homolog	11529	Nucleus	Transcription factor	-
		CREM_HUMAN	Isoform 20 of cAMP-responsive element modulator	11535	Nucleus	Spermatogenesis	-
17287	↓	BT3L4_HUMAN	Transcription factor BTF3 homolog 4	17271	Nucleus	Transcription factor	-
		XG_HUMAN	Glycoprotein Xg.	17272	Cell membrane	Xg blood group system	-
		3HAO_HUMAN	Isoform 2 of 3-hydroxyanthranilate 3,4-dioxygenase	17275	Cytoplasm	Fe^{2+} ion synthesis	-
		YV012_HUMAN	Isoform 2 of Leucine-rich repeat-containing protein LOC40	17279	Cytoplasm	-	-
		NTCP7_HUMAN	Isoform 5 of Sodium/bile acid cotransporter 7	17283	Membrane	Ion transport	-
		MSRB2_HUMAN	Methionine-R-sulfoxide reductase B2, mitochondrial.	17288	Mitochondrial	Reduction of methionine sulf.	-
		FXYD5_HUMAN	**FXYD domain-containing ion transport regulator 5**	**17292**	**Membrane**	**Regulation of E-cadherin**	**Promotes metastases**
		TMM80_HUMAN	Isoform 2 of Transmembrane protein 80	17292	Membrane	-	-
		MFAP5_HUMAN	Microfibrillar-associated protein 5	17294	**Secreted**	Component of the elastin-associated microfibrils	-
		CX038_HUMAN	Isoform 3 of Uncharacterized protein CXorf38	17295	Uncharacterised	-	-
		TTC32_HUMAN	Tetratricopeptide repeat protein 32	17296	Repeat peptide	-	-

Table 4.25: Summary of *m/z* peaks previously identified as of interest in SELDI-TOF MS studies of HPB cancer

Legend: Peaks of interest are listed and highlighted in bold where the study analysis identified it to be of particular interest. Panc Ca – pancreatic cancer.

1st author[ref]	Disease	Year	Country	Journal	Chip type	Peaks of interest	Comment
Scarlett[360]	CC	2006	Australia	Hepatology	H50	**4462**, 4560, 5382, 17266,4462, 4560, 5382, 5660, 5947, 6433, 6632, 6841, 8528, 8691, **11535**, 11697	-
Liu[399]	CC	2008	China	J cell biochem.	Not known	**1376, 1388, 1404**	All 3 identified as transthyretin
Navaglia[400]	Panc Ca	2009	Italy	Clin Chem Lab Med.	Imac 30	1526, 1211, 3519	-
Zinkin[401]	HCC	2008	USA	Clin Cancer Res.	CM10, IMAC30 and H50	**13391**	ID: cystatin C
Okabe[402]	IHCC	2012	Japan	Cancer Research	-	**8360**	Cell culture study. ID: epithelial cell-derived neutrophil-activating peptide-78
Wang[403]	CC	2009	China	Eur J Cancer Care	IMAC30	**28078**	ID: apolipoprotein A-I (ApoA-I)
Wang	CC	2010	China	Cancers	IMAC30	2170, 2967, 3300 3400, 4188, 4503, 4906, 5630, 5681, 7598, 7797, 10875 11167, 11242	NB: Not on PubMed
Kikkawa[404]	CC	2012	Japan	International Journal of Proteomics	Weak ion exchange	**4204**	Used MALDI. ID: prothrombin fragement
Qian[405]	Panc Ca	2012	China	Asian Pacific J Cancer Prev	Weak ion beads, then plain gold chips	**7762, 8560, 11654**	-

Table 4.26: Peaks from SELDI-TOF MS study matched with *m/z* peaks previously associated with CC or pancreatic CC in SELDI studies.

Legend: Matched peaks fell within 0.15% of each other. Panc Ca – pancreatic cancer.

My peak	↑/↓ in CC	Matched	↑/↓ in CC	Paper	Disease	m/z protein ID?
8572	↓	8560	↑	Qian[405]	Panc Ca	No
8699	↓	8691	↓	Scarlett[360]	CC	No
11526	↑	11535	↑	Scarlett	CC	No
17287	↓	17266	↓	Scarlett	CC	No

Figure 4.1: Standard workflow for GeLC-MS/MS studies

Legend: Adapted from Griffiths *et al* 2009[370]

Figure 4.2: Overview of label-free proteomics (I)

Legend: Schematic of 1D SDS-PAGE and proteolytic digestion. Adapted from images provided courtesy of Dr V Horneffer-van der Sluis.

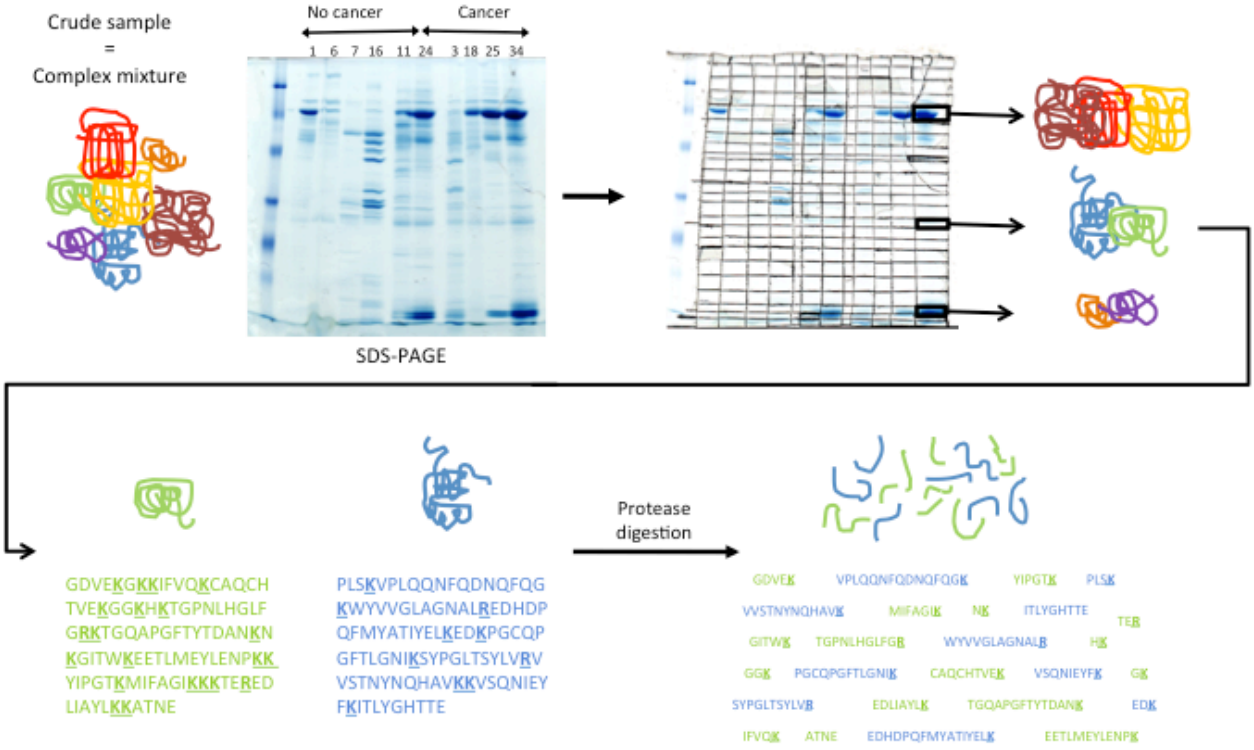


Figure 4.3: Overview of label-free proteomics (II)

Legend: LC-MS/MS to protein identification. Adapted from images provided courtesy of Dr V Horneffer-van der Sluis.

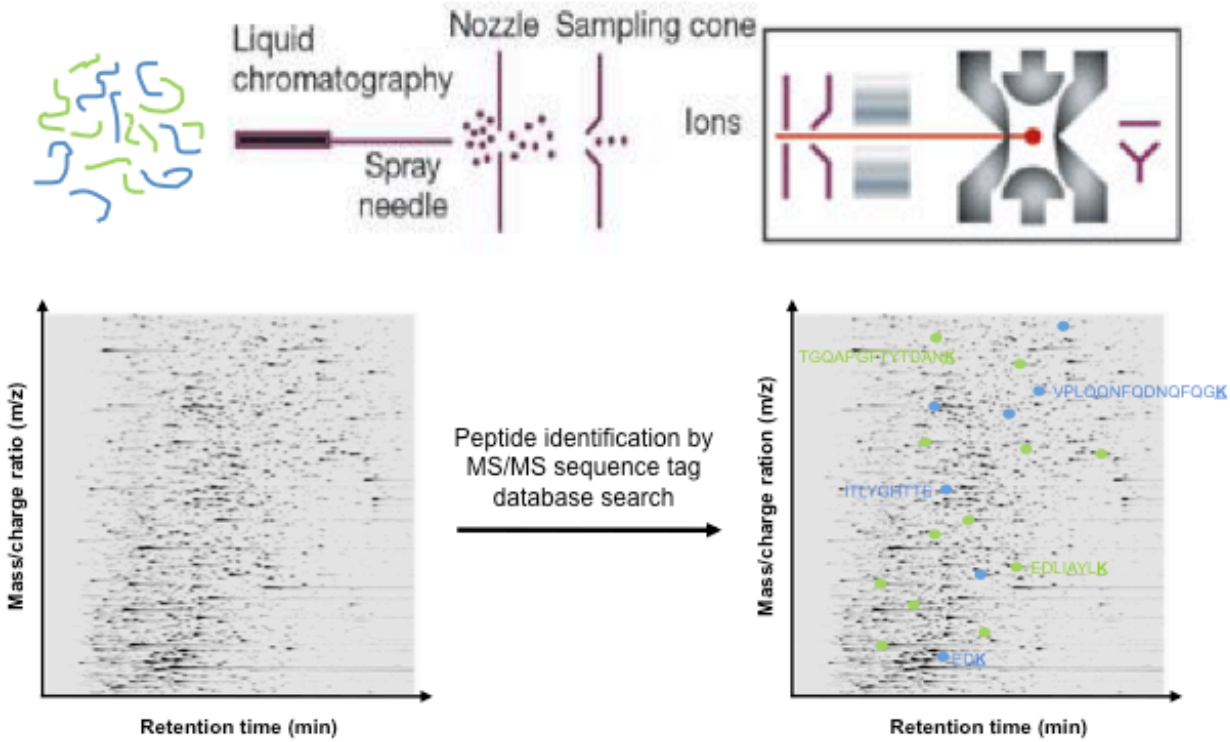


Figure 4.4: Illustration of available ProteinChip® active surfaces available for SELDI-TOF MS

Legend: The crude plasma sample is placed onto the ProteinChip spot. Proteins bind to the chemical or biological 'docking sites' on the surface. Adapted from images provided courtesy of Dr V Horneffer-van der Sluis.

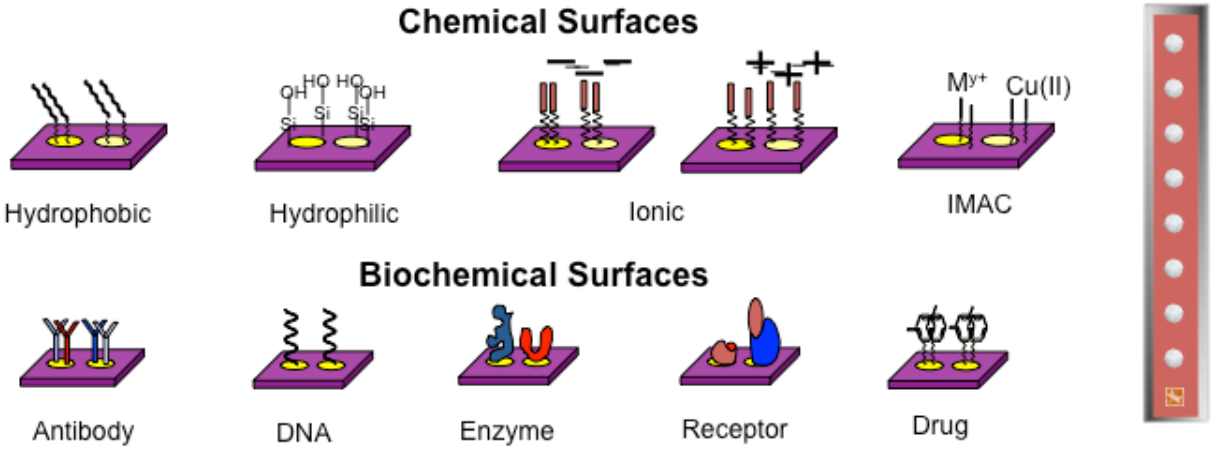


Figure 4.5: CM10 and H50 chip surface interactions

Legend: The properties of the two surfaces that I selected for use in my SELDI-TOF MS experiments. (A) CM10. An anionic surface with carboxylate groups that interact with positively charged residues on proteins (e.g. lysine, arginine, histidine). (B) H50. A hydrophobic surface that binds to hydrophobic residues on the protein (e.g. alanine, valine, leucine, isoleucine, phenylalanine, tryptophan, tyrosine). Adapted from images provided courtesy of Dr V Horneffer-van der Sluis.

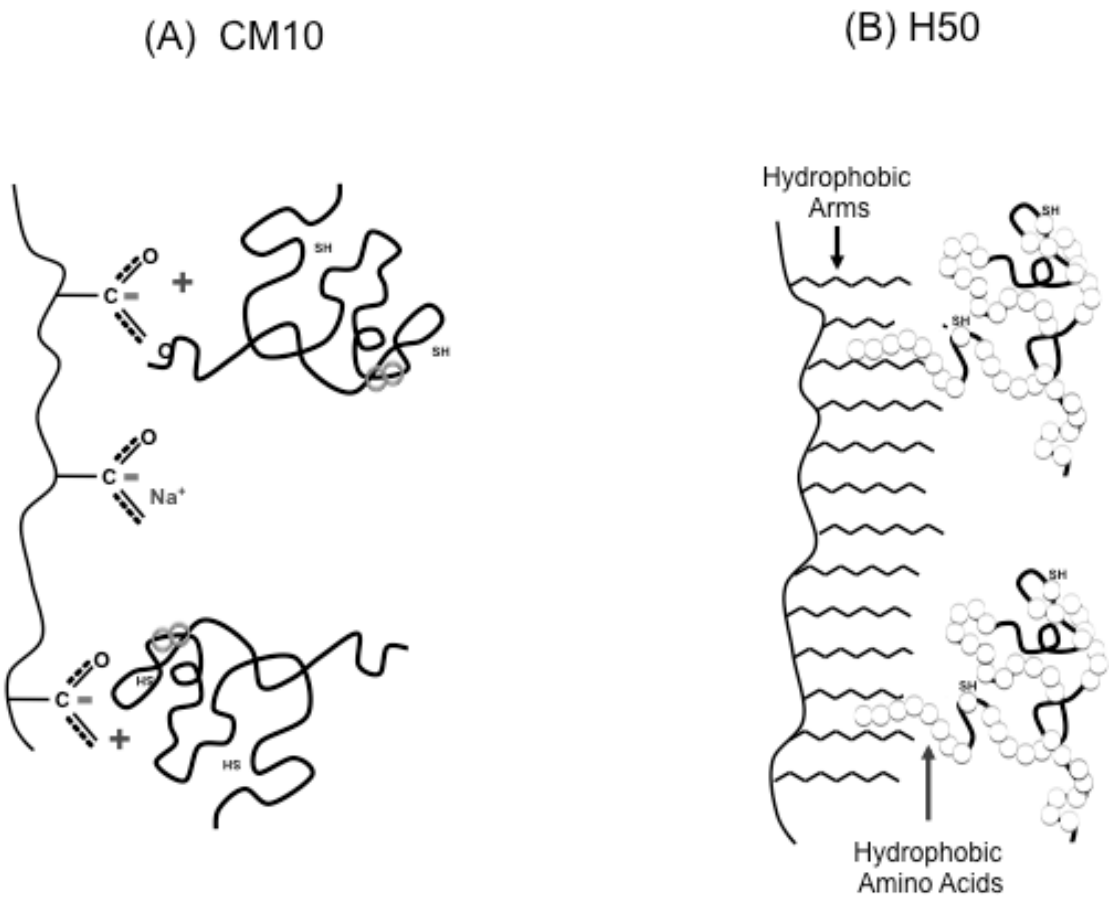


Figure 4.6: Schematic diagram of the Ciphergen® SELDI-TOF mass spectrometer

Legend: Showing the chip surface (Y), application of a laser and TOF MS measuring the molecular weights of the various proteins that are vaporised from the chip surface. The *m/z* ratios are represented in the spectra obtained. (Adapted from Issaq *et al* 2002)[406]

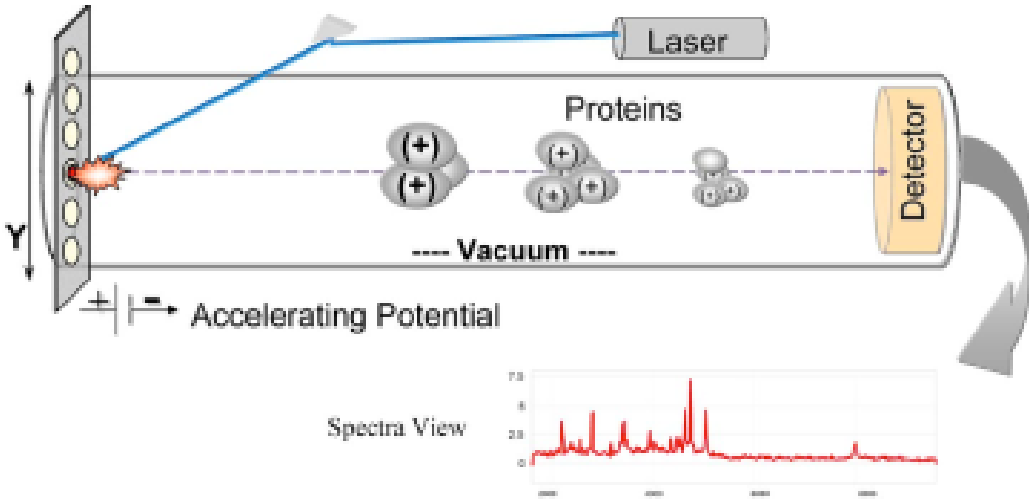


Figure 4.7: Workflow for protein depletion using the IgY12 Seppro Column

Legend: The 12 most abundant human plasma proteins are bound within the column, depleting the levels remaining within the plasma sample.

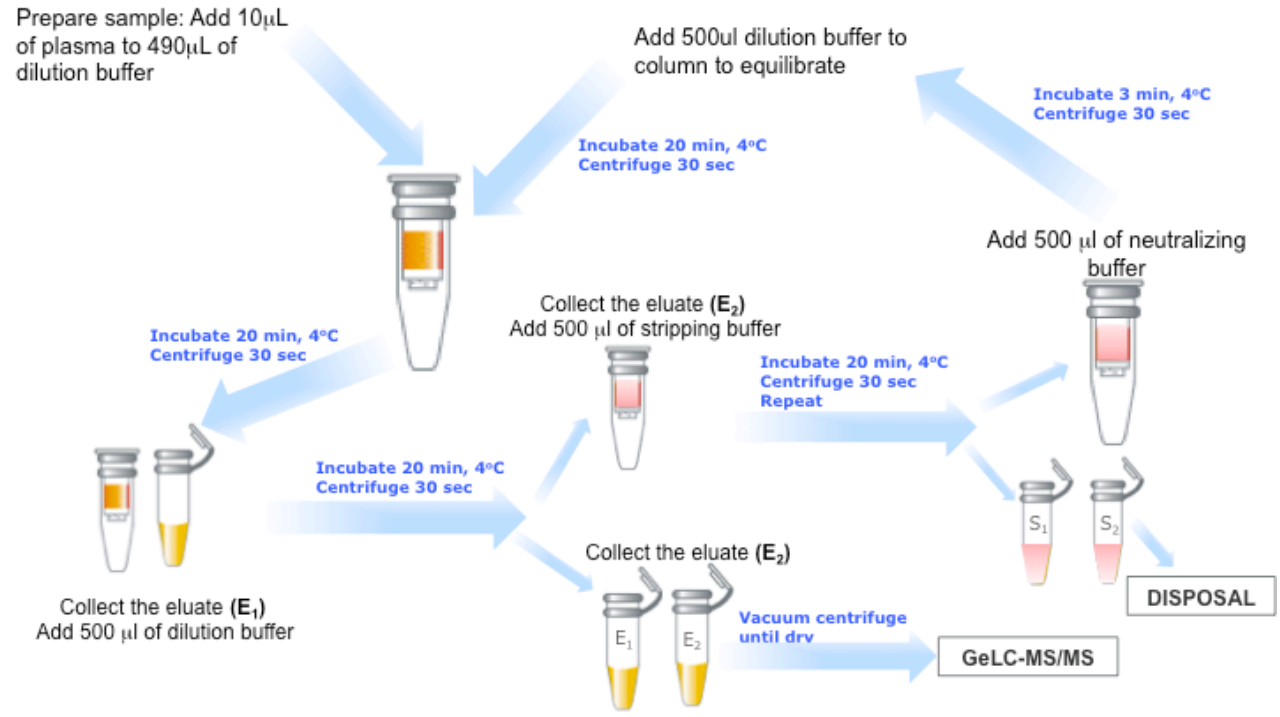


Figure 4.8: Layout of gels, GeLC-MS/MS study

Legend: Mrker – marker (protein standard (PageRuler, Fermentas, UK)). * - empty lane.

Gel 1

Lane	1	2	3	4	5	6	7	8	9	10
Sample	Mrkr	BB5	BB6	BB7	*	CC6	CC8	CC9	CC10	Mrkr
Volume	5µL	20µL	20µL	20µL	*	20µL	20µL	20µL	20µL	5µL

Gel 2

Lane	1	2	3	4	5	6	7	8	9	10
Sample	Mrkr	BB1	BB2	BB3	BB4	CC1	CC2	CC4	CC5	Mrkr
Volume	5µL	20µL	20µL	20µL	20µL	20µL	20µL	20µL	20µL	5µL

Figure 4.9: Gel run to test efficacity of IgY depletion process

Legend: Lanes 1 and 10 contain protein standard (PageRuler, Fermentas, UK). Paired samples of raw and depleted from 2 CC and 2 benign subjects were run in lanes 2-9. Substantial depletion of the albumin band (arrowed) can be seen.

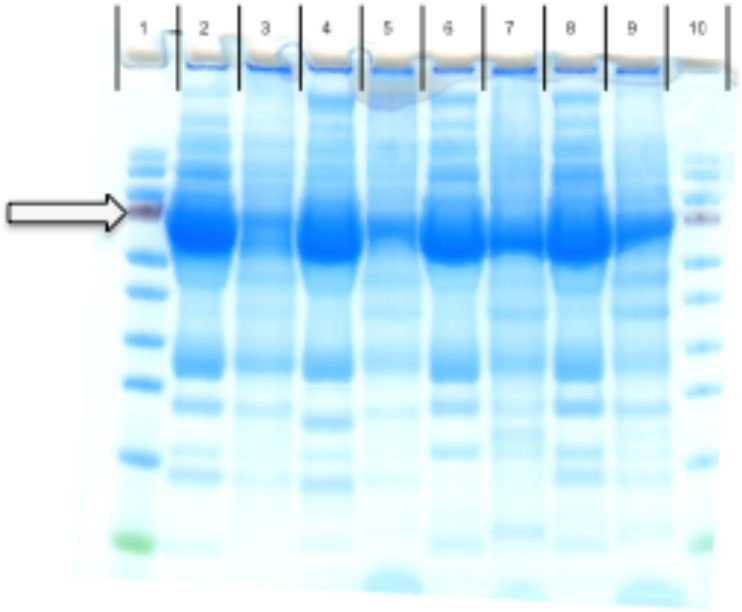


Figure 4.10: Image of gels from GeLC-MS/MS study

Legend: Samples were applied according to the plan in Figure 4.8. Planned cut up is demarcated according to protein standard lines in lanes 1 & 10.

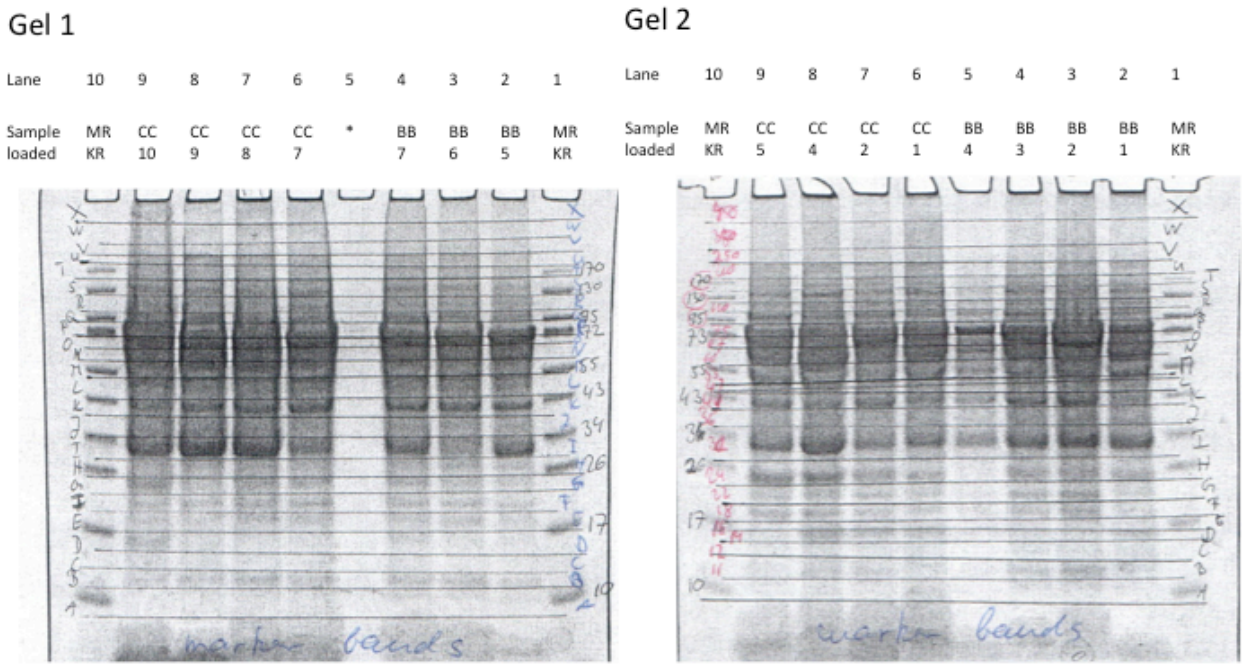


Figure 4.11: Representative results from preliminary SELDI-TOF MS CM10 condition optimisation experiments

Legend: The same sample has been applied to multiple spots on CM10 chips. Different combinations of pH and energy absorbing matrix (EAM) have been used, permitting comparison of spectral peak resolution and signal intensity.

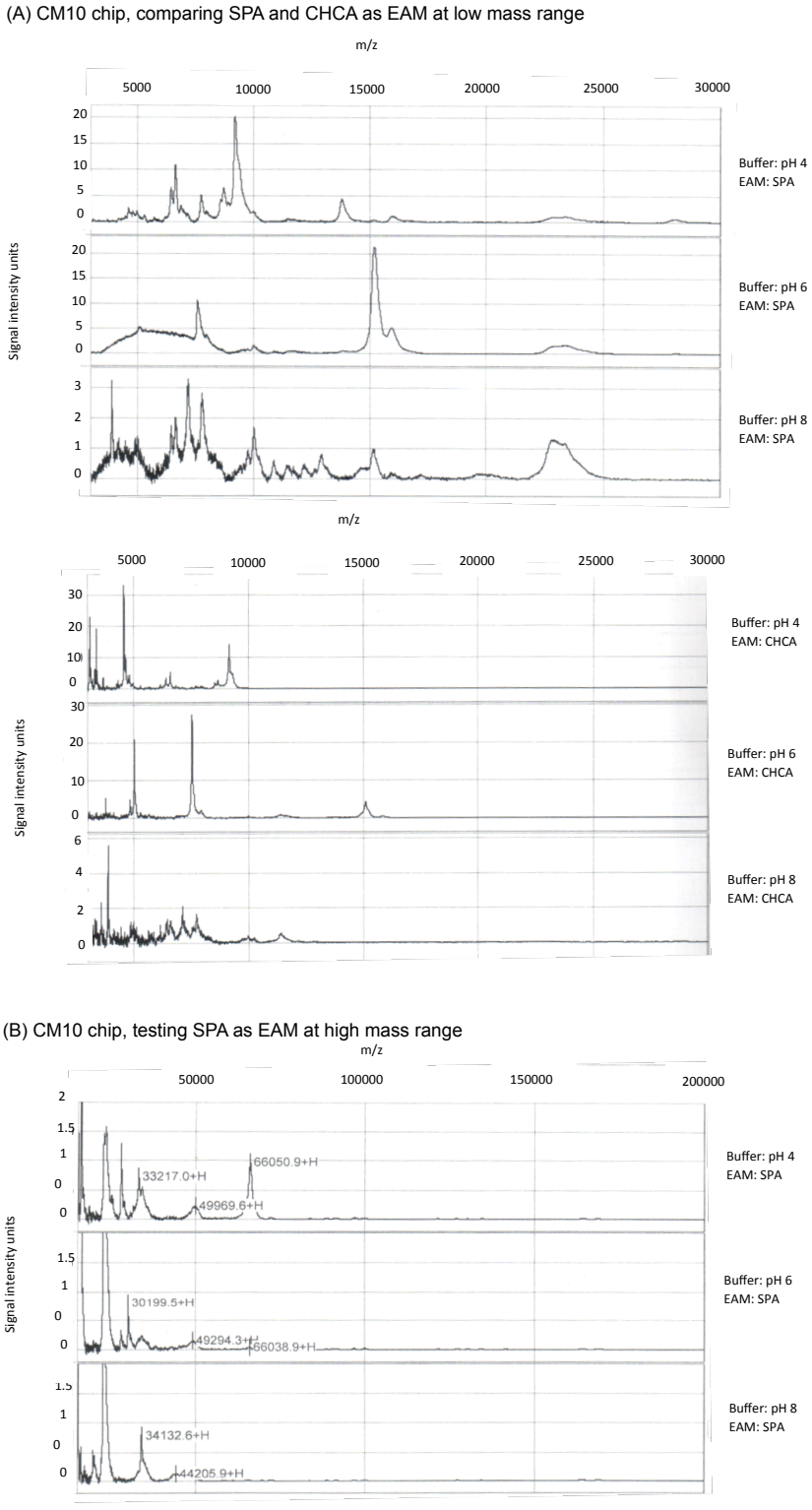
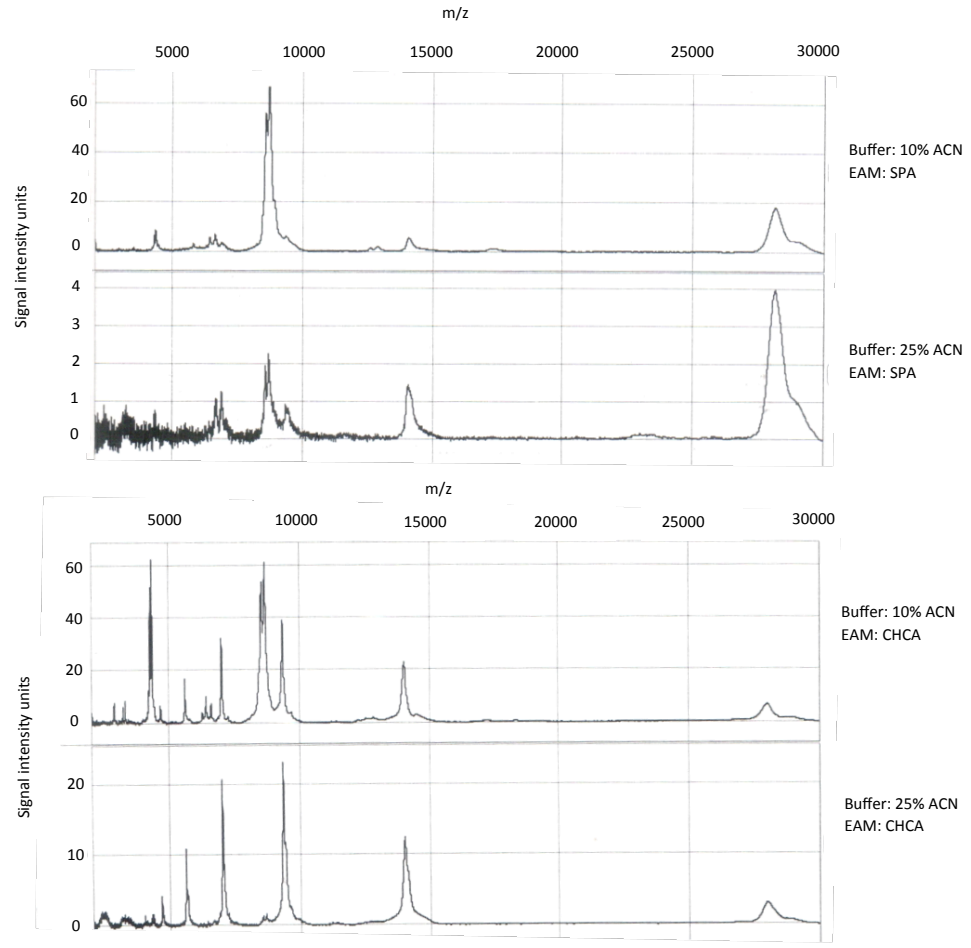


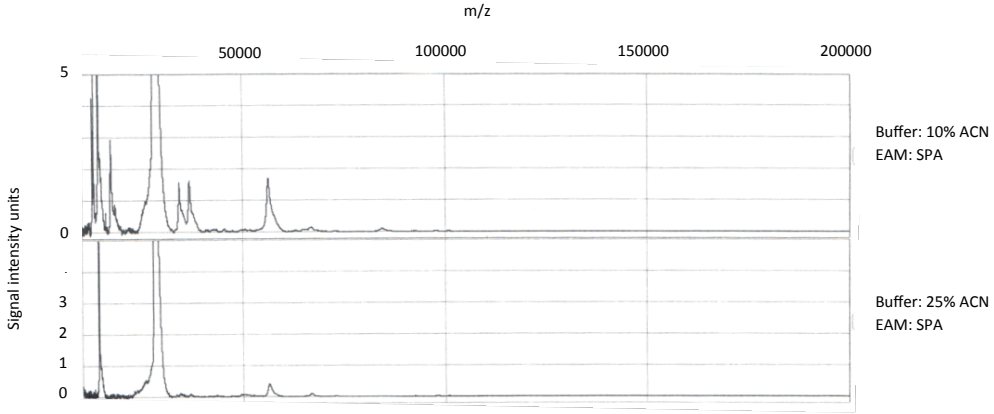
Figure 4.12: Representative results from preliminary SELDI-TOF MS H50 condition optimisation experiments

Legend: The same sample has been applied to multiple spots on CM10 ships. Different combinations of ACN concentration and EAM have been used, permitting comparison of spectral peak resolution and signal intensity.

(A) H50 chip, comparing SPA and CHCA as EAM at low mass range



(B) H50 chip, testing SPA as EAM at high mass range



5. Neutrophil gelatinase-associated lipocalin (NGAL) as a potential blood plasma biomarker for CC

5.1 Background

5.1.1 Current biomarkers

As discussed in Chapter 4, biomarkers circulating in the blood of patients with CC offer an attractive theoretical target for new diagnostic and prognostic tests. Current clinical biomarkers of CC, such as CA19-9, have suboptimal sensitivity and specificity, particularly in early stage disease, or on the background of PSC and other benign biliary conditions. There is, therefore, a need for new biomarkers for earlier diagnosis of CC and potentially for screening in high-risk groups. Previous work from the cholangiocarcinoma group at Imperial College London, to which I contributed, has identified neutrophil gelatinase-associated lipocalin (NGAL) in bile as a potential biomarker of CC.

5.1.2 Neutrophil gelatinase-associated lipocalin (NGAL)

NGAL is also known variously as lipocalin-2, siderocalin, oncogene 24p3, migration stimulating factor inhibitor, alpha1-microgobulin related protein and uterocalin.[407] NGAL is a member of the lipocalin family of small, secreted glycoproteins that transport small hydrophobic ligands. An extensive body of literature has emerged over the last 10 years, demonstrating the considerable interest in the function of this molecule in health and disease and its potential as a biomarker of disease.

5.1.2.1 Physiological functions of NGAL

NGAL is a 25kDa protein that is expressed by activated neutrophils, and a wide range of other cells, including epithelial cells of the breast, bone marrow, kidney, liver, lungs, small intestine and prostate.[408, 409] Human cell line and *in vivo* mouse studies have elucidated important roles of NGAL in health and disease. The main known function of NGAL is the transport of iron particles into the cell cytoplasm via interaction with specific membrane receptors (24p3R).[410] NGAL has a strong affinity

for bacterial iron binding siderophores, which are critical for the proliferation of many gram-negative bacterial species. NGAL bound siderphores are transported into the host cell cytoplasm, reducing the free iron available to bacteria and thereby exerting a bacteriostatic effect.[411, 412] Some authors have described NGAL as a stress protein or cytokine, as it appears to have pro-inflammatory effects. Increased expression in mouse and human models of lung injury has been demonstrated, as has a strong chemoattractive effect on neutrophils.[408, 413] After intravenous injection of *E. coli* into mice, NGAL levels in the plasma and liver rapidly become elevated.[414] In human cell line and *in vivo* murine experiments, NGAL seems to have strong antioxidant effects, counteracting oxidative stress and thereby increasing cellular resistance to cytotoxic species and apoptotic signals.[415-418]

5.1.3 NGAL in benign disease

NGAL has been shown to modulate several benign diseases. Its expression appears to promote inflammation in atherosclerotic plaques, likely through its association with MMP8 and MMP9, promoting proteolytic activity of MMP.[419] NGAL is known to mediate proliferation of renal tubules in mice and reduced NGAL expression has been implicated in some disorders of renal tubule proliferation.[420] NGAL has also been shown to suppress red blood cell production in models of anaemia.[421]

Elevated NGAL levels have been demonstrated in human psoriatic skin, myocarditis and the plasma of patients with untreated HIV disease.[422-425] Elevation of plasma NGAL has also been demonstrated in patients with hypertension, severe adult respiratory syndrome (SARS), carotid atherosclerosis and acute myocardial infarction.[426-430] Obstetric studies have shown increased plasma NGAL levels in women with gestational diabetes and preeclampsia.[431, 432] NGAL levels rise after administration of some drugs, including alcohol and methamphetamine.[433-435] In a mouse study, deliberate administration of hepatotoxic agents caused an increase of NGAL in the hepatocytes, biliary epithelial cells, plasma and urine. The NGAL rise

was proportional to the degree of liver injury induced.[435] Some utility of plasma or urinary NGAL as a diagnostic or prognostic biomarker has been demonstrated in a variety of inflammatory conditions including peritonitis, sepsis of any cause and Crohn's disease.[436-438]

The role of NGAL in acute kidney injury (AKI), and its potential as a diagnostic and prognostic biomarker, has been most extensively investigated.[439] Animal studies demonstrate NGAL release from injured renal tubules, with subsequent transport into the blood stream and urine.[430, 440] Creatinine is the most commonly used index of renal function, but lacks sensitivity in early AKI.[441] Plasma and urinary levels of NGAL become elevated very early in the course AKI and correlate with degree of renal failure as well as poor 28-day survival – with diagnostic and prognostic performance superior to that of creatinine.[442-444] Plasma and urinary NGAL assays are now available commercially, in kit form, and are being utilised in clinical practice in some renal units.[445]

5.1.4 NGAL in malignant disease

The role of NGAL in carcinogenesis has been explored extensively and a variety of influences on tumour development have been shown. Some effects appear contradictory and many vary between cancer types. A pro-proliferative and anti-apoptotic influence has been illustrated in *in vivo* murine experiments with thyroid cancer cells. Silencing of NGAL resulted in marked reduction in tumour clonogenic growth and tumour size. *In vitro* studies on the same NGAL deplete cells demonstrated marked increase in apoptosis.[446] Similar effects have been seen in models of endometrial cancer.[447] Rapid increased expression of NGAL in breast and lung cancer as a response to various different apoptosis inducing stimuli has been shown, with consequent significant reduction in the rate of apoptosis.[448, 449] In colon cancer cell lines and mouse models, NGAL expression appears to decrease invasiveness and liver metastasisation.[450, 451]

Elevated plasma NGAL has been investigated as a potential biomarker in a number of malignant conditions. In clinical studies of breast[452], ovarian[453, 454] colorectal[455], gastric[456] and renal cell carcinomas[457], plasma NGAL was significantly increased in malignant groups. In breast, colorectal and renal cancer studies, some utility in prognostication was also demonstrated.

5.1.5 Biliary NGAL in hepatopancreatobiliary malignancy

In vivo studies of murine models of HCC show that NGAL becomes markedly upregulated in the liver tissue early on in the process of malignant transformation.[458] Functional studies of HCC cells lines show that expression of NGAL actually decreases proliferation and invasiveness. NGAL expressing HCC cells implanted into mice show markedly reduced tumour growth.[459][55]

Ectopic over-expression of NGAL in pancreatic cancer cell lines resulted in reduction of cell adherence and invasion. No effect on proliferation or sensitivity to chemotherapeutic drugs was seen *in vitro*. However, when implanted, pancreatic cancer cells expressing NGAL produced smaller tumours, with markedly reduced metastatic spread.[448]

NGAL is expressed very weakly, or not at all, in normal pancreatic epithelial cells. It is universally expressed at high levels in pancreatic cancer.[460, 461] A study of NGAL in pancreatic juice from patients with pancreatic cancer or chronic pancreatitis and healthy controls was able to discriminate cancer patients from healthy controls with a receiver operator characteristic area under the curve (ROC AUC) of 0.9, and chronic pancreatitis from healthy controls with a ROC AUC of 0.87, but could not discriminate chronic pancreatitis from pancreatic cancer. Plasma NGAL levels were also elevated in pancreatic cancer and chronic pancreatitis, when compared to healthy controls (mean values of 103.5, 108.9 and 67.4 ng/mL respectively). Pancreatic cancer and chronic pancreatitis patients could be discriminated from healthy controls, but not from each other.[407][54]

5.1.6 Prior NGAL work by my group

Recently published work from the Imperial College cholangiocarcinoma research group, to which I contributed, utilised a label-free proteomic approach on bile samples collected from patients with pancreatobiliary disease and identified a promising new biliary biomarker for malignancy. In a discovery cohort of 22 benign cases and 16 subjects with malignant disease, gel electrophoresis followed by LC-MS/S was used to identify proteins that were differentially expressed in the malignant group. NGAL was found to be elevated, and this was confirmed with further immunoblotting studies. Our findings were then confirmed in a validation cohort of 14 benign and 7 malignant cases. On its own, NGAL concentration had a ROC-AUC of 0.76, sensitivity of 94% and specificity of 55%. PPV was 60% and NPV 92%. These findings were independent of plasma markers of pancreatobiliary disease such as CA19-9. Combining biliary NGAL levels with plasma CA19-9 levels improved the diagnostic performance with sensitivity of 85%, specificity of 82%, PPV of 79% and NPV of 87%.[364]

5.1.7 Rationale for study

The hypothesis free, proteomic study described above identified NGAL as a promising marker of pancreatobiliary malignancy in bile. Although potentially very useful, collecting human bile requires invasive, and expensive, procedures such as ERCP or PTC, which are not without risk to the patient. A blood test would be more acceptable to patients and healthcare funders, especially if it could be used in a screening context. NGAL is a small, well-conserved molecule. Plasma NGAL is now regarded as a robust clinical biomarker of acute kidney injury and some utility as a biomarker has also been demonstrated in a variety of malignant diseases. This offers clear evidence that NGAL can be shed from a primary disease site and then be usefully detected in downstream biofluids, such as blood (plasma).

5.2 Hypothesis and aim of study

5.2.1 Hypothesis

I hypothesised that plasma NGAL would be elevated in patients with CC, compared to healthy controls and to patients with PSC.

5.2.2 Aim

To determine the potential utility of plasma NGAL concentrations in the diagnosis of cholangiocarcinoma.

5.3 Materials and methods

5.3.1 Ethics

The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki. Approval from the local Research Ethics Committee and NHS R&D Department was sought and obtained for this study (Ref 09/H0712/82). All participants provided written, informed consent.

5.3.2 Subjects

5.3.2.1 CC patients

I recruited subjects with CC from the gastroenterology clinics and ERCP lists at Imperial College London Healthcare NHS Trust (ICH) and University College Hospitals NHS Trust (UCH). CC patients were considered for recruitment if their diagnosis was secured on the basis of a) pre- or post-operative histology or b) multidisciplinary team consensus on the basis of ≥2 imaging modalities, clinical course and plasma markers.

5.3.2.2 PSC patients

I recruited PSC patients in a similar fashion, from specialist hepatology clinics at ICH and UCH. The diagnosis of PSC was determined on the basis of clinical course, prior imaging, consistent autoantibody titres and, where available, biopsy results. Patients with advanced cirrhosis (Child-Pugh B or C, Table 4.1) were excluded, as were

patients who had undergone previous liver transplantation. Patients currently being investigated for suspected cholangiocarcinoma were excluded. However, patients with known dominant strictures that had been fully investigated were included. To reach an adequate cohort size, I approached collaborators at the Norwegian PSC Research Center (NoPSC), Rikshospitalet, Oslo, Norway. They were able to provide plasma samples from well-characterised PSC patients, with on going consent in place for use in studies of liver disease.

5.3.2.3 Healthy controls

Healthy volunteers were recruited at ICH and UCH. After completing a short, confidential health questionnaire, individuals with no current inflammatory or malignant disease, or liver disease of any sort, were recruited. Healthy controls had blood samples collected according to the same SOP and disease subjects. As these were healthy controls, undergoing no current investigation in the NHS, standard laboratory blood test results could not be collected.

5.3.3 Sample collection

Where samples were collected prospectively, I used my standard operating procedure (SOP, Appendix 1). For protein studies, 4mL of whole blood was drawn into an EDTA tube, immediately placed onto ice and processed within 1 hour.

I collected demographic data, including age, sex, ethnicity, and short medical history onto my case record form (CRF, Appendix 2). Where available, the most recent routine clinical laboratory indices were also documented, including white cell count, C-reactive protein (CRP), CA19-9, alkaline phosphatase, bilirubin and creatinine. Where available for CC cases, I collated imaging information to determine the anatomical location of the CC.

Plasma samples from the NoPSC archive had been collected according to a strict protocol that did not diverge from my SOP significantly.

5.3.4 ELISA

Plasma NGAL (lipocalin-2) quantification was performed in duplicate on plasma from each subject using a Quantikine ELISA kit (R&D Systems, Minneapolis, USA). This is a quantitative sandwich enzyme immunoassay technique. 96 well plates, pre-coated with monoclonal antibody to NGAL, were used. Samples and standards were pipetted into the wells where any NGAL is bound. After rinsing, an alkaline phosphatase linked monoclonal antibody specific for NGAL was applied to the samples. After a further wash, a substrate solution (nitroblue tetrazolium) was added to each well and a colour change, proportional to the amount of NGAL initially bound, occurred.

5.3.4.1 Preparation of reagents

R&D Systems supplied all reagents, in kit form. I added deionised water to 40mL of wash buffer concentrate to prepare a total volume of 1L of wash buffer. I diluted 40mL of calibration diluent concentrate in 160mL of deionised water to produce 200mL of calibration diluent. I prepared a series of calibration solutions by initially diluting 100μL of a NGAL standard solution (100ng/mL) with 900μl of calibration diluent, followed by sequential dilutions with 500μL of diluent (Figure 5.1). I prepared my active substrate solution 5 min before use, by combining colour reagents A and B. The stop solution and assay diluent that I used were already prepared in the kit.

5.3.4.2 Sample preparation

I processed the ELISA plates in batches of two. As NGAL is found in saliva, I used gloves and mask when handling all samples and reagents. Aliquots of each sample to be assayed were thawed on ice. After placing 20μL of each plasma sample into a fresh Eppendorf tube (Eppendorf Ltd, Stevenage, UK), I added 380μL of calibration diluent to each sample, resulting in a 20-fold dilution. Samples were then kept on ice until applied to the ELISA plates.

5.3.4.3 Assay procedure

i. 100μL of assay diluent was added to each well on the 8 (A-H) by 12 (1-12) 96 well plates. Columns 1 and 12 of each plate were used for NGAL calibration standards. Each plasma sample was run in duplicate. Therefore, a total of seven plates were used

ii. 50μL of each of the calibration standards were added to wells 1A-1G and 12A-12G and 50μL of pure calibration diluent was added to wells 1H and 12H

iii. 50μl of each plasma sample was added to each of a pair of adjacent wells. A proforma was used to document the sample ID and corresponding wells

iv. Once samples had been applied to each well, the plates were covered with an adhesive film and incubated at 4^{o}C for 2h

v. The contents of each well were then disposed of by removing the film and inverting the plate over a waste sink. All wells were washed 4 times by alternately filling with wash buffer and inverting the plate over the sink. After the fourth wash, the plates were blotted dry on clean paper towels

vi. 200μL of NGAL conjugate was then applied to each well, the plates were recovered and incubated for a further 2h at 4^{o}C

vii. The 4x-rinsing step (step v., above) was repeated

viii. 200μL of substrate solution was added to each well. The plates were again covered and incubated for 30min at room temperature in the dark

ix. 50μL of stop solution was added to each well. The wells were gently agitated to ensure full mixing

x. The absorbance of each well was read within 10min. An optical calorimeter plate reader (Fisher Scientific UK Ltd, Loughborough, UK) set at 450nm was used

5.3.4.4 Calculation of NGAL concentrations

The absorbance recorded for each well was corrected by subtracting the averaged values from the zero wells (1H and 12H) on each plate. A standard curve was then produced, by plotting the absorbance value of each standard against its known NGAL concentration and defining a line of best fit (Excel, Microsoft). An example of such a standard curve, and its defining equation can be seen in Figure 5.2. The concentration of NGAL in each sample was determined by reading each absorbance value against this graph and multiplying by a factor of 20, to correct for the original twenty-fold dilution of each sample.

5.3.5 Statistical techniques

5.3.5.1 Descriptive statistics

Median levels and interquartile range for NGAL, other laboratory values and demographic factors were calculated using GraphPad Prism v5 (GraphPad Software Inc, La Jolla, CA, USA).

5.3.5.2 Comparison of NGAL levels

As the results were from non-paired groups, and were not expected to be normally distributed, I used the two-tailed Mann-Whitney U test to compare groups. This is a non-parametric test, and was undertaken using GraphPad Prism v5. CC and healthy control cohorts were compared first. Differences between CC and PSC cohorts, and between PSC and healthy control cohorts, were then sought. Two combined cohorts were then investigated; the CC cohort was compared with a combined benign control group of PSC and healthy subjects and a combined disease cohort of CC and PSC cases was compared to the healthy control cohort. Finally, NGAL levels in the

different anatomical subgroups of CC (IHCC, hilar CC and distal CC) were compared.

5.3.5.3 ROC AUC

I performed a receiver operator characteristic, area under the curve (ROC AUC) analysis using GraphPad Prism v5. The area under a ROC curve quantifies the overall ability of a test to discriminate between those individuals with the disease and those without. The AUC has no units, but is the fraction of the maximum possible AUC. The maximum possible AUC, that of an ideal test, is 1.0 and an entirely non-discriminatory test has a ROC AUC of 0.5. CC and healthy control cohorts were compared first. Differences between CC and PSC cohorts, and between PSC and healthy control cohorts, were then sought. Two combined cohorts were then investigated; the CC cohort was compared with a combined benign control group of PSC and healthy subjects and a combined disease cohort of CC and PSC cases was compared to the healthy control cohort.

5.3.5.4 Correlation with other laboratory indices

I performed Pearson's correlation analysis to assess relationships between the levels of NGAL and other plasma markers, results of which were collected from clinical systems at the time of sample collection. Pearson's generates an R squared (R^2) value, termed the coefficient of determination. It has a value that ranges from zero to one, and is the fraction of the variance in the two variables that is shared. For example, if R^2=0.32, then 32% of the variance in A can be explained by variation in B. Conversely, 32% of the variance in B may be explained by variation in A. The p-value generated is the result of testing the null hypothesis that there is no correlation between A and B in the overall population. It is the chance that random sampling would result in a correlation coefficient as far from zero as is observed. Pearson's correlation test was performed with GraphPad Prism v5.

5.3.5.5 Exploration of potential confounders

Non-disease factors that differed between my CC, PSC and HC cohorts and might represent confounding factors were explored. Such factors included age, sex, ethnicity and site of collection. I conducted Pearson's correlations of age with NGAL in each study cohort to investigate the influence of age on NGAL level. Where appropriate, I conducted M-W U testing or Kruskall-Wallace ANOVA as appropriate to seek any differences between NGAL across sex, ethnicity and site of collection subgroups. Finally, I repeated M-W U to confirm that any differences in NGAL levels between the principal cohorts were reproducible in subgroups that excluded each confounding factor in turn. These factors were further explored though multiple linear regression analysis, detailed in section 5.3.5.7, below.

5.3.5.6 Performance of routine laboratory indices as biomarkers of CC

I performed exploratory ROC-AUC analysis of each of the routine laboratory indices collected, to permit comparison with NGAL as a putative new marker. As no routine laboratory data existed for my healthy controls, I created a hypothetical data set using Excel (MS). For each parameter, random values were created that fitted a binomial distribution, within that parameter's normal range. ROC-AUC analysis was performed for each laboratory parameter, comparing CC with HC, CC with PSC and then PSC with HC.

5.3.5.7 Multiple linear regression

5.3.5.7.1 Combination of NGAL with routine laboratory indices

The hypothecated healthy control indices were also used for this section of analysis. I used multiple linear regression to perform a general discriminant analysis using STATISTICA (StatSoft Ltd., Milton Keyenes, UK). A forward stepwise technique was used with F_1 of 1.1 and F_2 of 1.0, with a maximum number of steps of 15. Included parameters were plasma urea, creatinine, CRP, ALT, ALP, bilirubin, CA19-9, white cell count, albumin and NGAL. The final model reached was then applied to the subjects in each cohort to create a set of values that could be subject to repeat ROC-

AUC analysis (GraphPad Prism). The CC cohort was first compared to the HC cohort. The analysis was then repeated to compare the PSC and HC cohorts.

5.3.5.7.2 Combination of NGAL and CA19-9

Multiple linear regression was repeated with just two indices available for inclusion, NGAL and CA19-9, to assess the utility of these two in combination.

5.3.5.7.3 Exploration of potential confounders

I performed repeated multiple linear regression analyses, each time weighting for potential confounding factors (age, sex, ethnicity, site of collection). This technique compensates for any influence of the potential confounding factor on NGAL level. The β-value and p-value for NGAL at the end of each regression was noted to quantify any modification in the diagnostic utility of NGAL.

5.4 Results

5.4.1 Subjects recruited

A total of 243 plasma samples were included in this study: 97 samples from patients with confirmed CC, 64 from patients with PSC and 82 from healthy controls with no CC or known liver disease.

There were significant differences in age, male-to-female ratio and ethnicity between the three groups (Table 5.1).

5 (5%) of CC subjects had a prior diagnosis of PSC. Anatomical classification of CC was possible in 84 (87%) of CC cases; 9 IHCC, 59 hilar and 16 distal lesions. All the PSC patients had no cirrhosis, or fully compensated Child Pugh A cirrhosis, and none had a history of liver transplantation.

Routine laboratory data were available on 64 (66%) of the CC cohort and 57 (92%) of the PSC cohort. These represented the most complete set of matched results that our collaborating sites could extract from their archives. As discussed in 5.3.2.3 above, no routine laboratory data were available on the healthy control cohort.

5.4.2 Plasma NGAL levels

Median NGAL concentrations (IQ range) in CC, PSC and healthy controls were 91 (65-154), 83 (66-109) and 64 (52-78) ng/mL respectively (Table 5.2). The scatter plot in Figure 5.3 illustrates the distribution of NGAL values in all three cohorts.

5.4.2.1 CC cohort vs HC cohort

NGAL levels were significantly higher in plasma from CC patients compared with healthy controls ($p<0.0001$) (Table 5.3). The area under the ROC curve was 0.71 (95% CI 0.64-0.79 $p< 0.0001$) (Figure 5.4a).

5.4.2.2 CC cohort vs PSC cohort

NGAL levels were not significantly higher in plasma samples from the CC cohort than those from the PSC cohort ($p=0.17$) with a ROC-AUC of 0.57 (95% CI 0.48-0.65, $p=0.167$) (Figure 5.4b).

5.4.2.3 PSC cohort vs healthy control cohort

NGAL levels were significantly higher in the PSC group (median 83.1, IQ range 65.7-108.2), when compared to healthy controls (median 64.4, IQ range 51.4-77.7, $p<0.0001$). The ROC-AUC was 0.72 (95% CI 0.64-0.80, $p<0.0001$) (Figure 5.4c).

5.4.2.4 CC cohort vs combined benign cohort

NGAL levels were significantly higher in the CC group compared to the combined benign cohort ($p<0.0001$). The ROC-AUC was 0.65 (95% CI 0.57-0.72, $p<0.0001$). (Figure 5.4d)

5.4.2.5 Combined CC and PSC cohort vs healthy controls

NGAL levels were significantly higher in the combined disease cohort of CC and PSC (median 85.5, IQ range 65.2-128.1), when compared to healthy controls (median 64.4, IQ range 51.4-77.7, $p<0.0001$). The ROC-AUC was 0.71 (95% CI 0.65-0.78, $p<0.0001$) (Figure 5.4e).

5.4.2.6 NGAL levels by anatomical subtype of CC

Separate Kruskal-Wallis ANOVA testing of NGAL levels between anatomical subtypes of CC (hilar vs intrahepatic, distal or not known) demonstrated no significant differences (p= 0.17, 0.79, 0.14 respectively) (Table 5.4).

5.4.3 Relationship between NGAL and other indices

There were significant differences in ALT, albumin, CRP, CEA and CA19-9 between the CC and PSC cohorts. There were no significant differences in bilirubin, ALP, white cell count or AFP between the CC and PSC cohorts. (Table 5.5).

There was no relationship between NGAL levels and bilirubin (R^2=0.01), ALP (R^2=0.02), urea (R^2=0.07) or creatinine (R^2=0.03). There was little or no correlation between NGAL and CRP (R^2=0.14) and white cell count (R^2=0.09) and moderate correlation with CA19-9 concentrations (R^2=0.38) (Table 5.6).

5.4.4 Exploration of potential confounders

5.4.4.1 Age

There were significant differences in age between the three cohorts (Table 5.1). Age positively correlated with plasma NGAL concentration in CC and HC cohorts. There was no correlation between age and NGAL in the PSC cohort. When all cohorts were combined, the correlation between age and NGAL level persisted (p<0.000001) (Table 5.7).

5.4.4.2 Sex

There were significant differences in the male to female ratio in the three cohorts (Table 5.1). There was no difference in plasma NGAL concentration between male and female subjects in the PSC and HC cohorts. However, NGAL levels were significantly lower in females in the CC group (p=0.03) (Table 5.8A). Despite this, the significant difference between CC and HC groups, and between PSC and HC groups, persisted in both male and female subgroups (Table 5.8B).

5.4.4.3 Ethnicity

There were significant differences in the ethnicity of subjects in the three groups (Table 5.1). Plasma NGAL level did not vary significantly with ethnicity in the CC or PSC cohorts. However, plasma NGAL did vary significantly with ethnicity in the HC cohort (p=0.00002) (Table 5.9A). In the two ethnic subgroups that are big enough to analyse, Caucasian and Asian, the significant differences between CC and HC groups persists. However, any difference between PSC and HC becomes non-significant in the Caucasian subgroup (p=0.08) (Table 5.9B).

5.4.4.4 Site of sample collection

There were no significant differences in plasma NGAL levels in CC and PSC subjects recruited at the three sites. However, there was a significant difference in plasma NGAL concentrations between HC subjects recruited at the ICH and UCH sites (p=0.0004) (Table 5.10A). In the ICH subgroups, the significant difference between CC and HC cohorts persisted. In the UCH subgroup, the difference between PSC and HC cohorts was sustained (Table 5.10B). However, this could not be tested in the ICH subgroup, as the number of PSC patients was too small.

5.4.5 Performance of routine laboratory indices as biomarkers

In discriminating CC from HC subjects, NGAL (ROC-AUC 0.72) was outperformed by bilirubin, ALP, ALT, albumin, CRP and CA19-9. It was underperformed by creatinine. Urea and white cell count were of no utility in discriminating these two groups. (Table 5.11)

In discriminating CC from PSC subjects, ALT, albumin, CRP and CA19-9 were of some utility. Like NGAL, bilirubin, ALP, white cell count proved to be of no utility in discriminating these two groups (Table 5.12).

In discriminating PSC from HC subjects, NGAL (ROC-AUC of 0.73) was outperformed by bilirubin, ALP, ALT and CA19-9. NGAL was underperformed by

albumin and CRP. White cell count was of no utility in discriminating these two groups. (Table 5.13)

5.4.6 Multiple linear regression

5.4.6.1 Exploration of potential confounders

Multiple linear regression was also used to further explore the potential confounding factors considered in 5.4.4. The uncorrected β-value for plasma NGAL alone as a discriminator of CC from HC subjects was 0.359. This was diminished, very slightly, by correction for age. It was unchanged by correction for sex and was enhanced, very slightly, by correction for ethnicity or site. (Table 5.14)

The same analysis was performed to investigate the influence of confounders on the difference between NGAL in PSC and HC cohorts. The uncorrected β-value for plasma NGAL was 0.410. This was not changed by correction for sex, ethnicity or site of collection. It was diminished somewhat after correction for age (β=0.341) (Table 5.15).

5.4.6.2 Combination of NGAL with routine laboratory indices

A general discriminant function analysis was used to explore the utility of combinations of routine laboratory indices and NGAL in discriminating groups. The optimum model to discriminate CC from HC included 6 parameters: albumin, ALT WCC, CRP, NGAL and CA19-9. The relative β-values for each of the 6 parameters in the model are detailed in Table 5.16. Using this model, a ROC-AUC of 0.98 was achieved (95% CI 0.97-1.0, $p<0.0001$) (Figure 5.5).

When just plasma NGAL and CA19-9 were subjected to multiple linear regression, the optimum model to discriminate CC from HC included NGAL only, with no added value for inclusion of plasma CA19-9. The ROC-AUC generated from this model was 0.73, consistent with the ROC AUC from the raw NGAL data.

General discrimination analysis was also applied to the PSC and HC cohorts. The optimum model here was reached after 4 steps and comprised the following 4 parameters: ALT, WCC, CA19-9 and NGAL. β -values and p-values for each of the parameters in this model are detailed in Table 5.17. Using this model, a ROC-AUC of 0.98 was achieved (95% CI 0.97-1.0) (Figure 5.6)

As there was no diagnostic utility in NGAL alone in discriminating CC from PSC, multiple linear regression analyses were not undertaken on this pairing of groups.

5.5 Discussion

This is the first reported study to investigate plasma NGAL levels in patients with CC and PSC. NGAL is expressed at significantly higher concentrations in blood plasma samples from patients with CC, when compared to plasma samples from healthy control subjects. NGAL is not significantly elevated in CC patients, when compared to patients with PSC. There appears to be no difference in the degree of elevation of NGAL between different anatomical subtypes of CC. NGAL is significantly elevated in PSC patients, when compared to healthy controls. Plasma NGAL appears to increase with age, regardless of disease status.

5.5.1 Potential mechanisms of NGAL elevation

As one would expect from its name, the initial discovery of constitutively expressed NGAL was in neutrophils, with marked increase in expression during neutrophil activation.[462, 463] Studies of activated neutrophils demonstrate 200 fold higher levels in the cell culture media than within the cells, suggesting a mechanism by which NGAL is actively secreted by neutrophils.[408] As already described in section 5.1, clinical studies have shown an increase in plasma NGAL during induced gram negative sepsis in mice, and in a host of human studies of acute infection or inflammation. None of the CC, PSC or healthy control subjects had current active sepsis or were receiving antibiotics at the time of sample collection. However, many of the patients in my CC cohort had undergone instrumentation of the biliary tree with

biliary stenting and therefore had, probably invariably, some degree of post-procedural cholangitis. Although much less likely to have been subject to ERCP and placement of biliary prosthesis, some of the PSC cohort had had instrumentation of the biliary tree, and secondary bacterial cholangitis is one of the known sequelae of PSC itself. None of the healthy control cohort had biliary prostheses or cholangitis of any sort. In my study, plasma NGAL is clearly elevated in both CC and PSC patients compared to healthy controls. The question therefore arises of whether this NGAL rise could simply be evidence of systemic neutrophil activation, due to low-grade sepsis from cholangitis, or a non-infective inflammatory response. The correlation analyses with CRP (R^2=0.14, p<0.001) and WCC (R^2=0.09, p<0.002), both of which are surrogate markers of neutrophil activation, are helpful here. The p-values certainly indicate significant correlation, although the R^2 values suggest that this correlation is modest.

Another factor that I anticipated would differ between my cohorts was the degree of biliary obstruction or cholestasis. Both CC and PSC are cholestatic diseases, but I expected the CC cohort to have more pronounced biliary outflow obstruction. In fact, no difference in the standard markers of cholestasis (bilirubin and ALP) was seen between CC and PSC cohorts. It is reasonable to assume that there would be a difference in bilirubin and ALP levels between CC patients and healthy controls, but the standard laboratory data on my healthy control specimens does not exist to confirm this. The fact that plasma NGAL was significantly elevated in both CC and PSC cohorts might fit with cholestasis being a common cause of the NGAL rise in both groups. However, there was no correlation between NGAL level and bilirubin (R^2=0.01, p=0.53) or ALP (R^2=0.02, p=0.15), suggesting that any NGAL rise is independent of cholestasis.

Hyperbilirubinaemia, infection, cachexia and malignancy are all known risk factors for renal impairment. Clearly, all these factors are more common in CC patients than in

my healthy controls. Urinary NGAL (and to a lesser extent, plasma NGAL) have repeatedly been shown to be elevated early in acute kidney injury, and in chronic renal impairment, before a creatinine rise is seen. I was not able to perform formal creatinine clearance or urinary analysis on the subjects in my study, to definitively assess their renal function, and such tests do not form part of the routine clinical assessment of our patients. However, correlation of plasma NGAL and creatinine in my subjects was, if not non-existent, minimal (R^2=0.03, p=0.27). Urea, which becomes elevated before the creatinine in hypovolaemic states but is a less specific marker of AKI than creatinine, was minimally correlated with NGAL (R^2= 0.07, p=0.051). Very early kidney injury, without a creatinine rise, cannot be excluded as a potential confounder. However, it seems unlikely that early AKI would account for the NGAL rise seen in my CC cohort, without any association with the creatinine level being seen.

Finally, NGAL may be higher in the plasma of CC patients because it is being shed from the site of biliary malignancy into the circulation. NGAL is a small, stable molecule and such a phenomenon has been demonstrated in a number of other malignancies.[407] The fact that, in my study, the laboratory marker with which NGAL most strongly correlates is CA19-9 (R^2=0.38, p<0.0001) supports this. Although it has critical deficiencies, CA19-9 is the current best plasma marker for CC and is known to originate from the biliary tree itself. Our previous work in bile from patients with pancreatobiliary malignancy, where plasma and urinary NGAL levels were *not* elevated, also points to local biliary generation of NGAL in CC.[364] Whether NGAL is elevated in CC simply as a 'bystander', due to local biliary inflammation and neutrophil activation, or is constitutively expressed by malignant biliary cells themselves is not deducible from my data. Work on NGAL expression in CC cell lines and excision specimens would be required to elucidate this. Studies in the pathogenesis of pancreatic and hepatocellular cancers, which are both increasingly

shown to share common pathogenic features with cholangiocarcinoma, do suggest a functional role for NGAL in cancer development.[448, 459] A reduction in proliferation, invasion and metastatic activity of pancreatic cancer has been shown in mice, when NGAL is constitutively expressed by the cells.[55] However, given the range of conflicting effects of NGAL demonstrated in different malignancies, it cannot be assumed that tumour effects seen in HCC or pancreatic cancer can be extrapolated to CC. Investigation in CC cell lines, and CC tissue may elucidate this further, and permit investigation of the potential prognostic effects hinted at by mouse work on HCC and pancreatic cancer cells.

5.5.2 NGAL as a discriminatory test

The ROC AUC analysis suggests that NGAL discriminates between my CC and healthy control cohorts to a reasonable extent (ROC AUC=0.71, $p<0.0001$). NGAL performs very similarly in differentiating PSC from healthy control subjects (ROC AUC 0.72, $p<0.0001$). However, this does not extend to discriminating between CC and PSC cohorts (ROC AUC=0.56, $p=0.17$). As discrimination of benign from malignant strictures is the most difficult clinical challenge, this is perhaps a more important test of the potential utility of NGAL as a diagnostic biomarker, where the ROC-AUC of 0.56 suggests a very weak diagnostic test. Combination of NGAL with other markers, such as CA19-9, in a panel of tests might be expected to improve performance, as it did in our group's earlier work on biliary NGAL. In that study, combining plasma CA19-9 and biliary NGAL improved diagnostic accuracy from a sensitivity of 94%, specificity of 55%, positive predictive value of 60% and a negative predictive value of 92% to a sensitivity of 85%, specificity of 82%, positive predictive value of 79%, and a negative predictive value of 87%. Unfortunately, I did not have routine laboratory data available for my healthy control cohort and so I could not explore this directly. Although I was able to look for association between parameters of infection, inflammation and renal failure and NGAL concentration, it was not

possible for me to look at the utility of combinations of NGAL with other markers. It may be reasonable to assume that all the healthy controls had normal blood parameters. Therefore I generated a hypothetical set of normal laboratory values for the healthy control cohort, to permit further exploratory analyses. I was first able to investigate routine laboratory markers as stand alone discriminatory tests, with some, such as ALT and albumin, performing substantially better than NGAL or CA19-9. I was then able to undertake explorative multiple linear regression modelling to explore combinations of these markers with NGAL. In this analysis, NGAL proved slightly superior to CA19-9, with no added benefit in combining the two. A panel of six markers, including NGAL, was able to discriminate CC from healthy control subjects with a ROC-AUC of 0.98 (i.e. sensitivity and specificity approaching 100%). As there was no diagnostic utility in NGAL alone in discriminating CC from PSC (ROC AUC=0.56, p=0.17), multiple linear regression analyses were not undertaken on this pairing of groups, as such analyses would be futile.

I subsequently undertook the same analyses on the PSC and healthy control cohorts. A panel of 4 parameters, including NGAL, was able to discriminate PSC from healthy controls – again with an impressive ROC-AUC of 0.98.

5.5.3 Technical issues encountered

The sample processing and ELISA techniques proved uncomplicated and reliable, with no significant variation between duplicate samples and minimal assay failures requiring repeat. The manufacturers of the ELISA kit recommended the use of heparinised samples over EDTA samples as EDTA has a theoretical chelating effect on NGAL. However, longstanding evidence exists suggesting that the stability of NGAL is reduced after collection using lithium heparin leading to degradation and subsequent reduced detectable NGAL levels.[436] Moreover, the archived PSC specimens were available only as EDTA-collected plasma. I therefore performed a small pilot experiment to confirm reproducibility of NGAL results in a dozen samples

where both EDTA and lithium heparin samples were available (results not shown). This confirmed reproducibility and concordance of NGAL measured in samples of both formats and I therefore elected to use EDTA specimens in my full experiment.

5.5.4. Comparability of groups

Significant differences existed in age, sex and ethnicity characteristics between my three cohorts. When comparing PSC patients (without cirrhosis or previous liver transplant), who are generally in their 3rd or 4th decade, with CC patients generally in their 6th or 7th decade, it is essentially impossible to age match. I demonstrated a correlation between plasma NGAL level and increasing age in both CC and healthy control groups. This was not the case in the PSC cohort, but this is likely to be a function of the lower, and narrower range, of age in the PSC subjects. Multiple linear regression with correction for age diminished the diagnostic power of NGAL negligibly in discriminating CC from healthy controls, and trivially in the discrimination of PSC subjects from healthy controls. Men are at much higher risk of PSC than women and a similar, but lesser, tendency exists in CC. It is therefore hard to collate fully sex-matched cohorts. Male and female subgroup analysis demonstrated the same significant findings found in the mixed sex groups and multiple linear regression, with correction of NGAL for sex, made no difference to NGAL's performance. Subgroup analysis of Caucasian and Asian subjects demonstrated the same significant findings found in the combined groups and correction for ethnicity in the multiple linear regression analyses did not significantly weaken the finding. My CC cohort does broadly reflect the epidemiology of different disease phenotypes seen in the UK population, with a preponderance of hilar, extrahepatic lesions. Intrahepatic CC may be somewhat underrepresented. This is likely because my CC cases were recruited from medical gastroenterology clinics and ERCP lists, where cases tend to be those amenable to endoscopic therapy, and less IHCC is seen. Testing of NGAL levels between anatomical subgroups of CC demonstrated no

difference, suggesting that NGAL elevation is independent of CC anatomical subtype. Collection of samples at three different sites may also have introduced some bias, although strict, consistent sample collection protocols were observed at all sites. Analysing subgroups from ICH and UCH confirmed the strong significant difference between CC and healthy control cohorts in the ICH cohort. However, this was not a significant finding in the UCH subgroups. The PSC vs healthy control finding was also not demonstrated in the UCH subgroup, however this included only 12 PSC subjects and so was a very limited analysis. It was not possible to make similar analyses of the Norwegian subgroup as I had neither healthy control nor CC subjects from that site. Correction of NGAL for site of collection in multiple linear regression modelling did not modify NGAL's diagnostic performance.

5.5.5 Ongoing and proposed work resulting from my study

Further work resulting from my data is on going. I successfully applied for UKCRN adoption of our on going biomarker discovery programme and currently 6 UK HPB centres are prospectively recruiting CC cases, using our protocol, for inclusion in validation cohorts and further discovery studies. This covers tissue, bile, plasma and urine collection.

Simply completing assays for routine clinical markers in the balance of my samples (principally healthy controls) would allow more complete analysis of confounding effects as well as permitting me to examine the diagnostic utility of combinations of more markers. Obtaining samples from healthy control and CC subjects from the NoPSC Norwegian archive and analysing their NGAL and routine biochemical levels would permit me to look more comprehensively at the confounding effect of the different sites of collection.

We intend to examine the expression of NGAL in CC cell lines and human excision specimens. Given the elevation in PSC, investigation of NGAL expression in biliary tissue from PSC patients may also prove informative

5.5.6 The future of NGAL in clinical practice

NGAL has already found a definite role in the diagnosis of acute kidney injury. The use of NGAL as a clinical biomarker in other diseases is subject to a significant amount of on going work around the world. A diagnostic role in pancreatic and hepatocellular carcinomas is emerging, and may also emerge in CC and PSC.

Beyond use as a biomarker, a number of studies have investigated the potential imaging and therapeutic opportunities associated with NGAL's discovery in tumours and other diseases. Antibody to NGAL conjugated with a near infrared dye (IRDye800CW) has been shown to be feasible in imaging *in vitro* and *in vivo* models of pancreatic adenocarcinoma cells.[464] A further murine study of atherosclerotic disease also demonstrated increased uptake of a labelled NGAL antibody in a carotid plaque model.[419] Although these studies have demonstrated internalisation of the NGAL-dye conjugate into the target cells, the avid uptake of NGAL into non-target tissues, specifically the liver and urinary tract, have limited clinical application to date.[407]

The therapeutic opportunities in down- or up-regulating NGAL expression have been speculated upon. There are currently no specific molecules known to directly modulate NGAL expression. However, recombinant human and mouse NGAL is available as are respective, highly specific antibodies. Evidence that NGAL inhibits proliferation and invasion in HCC and pancreatic cancer offers a potential target for recombinant NGAL therapy.[459] The systemic effects of exogenous NGAL administration have yet to be explored, let alone any efficacy in treating disease states. Conversely, the upregulation of NGAL in the inflamed colonic mucosa of ulcerative colitis has led some authors to speculate that NGAL-antibody is worthy of investigation as a therapeutic agent, citing the efficacy of IL-2 in renal cell carcinoma and anti-TNF therapy in Crohn's disease.[465] Such approaches may be limited by the generation of a host immune response to extrinsic antibodies to NGAL, or by

unintended consequences of NGAL suppression in the target or bystander tissues. To date, the conjugation of NGAL antibodies with chemoactive drugs, aiming to deliver high doses to tumour tissue, has not been investigated. Such an approach would increase dose delivery to tumour tissue whilst reducing systemic doses and toxicity. However, a weakness of such an approach may be the extensive expression of NGAL in diverse cells and it's uptake by the liver and urinary tract – as illustrated by studies of NGAL antibody-IR marker conjugate imaging studies.

Clearly NGAL plays a significant role in health and disease. A large body of *in vivo* and *in vitro* studies of global expression after gene knock in, knock out or modification of cytokine milieu, have confirmed that individual genes and proteins rarely function in isolation in either health or disease.[407] Genes and their proteins function in groups, and studies of NGAL within its wider group of associated genes is likely to prove more fruitful than study in isolation. With burgeoning data of utility as a biomarker in benign and malignant disease, the most exciting clinical opportunities may be yet to emerge from studies of the *function* of NGAL in disease - if they lead to functional imaging or even therapeutic applications.

5.5.7 Conclusion

This is the first study to examine the utility of plasma NGAL as a biomarker in CC. NGAL is elevated in the plasma of patients with CC, and patients with isolated PSC, compared to healthy controls. Elevation of NGAL appears to be generally independent of renal impairment, cholestasis or systemic inflammatory response, suggesting that NGAL could represent an independent plasma biomarker of CC and PSC. Due to the nature of the diseases under investigation, a number of confounding factors have been incurred whilst recruiting subjects for this study. I could, for instance, have restricted my cohorts to Caucasian subjects only, but this would have substantially reduced my sample sizes and generalisability of my findings. Age, sex and ethnic matching of cases and controls would have been advantageous, but

proved impossible in practice. The impact of such factors was investigated as thoroughly as possible, with a range of statistical tools, without undermining any of the principal positive findings. Although my study involved prospective sample collection, all patients recruited had already been extensively investigated and followed up, ensuring clarity of diagnosis. This may mean that my CC cohort represent quite advanced disease and so cannot be directly applied to, say, patients with early, asymptomatic CC. In a PSC cohort, repeated longitudinal testing of plasma NGAL and correlation with the evolution of superadded CC would be required, if the potential as a screening test for CC in PSC were to be tested. On the basis of my findings of no difference in NGAL levels between PSC and CC subjects, it seems unlikely that plasma NGAL would form a useful test for identifying CC developing on a background of PSC. Assessment of the wider utility of plasma NGAL as a diagnostic test would be best achieved by prospective recruitment of all patients with a mass forming or stricturing biliary lesion at the outset of investigation. Serial measurement of NGAL with concomitant routine laboratory markers could then be related to eventual diagnosis and prognosis. This would be best undertaken as a multicentre study, including non-referral centres to reduce the risk of recruitment bias.

5.6 Chapter 5 Tables and Figures

Table 5.1: Subject characteristics, plasma NGAL study

Legend: *n* – number of subjects in cohort. Age shown is median value with interquartile (IQ) range in parentheses. CC – cholangiocarcinoma cohort, PSC – PSC cohort, HC – healthy control cohort.

Cohort	**CC**	**PSC**	**HC**
n	97	64	82
Age	65.8 (56.8-73.8)	42.2 (31.4-52.7)	48.3 (39.8-62.8)
Sex (M : F)	1.6 : 1	3.6 : 1	1 : 1
Ethnicity *n* (%)			
Caucasian	69 (71%)	58 (90%)	41 (50%)
African	8 (8.3%)	0	1 (1.2%)
Asian	15 (15.5%)	5 (8%)	38 (46.3%)
Not known	5 (5%)	1 (1.6%)	0
PSC *n* (%)	5 (5%)	64 (100%)	0

Table 5.2: NGAL concentrations by cohort

Legend: *n* - cohort size, the median value of NGAL concentration is given with IQ range in parentheses, CC – cholangiocarcinoma cohort, PSC – PSC cohort, HC – healthy control cohort.

Cohort	***n***	**Median NGAL concentration, ng/mL**
CC	97	91 (65-154)
PSC	64	83 (66-109)
HC	82	64 (52-78)

Table 5.3: Comparison of NGAL concentration between CC, HC and PSC groups

Legend: Results of Mann-Whitney (M-W) U-test and receiver operator characteristic area under the curve (ROC-AUC) analyses. CC – cholangiocarcinoma cohort, PSC – PSC cohort, HC – healthy control cohort.

Groups compared	M-W U-test p-value	ROC-AUC	ROC p-value
CC to HC	<0.0001	0.71	<0.0001
CC to PSC	0.17	0.56	0.17
PSC to HC	<0.0001	0.72	<0.0001
CC to HC *plus* PSC	<0.0001	0.65	<0.0001
CC *plus* PSC to HC	<0.0001	0.71	<0.0001

Table 5.4: Breakdown of CC cohort by anatomical subtype

Legend: The number of subjects with each subtype of CC is shown, along with the median plasma NGAL level (IQ range in parentheses) for each subgroup. The p-value shown relates to a Kruskal-Wallis ANOVA test comparing all subtype groups.

	n (%)	Median [NGAL], ng/mL	p-value
Hilar	59 (61%)	85.4 (64.4-140.7)	0.53
IHCC	9 (9%)	128.6 (76.6-164.9)	0.53
Distal CC	16 (16%)	94.6 (43.2-143.2)	0.53
Not determined	13 (13%)	108 (67.4-184.9)	0.53

Table 5.5: Routine laboratory parameters by study group

Legend: Quantitative data are shown as median values and interquartile range in parentheses. P-value from Mann-Whitney U test comparison of CC and PSC cohorts. No laboratory data were available for healthy controls.

Parameter	CC	PSC	p-value
Ur	5.0 (3.7-7.6)	n/a	-
Cr	68.5 (62-82.5)	n/a	-
Bili	21.5 (11.0-77.0)	24 (12.3-41.0)	NS
ALP	287 (152.3-625.5)	216 (122.3-405.8)	NS
ALT	45.5 (24.0-80.0)	92 (51-131)	$3x10^{-4}$
Alb	30.0 (26.8-38.0)	40 (36-43)	$1.3\ x10^{-7}$
CRP	47.5 (17.7-109)	4.25 (1.7-11)	$1.5x10^{-6}$
WCC	6.4 (5.5-9.6)	6.24 (4.7-7.7)	NS
AFP	5.0 (2.5-5.5)	4.0 (3.0-5.0)	NS
Cea	3.0 (2.0-15.3)	1.45 (1.2-7)	$4.0x10^{-3}$
CA125	51.5 (23.0-213)	n/a	-
CA19-9	268 (21.5-1607)	24 (16.75-39)	$1.8x10^{-3}$

Table 5.6: Pearson's correlation analysis of NGAL concentration to routine laboratory parameters

Legend: R^2 – correlation coefficient

Laboratory parameter	R^2	p-value
Bilirubin	0.01	0.53
ALP	0.02	0.15
Creatinine	0.03	0.27
Urea	0.07	0.05
WCC	0.09	<0.002
CRP	0.14	<0.001
Ca19-9	0.38	<0.0001

Table 5.7: Pearson's correlation analysis of age with plasma NGAL concentration

Legend: Values given for each cohort and for a combination of all three cohorts. Age given as median value with IQR in parentheses. NGAL concentration, [NGAL], given in ng/mL as median value with IQR in parentheses.

	n	Age	[NGAL]	R^2	p-value
CC	97	66 (57-74)	91 (65-154)	0.07	0.004
PSC	64	42.24 (31-53)	83 (65-108)	0.00014	0.32
HC	82	48 (40-63)	64 (52-78)	0.31	$0.3x10^{-8}$
Combined	243	55 (41-66.5)	75 (59-109)	0.1	0.000001

Table 5.8: NGAL concentrations in male and female subgroups

Legend: (A): NGAL concentrations in male and female subgroups with p-value from M-W U test between male and female values in each row. (B): Results of M-W U testing between CC, HC and PSC subjects within male and female subgroups. NGAL concentration, [NGAL], given in ng/mL as median value with IQR in parentheses.

(A)

	Male (n)	[NGAL]	Female (n)	[NGAL]	p-value
CC	59	108.4 (70-165)	38	75.8 (55.4-118.6)	0.03
PSC	50	82.4 (66-117)	14	84 (72-107)	0.38
HC	41	71 (55-88)	41	63 (50-75)	0.09

(B)

	Male	Female
CC vs HC p-value	<0.0001	0.01
CC vs PSC p-value	0.04	0.66
PSC vs HC p-value	0.005	0.001

Table 5.9: NGAL plasma concentrations by ethnicity and study cohort

Legend: (A) NGAL plasma concentrations by ethnicity and study cohort, with non-parametric statistical comparison of NGAL concentration between ethnic subgroups. (B) Results of M-W U testing between CC, HC and PSC subjects within ethnic subgroups. NGAL concentration, [NGAL], given in ng/mL as median value with IQR in parentheses. Cauc – Caucasian.

(A)

	Caucasian (n)	[NGAL]	African (n)	[NGAL]	Asian (n)	[NGAL]	Not known (n)	[NGAL]	p-value	Test used
CC	69	93 (66-164)	8	60 (40-87)	16	118 (84-135)	5	91 (59-150)	0.13932	KW ANOVA
PSC	58	83 (66-112)	0	-	5	76 (67-101)	1	115 (n/a)	0.99	M-W U
HC	41	*75* (63-89)	1	84 (n/a)	40	57 (44-65)	0	-	0.00002	MW-U

(B)

	Caucasian subgroup	Asian subgroup
CC vs HC P-value	<0.002	<0.00001
CC vs PSC P-value	0.19	0.07
PSC vs HC P-value	0.08	0.02

Table 5.10: NGAL plasma concentrations by site of collection and study cohort

Legend: (A) NGAL plasma concentrations by site of collection and study cohort, with M-W U test comparison of NGAL concentration between site of collection subgroups. (B) Results of M-W U testing between CC, HC and PSC subjects within site of collection subgroups. NGAL concentration, [NGAL], given in ng/mL as median value with IQR in parentheses. ICH – Imperial College Healthcare, UCH – University College Hospital.

(A)

	ICH (n)	[NGAL]	UCH (n)	[NGAL]	Norway (n)	[NGAL]	M-W U, p-value
CC	49	108 (71-167)	48	87 (60-137)	0	-	0.12
PSC	1	72 (n/a)	12	90 (69-115)	51	83 (65-110)	0.82
HC	56	59 (48-73)	26	75 (70-93)	0	-	0.0004

(B)

	ICH subgroup	UCH subgroup
CC vs HC P-value	<0.00001	0.38
CC vs PSC P-value	-	0.86
PSC vs HC P-value	-	0.34

Table 5.11: ROC-AUC analysis of routine laboratory markers, and NGAL, as discriminators of CC and HC subjects

	CC (n)	HC (n)	ROC-AUC	ROC-AUC 95% CI	p-value
Urea	42	82	0.61	0.50-0.72	0.05113
Creatinine	42	82	0.68	0.58-0.78	<0.001
Bilirubin	64	82	0.77	0.69-0.85	<0.0001
ALP	63	82	0.94	0.90-0.98	<0.0001
ALT	60	82	0.85	0.78-0.91	<0.0001
Alb	43	82	0.94	0.90-0.98	<0.0001
CRP	36	82	0.89	0.81-0.97	<0.0001
WCC	52	82	0.55	0.44-0.66	0.32
CA19-9	47	82	0.79	0.69-0.90	<0.0001
NGAL	*65*	*82*	*0.72*	*0.63-0.81*	*<0.0001*

Table 5.12: ROC-AUC analysis of routine laboratory markers, and NGAL, as discriminators of CC and PSC subjects.

Legend: Urea and creatinine could not be included in this analysis as data were not available on the PSC cohort.

	CC (n)	HC (n)	ROC-AUC	ROC-AUC 95% CI	p-value
Urea	-	-	-	-	-
Creatinine	-	-	-	-	-
Bilirubin	64	57	0.51	0.41-0.62	0.79
ALP	63	57	0.6	0.50-0.70	0.06
ALT	60	56	0.72	0.63-0.82	<0.0001
Alb	43	49	0.84	0.77-0.92	<0.0001
CRP	36	48	0.86	0.78-0.94	<0.0001
WCC	52	56	0.6	0.49-0.70	0.09
CA19-9	47	51	0.72	0.60-0.84	<0.001
NGAL	*65*	*57*	*0.55*	*0.45-0.65*	*0.3233*

Table 5.13: ROC-AUC analysis of routine laboratory markers, and NGAL, as discriminators of PSC and HC subjects

Legend: Urea and creatinine could not be included in this analysis as data were not available on the PSC cohort.

	CC (n)	HC (n)	ROC-AUC	ROC-AUC 95% CI	p-value
Urea	-	-	-	-	-
Creatinine	-	-	-	-	-
Bilirubin	57	82	0.82	0.75-0.90	<0.0001
ALP	57	82	0.91	0.85-0.97	<0.0001
ALT	56	82	0.96	0.92-0.99	<0.0001
Alb	49	82	0.69	0.59-0.78	<0.001
CRP	48	82	0.52	0.40-0.64	<0.001
WCC	56	82	0.65	0.55-0.75	0.002
CA19-9	51	82	0.76	0.67-0.84	<0.0001
NGAL	*57*	*82*	*0.73*	*0.65-0.82*	*<0.0001*

Table 5.14: Multiple linear regression analysis, utilising plasma [NGAL] to discriminate CC from HC subjects

Legend: MLR settings: F_1=1.1, F_2=1.0, max steps 15. Uncorrected value is given with subsequent correction for various potential confounding factors.

Corrected for	β−value	p-value
Nothing	0.359	<0.0001
Age	0.339	<0.0001
Sex	0.359	<0.0001
Ethnicity	0.366	<0.0001
Site	0.36	<0.0001

Table 5.15: Multiple linear regression analysis, utilising plasma [NGAL] to discriminate PSC from HC subjects

Legend: MLR settings: F_1=1.1, F_2=1.0, max steps 15. Uncorrected value is given with subsequent correction for various potential confounding factors.

Corrected for	β-value	p-value
Nothing	0.410	<0.0001
Age	0.341	<0.0001
Sex	0.409	<0.0001
Ethnicity	0.400	<0.0001
Site	0.409	<0.0001

Table 5.16: Multiple linear regression analysis, utilising plasma [NGAL] in combination with routine laboratory parameters, to discriminate CC from HC subjects

Legend: MLR settings: F_1=1.1, F_2=1.0, max steps 15. This model was reached after 7 steps. ROC-AUC analysis of this model's performance can be found in Figure 5.5.

Parameter	β-value	F to remove	p-value
Albumin	0.48	26.6	0.000001
ALT	0.201	4.84	0.03
WCC	0.25	12.03	0.0000763
CRP	0.243	5.23	0.024
NGAL	0.21	5.06	0.026
Ca19-9	0.2	4.4227	0.037

Table 5.17: Multiple linear regression analysis, utilising plasma [NGAL] in combination with routine laboratory parameters, to discriminate PSC from HC subjects.

Legend: MLR settings: F_1=1.1, F_2=1.0, max steps 15. This model was reached after 6 steps. ROC-AUC analysis of this model's performance can be found in Figure 5.5.

Parameter	β−value	F to remove	p-value
ALT	0.49	1627	<0.0001
WCC	-0.19	462	<0.0001
CA19-9	0.151	244	<0.0001
NGAL	0.153	213	<0.0001

Figure 5.1: Workflow for dilution of calibration standards

Legend: Sequential dilution of calibration standard resulted in 7 solutions, the ELISA results from which form 7 of 8 points on the subsequent calibration curve. The eighth calibration solution was plain calibration diluent (i.e. zero NGAL).

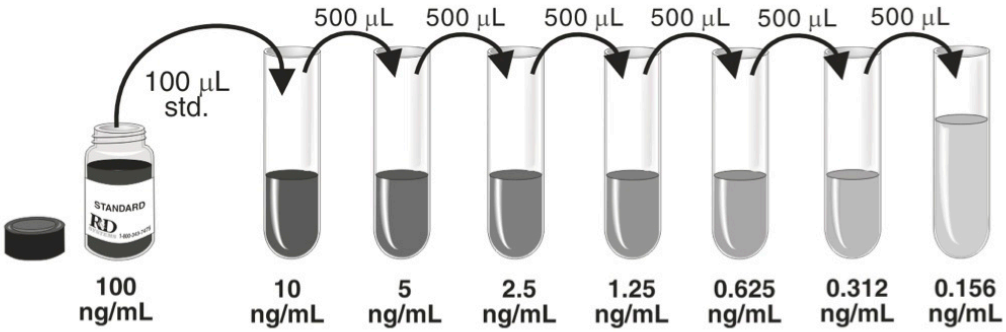


Figure 5.2: Example calibration curve (from plate 1 analysis)

Legend: Diamond points represent values derived using known concentration calibrant solutions. The line of best fit in shown, and is derived from the calculated polynomial formula shown. Unknown NGAL concentrations from study samples were subsequently calculated from the known absorbance value.

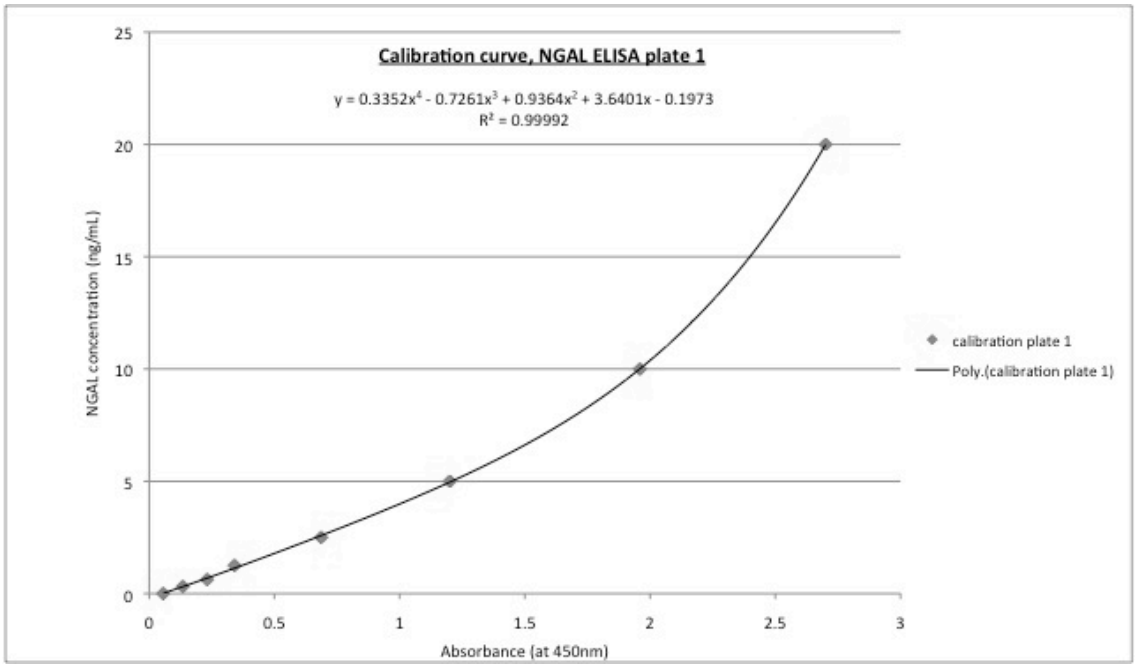


Figure 5.3: Scatter plot of plasma NGAL concentrations by group

Legend: ○ - CC subject, □ - healthy control subject, △ - PSC subject. Red line shows median value.

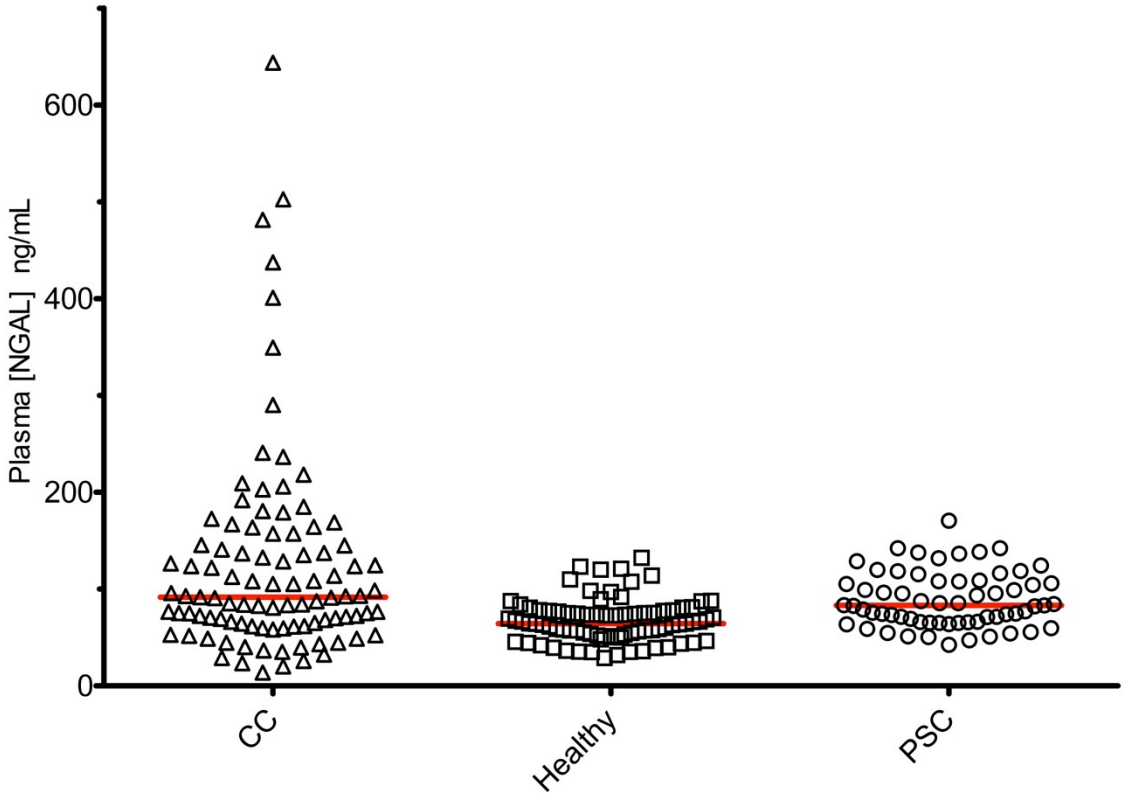
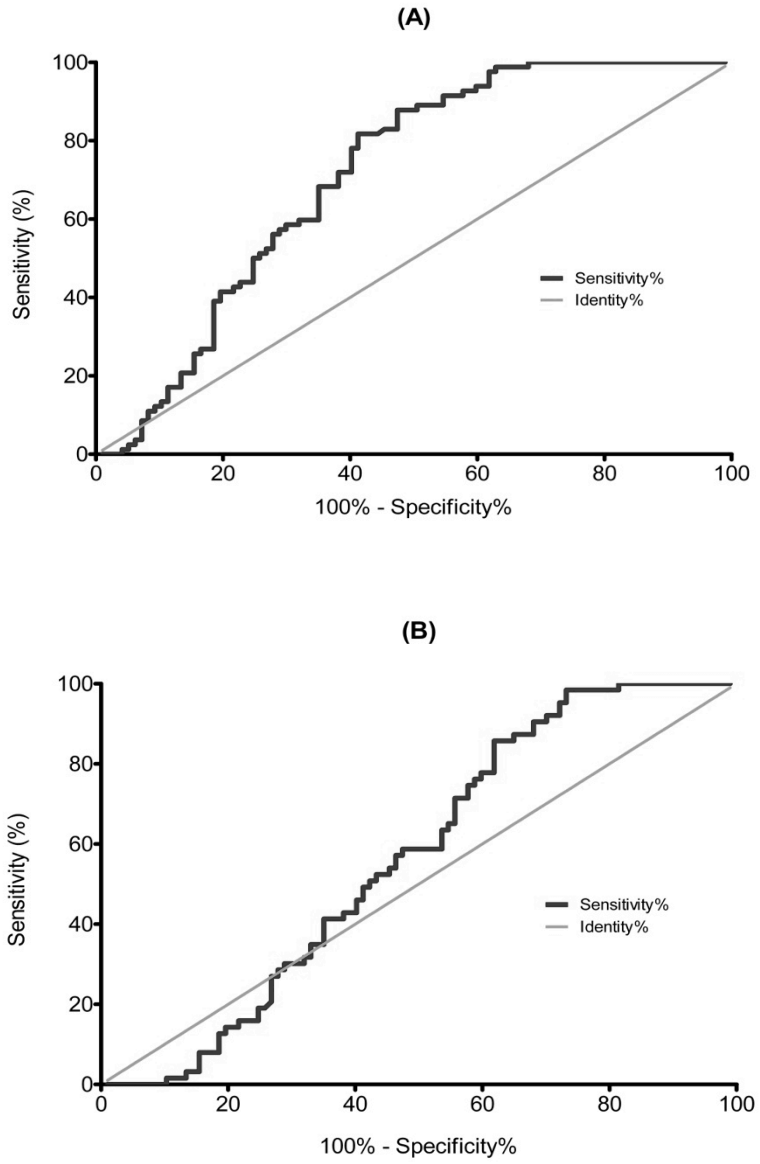
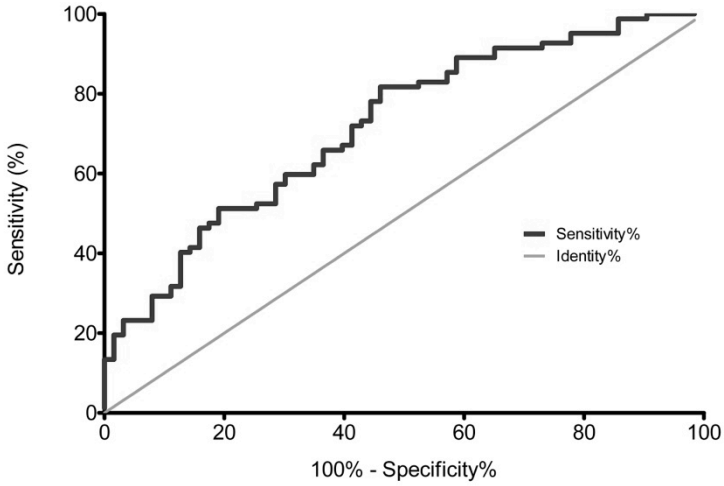


Figure 5.4: ROC AUC curves for inter-group discrimination

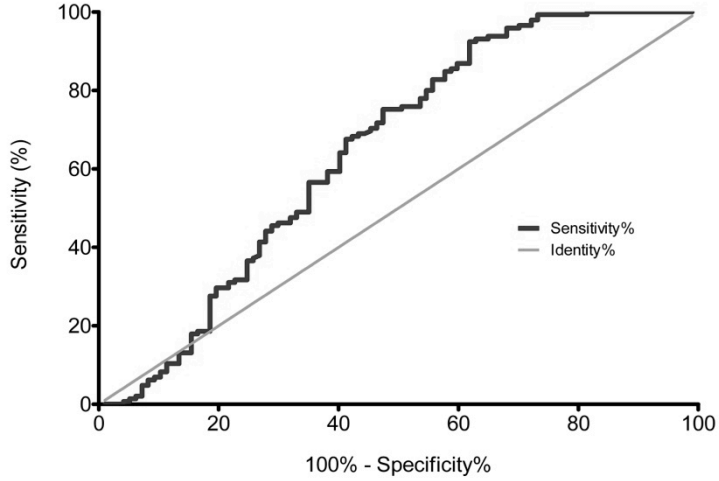
Legend: (A) CC vs healthy controls, (B) CC vs PSC, (C) PSC vs healthy controls, (D) CC vs combined benign groups, (E) Combined disease groups vs healthy controls.



(C)


Sensitivity (%)
100
80
60
40
20
0
0
20
40
60
80
100
100% - Specificity%
Sensitivity%
Identity%

(D)


Sensitivity (%)
100
80
60
40
20
0
0
20
40
60
80
100
100% - Specificity%
Sensitivity%
Identity%

(E)

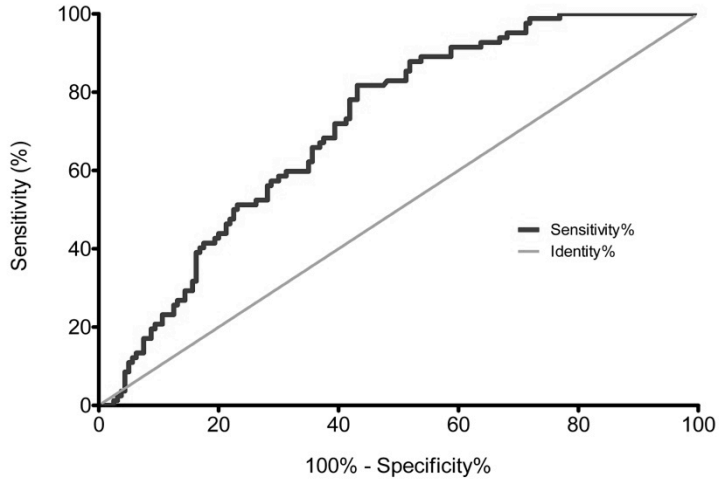
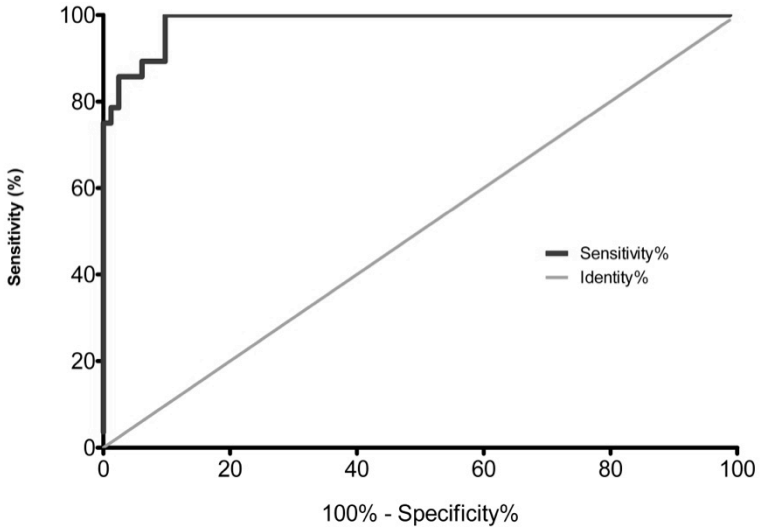

Sensitivity (%)
100
80
60
40
20
0
0
20
40
60
80
100
100% - Specificity%
Sensitivity%
Identity%

Figure 5.5: ROC curve, using values calculated from multiple linear regression modelling, comparing CC and HC groups.

Legend: The ROC-AUC is 0.98. This achieved by combining results from 6 laboratory parameters, including NGAL, in a multiple linear regression model.



6. Development of RF detector microcoils for MR imaging of the biliary tree

6.1 Background

6.1.1 ^{1}H MR physics

Hydrogen nuclei (^{1}H) are found very abundantly in water and, therefore, in the human body. As unpaired protons, hydrogen nuclei have two possible spin states, termed up and down. When placed in the magnetic field of a MRI scanner, the protons all align with, or against, that field. When exposed to radio-frequency (RF) energy, emitted by a transmitter coil in the scanner, the spin state of each proton changes. When the transmitter coil is switched off, the protons return to their original state, emitting RF energy as they do so. A RF detector coil can detect this emitted energy. The rate at which protons change spin state, and therefore emit this RF energy, varies with their chemical environment. This property permits the resolution of different tissues in human imaging studies.[466] The RF signal emitted by protons as they relax degrades exponentially with distance from the receiver coil. To maximise the signal to noise ratio, detector coils are therefore placed as close as possible to the tissue being studied. This permits collection of maximum signal and thereby maximum imaging data. Detector coils tend to be fitted closely around the body part being imaged but some are placed internal to the area of interest, for instance rectal and intravascular coils.[466, 467]

6.1.2 Imaging of the biliary tree

As discussed more extensively in Chapter 1, a range of imaging techniques is used in the investigation of CC. Current staging of CC is relatively poor, partly due to imperfect imaging resolution of the biliary tree and adjacent structures. Endoscopic ultrasound (EUS) has an established role in the assessment of pathology of the distal bile duct, ampulla and pancreas, with particular utility in the local staging of malignant HPB disease. Due to its limited range and resolution, surface ultrasound in unhelpful in the staging of CC. In EUS, the US probe is built into the tip of an endoscope that can be

passed into the upper gastrointestinal tract. The US system can then be closely apposed to the gastric or duodenal wall adjacent to the area of interest, including deep structures such as the pancreas, improving tissue visualisation. However, the resolution of endoscopic ultrasound remains intrinsically limited by available technology. ERCP offers the highest resolution cholangiographic images and the opportunity for therapeutic biliary intervention, such as stent placement, but gives no information external to the ducts and has the attendant risks of pancreatitis and visceral perforation. ERCP is technically demanding, as it requires selective cannulation of the common bile duct. This duct cannot actually be visualised until it is cannulated and radiopaque contrast is instilled. Its axis in a given individual is unknown and varies widely in the population. This leads to a relatively high procedural failure rate (*circa* 25%), and substantial risk of repeated pancreatic duct cannulation and subsequent pancreatitis.[468, 469] An MRI with MRCP is currently considered the first choice diagnostic imaging modality for malignancies of the biliary tree. It is very low risk, non-invasive and a detailed cholangiogram is obtained, along with high-resolution extraductal information such as tumour extent or hepatic metastases. Very high resolution biliary MRI is hampered by the distance of the duct from the surface of the patient's body, and therefore from the surface RF detector coil. An RF detector coil placed within the bile duct would reduce this distance hugely, allowing better signal acquisition and thereby improving image resolution. If a further RF detector coil were integrated into a duodenoscope, offering real time MR imaging of the ampulla and distal bile duct during ERCP, diagnosis of distal disease and deep cannulation rates might also be improved. Integration of a microcoil into the scope tip would also provide additional opportunities for scope tracking within the MR scanner bore.

6.1.3 ESPRC and Wellcome Trust Projects, overview

In the first year of my research fellowship I worked on an ESPRC funded project. This was multifaceted, comprising development of a new catheter mounted RF detector microcoil, a MR compatible side-viewing duodenoscope, a novel system for motion

artefact correction and a new biliary tissue sampling system. I was principally responsible for clinical input into the development of the catheter mounted RF microcoil, its initial testing on phantom models and subsequent testing on *ex vivo* tissue.

In the second and third years of my fellowship, I contributed to a Wellcome Trust funded project. Again, this had multiple subprojects including development of a somewhat larger RF detector microcoil and its integration into an MR compatible duodenoscope, robotised endoscope control systems and automated MR tracking systems. I was principally responsible for clinical input into the development of the detector coil, its testing in phantoms and *ex vivo* tissue, and initial integration of endoscopic and MR imaging in the MR scanner room.

In this chapter I will briefly review previously published work in this field, and then the work that I have personally conducted as a collaborator in these two larger projects.

6.1.4 Radiofrequency (RF) microcoils

6.1.4.1 Previous microcoil catheter technology

Small internal coils placed close to the field of view (FOV) of interest can have a much higher signal-to-noise ratio, demonstrated in vascular studies initially.[467] Catheters based on planar-loop, opposed-solenoid and twisted-pair coils have been developed for vascular imaging and catheter tracking.[470] Previous attempts have been made to develop receiver coils that might be placed into the biliary tree. High-resolution biliary images have been obtained with 'loopless catheter antennae', designed as vascular imaging coils, inserted into percutaneous transhepatic biliary drainage tubes.[471, 472] Microcoils have been produced on gallium arsenide, silicon and glass substrates but most demonstrations have been confined to the laboratory. These detectors have generally been handmade, bulky and rigid. Some examples of previous constructions are illustrated in Figure 1.1.[473] None of these could be adapted for per duodenoscope delivery.

6.1.4.2 Previous endoscope mounted RF detector coils

A commercial MR-imaging endoscope designed for ^{1}H imaging at 1.5 T and containing a single-turn RF detection coil embedded in its tip (the Olympus XGIF-MR30) was demonstrated in the 1990s.[474-478] A similar instrument, designed for research use at 0.5T, has also been investigated. This instrument initially utilised a detached, projecting detector coil[479, 480] and then an inductively coupled saddle coil for intrinsic electrical safety, a prerequisite for clinical application.[481, 482] These systems were developed for the diagnosis and staging of colonic or oesophageal cancer and were therefore based upon forward viewing endoscopes. Furthermore, in several of these instruments the endoscope biopsy channel was occupied with the RF coil cabling, preventing biopsy channel use for other purposes.[479-482] Such approaches would not permit ERCP with simultaneous MR image capture.

6.2 Devices tested

The two devices investigated have been designed and built by Prof. RRA Syms in a collaboration between the Department of Electronic and Electrical Engineering and the Hepatology Section, Department of Medicine, Imperial College London. I contributed to the design specification of these devices, and have been involved in the appraisal of different iterations.

6.2.1 Specification and design of catheter microcoil

Knowledge of previous published attempts to develop microcoils, the physical constraints on endoscopically delivered catheter devices and patient safety requirements informed the specification of the catheter microcoil. Requirements included:

- Introduction and removal through a 3.2mm endoscope channel
- Reproducible and consistent performance
- Economic to manufacture as a single use device
- MR compatible

- Sterilisable before single-use
- Watertight
- Electrically safe: including induced currents in MR fields
- Non-toxic
- Flexible, to make turns within the endoscope and the bile duct
- Maintain at least one internal catheter lumen

The microcoil detector is 60mm in length and 2.7mm in diameter and the whole catheter is 2m long. A prototype version is pictured in Figure 4.2. The detector coil is a two-turn planar coil with stacked windings. Both the detector coil and its output cable are printed lithographically as thin-film devices on a polyimide substrate of less than 100μm thickness (Figure 6.3a). This allows application of the whole detection and transmission system to the outside of a catheter, with subsequent shrink-wrapping (Figure 6.3b). This preserves the internal lumens without excessively increasing the catheter's diameter. Necessary capacitors are integrated into the printed components, maintaining the required smooth, narrow profile of the catheter tip. All materials are non-magnetic; the catheter is watertight and electrically safe. The catheter can be passed through an endoscope biopsy channel and is flexible enough to withstand deflection by the duodenoscope's elevator bridge. It is non-toxic and the cost and difficulty of volume production would be small.

6.2.2 Specification and design of an MR duodenoscope with integrated RF detector microcoil

There were several further considerations that emerged during the design and integration of a microcoil detector into the tip of a duodenoscope:

- Space within the shaft of the endoscope is already almost maximally exploited, with transmission of tip actuation forces, light, air, optics, suction and biopsy channels. Transmission of signal from the RF coil could not displace any of these functions, or diminish the performance of the endoscope

- The scope tip is particularly crowded and there is very little space for integration of further detection systems
- The scope is side viewing and the MRI field of view must encompass the endoscopic field of view
- Integration of a fiducial in the scope tip is necessary to permit MR tracking of the scope tip with a proposed MR tracking system
- Electrical safety is of critical importance

The duodenoscope design is based on a 13 mm diameter non-magnetic side-viewing duodenoscope previously developed by Endoscan (Endoscan Ltd, London, UK). Non-ferrous materials such as polymers, titanium metal, and glass are to be used for the flexible elements, the actuation wires, pivots and fixtures, and the imaging components, respectively. Unlike current clinical endoscopes, illumination will be provided using a remote white-light source and a fluidic optical fibre, and viewing will be carried out using a coherent fibre bundle that routes the image back to an eyepiece or a video camera mounted on the operating handle. Other normal endoscope functions are to be preserved, including water/air, suction and the biopsy channel. In addition, a sub-miniature cable will be integrated into the shaft, with an RF connector on the endoscope's handle.

The endoscope tip is the principal area of modification. A simplified diagram is shown in Figure 6.4, highlighting its most significant mechanical and optical features. The only space available for additional internal components in the tip lies beyond its distal end. Significantly increasing the length of the tip would reduce its mechanical performance, and might also increase the risk of inadvertent visceral perforation during simultaneous scope torqueing and tip deflection. Therefore, the size of the additional space added was kept to an absolute minimum.

The tip design includes two coils. The 1st is the primary detector coil. This is a saddle coil, based on the optimum Ginsberg-Melchner design.[483] It is fabricated as a thin-film

printed circuit board (PCB). Its configuration on the PCB is illustrated in Figure 6.5a. The entire circuit is formed into a sleeve by sandwiching the PCB between two cylinders of plastic, with an opening to allow a clear endoscopic view and passage of catheter tools through the biopsy channel. This first coil is designed to be temporarily applied to the outside of the endoscope, to be removable between procedures and potentially disposable. To achieve intrinsic electrical safety, there is no direct connection between this first coil and the rest of the MR system. The second, much smaller, 'pick-up' coil is located in the small extension to the scope's tip. This is sealed within the scope, but can be accessed by unscrewing the tip cap. Mutual coupling between the two coils occurs through the tip. The signal induced in this second coil by the first is transmitted via the coaxial cable in the scope shaft to the scope handle, and then on to the MR system. The 2nd coil is another thin-film PCB, arranged as a two-turn thin-film saddle coil, and is illustrated in Figure 6.5b. The internal coil is wrapped around a cylindrical plastic shell. The shell also contains a volume of a polymer-based wet burn pad (Spenco Healthcare International, UK), which provides an internal signal source to act as a fiducial marker.[484] Overall placement of the two coils on the tip of the duodenoscope is as shown in Figure 6.5c. The external coil is arranged to overlap both the internal signal source and the internal coil.

Final build of the purpose-designed endoscope is still underway. Scope tip coil experiments described here were performed with the detector coil built into dummy scope tips, or applied temporarily to the outside of a previously constructed MR compatible duodenoscope.

6.3 Hypothesis & Aims

6.3.1 Hypothesis

Intraductal and intraduodenal MR detector coils will increase the resolution of MR imaging of biliary and peri-ampullary anatomical structures

6.3.2 Aims

1) To confirm the utility of a catheter based RF microcoil by:

i) quantifying the resolution of images obtained

ii) defining the sensitive range and signal uniformity

iii) proving that the coil can function in wet conditions

iv) demonstrating high resolution imaging of *ex vivo* pancreatobiliary tissue

v) comparing the images obtained with those from the standard MR coil

vi) quantifying signal-to-noise ratios achieved

2) To confirm the utility of a prototype duodenoscope tip RF microcoil by:

i) defining the sensitive range and signal uniformity

ii) demonstrating high resolution imaging of *ex vivo* duodenal tissue

iii) comparing the images obtained to equivalent EUS images

iv) assessing the impact of simulated unavoidable (respiratory) motion artefact

v) performing simultaneous endoscopic and MR imaging

6.4 Catheter mounted microcoil testing

^{1}H MR imaging was performed using was performed using a 1.5 T HD Signa Excite scanner (GE Healthcare, Little Chalfont, UK). The standard gantry body coil was used for transmission and the microcoil was connected to the auxiliary coil input for reception.

6.4.1 Resolution testing

6.4.1.1 Method

The microcoil was placed at the isocentre in the coronal plane as shown in Figure 6.6A. The microcoil was auto-tuned using a large spherical phantom, using a fast recovery fast spin echo (FRFSE) sequence, and the transmission gain was gradually increased to the maximum allowed by the system.

A glass cuvette containing a nylon nut and cheesehead bolt in copper sulfate solution was used as a resolution test subject. The pitch of the bolt thread was known to be 0.7mm. The microcoil was placed between the cuvette and the spherical MR phantom, as shown in Figure 6.6B. Localising scans were used to centre the image. Images were acquired using a T_2-weighted FRFSE sequence. Imaging was carried out using a relaxation recovery time (TR) of 33 ms, an echo time (TE) of 15ms, flip angle 10^o, FOV of 80x40 and 1.2mm slice thickness.

6.4.1.2 Results

Images were acquired. Figure 6.7 shows a sagittal image, in which the 0.7mm pitch teeth of the bolt are clearly resolved. This demonstrates strongly sub-millimetre resolution imaging. The sensitive range was quantified as 15mm radius around the length of the microcoil.

6.4.2 Field uniformity testing

6.4.2.1 Method

A test rig was constructed and used to hold the microcoil probe immersed in copper sulphate solution, Figure 6.8. A sagittal scout image was obtained before axial images were acquired along the length of the microcoil, using the MR settings detailed in 6.4.1.1.

6.4.2.2 Results

The microcoil functioned well when immersed in liquid. Figure 6.9 shows a sagittal image from a localising scan. The microcoil can be seen running vertically, together with the support rod and co-axial wire. Figure 6.10 shows 28 consecutive axial slices along the microcoil. The central dark area represents the catheter. The signal is highly uniform along the length of the coil. Degradation is apparent in the early images, where the coil is attached to the rig. Two small defects can also be seen along the length of the coil, due to the the coil conductors.

6.4.3 Tissue contrast testing

6.4.3.1 Method

A number of different tissues were used. Butchered porcine material was obtained from a specialist supplier (Medical Meats, Rochdale, UK), so that *en bloc* specimens could be used. The MR settings from 6.4.1.1 were used.

6.4.3.1.1 Specimens examined

i) Bile duct. The cystic duct was incised and the microcoil placed within it, as shown in Figure 6.11. Axial images were obtained along the length of the coil.

ii) Gallbladder fossa. The microcoil was placed on the surface of the liver, as shown in Figure 6.13A&B. Axial images were first obtained with the standard body coil. Imaging was then repeated with the same settings, but using the microcoil for detection.

iii) A pancreas. The microcoil was placed alongside the specimen, with a cuvette phantom as a marker, as shown in Figure 6.16A and B. Axial images were first obtained with the standard body coil. Imaging was then repeated with the same settings, but using the microcoil for detection.

6.4.3.1.2 Signal-to-noise ratio (SNR)

In experiments ii and iii, the signal-to-noise ratio (SNR) was calculated. To calculate SNR, the scanner workstation and software were used (G.E.). Two 40-pixel diameter regions of interest (ROIs) were defined. The first ROI was selected within the most homogeneous area of the tissue scanned. An ROI was then defined outside the tissue, in the image background (i.e. air). The standard deviation of signal strength at this ROI was calculated. Where S is the signal mean from the ROI in the tissue and SD_{air} is the standard deviation in the ROI in air:[485]

$$SNR = S / SD_{air}$$

6.4.3.2 Results

6.4.3.2.1 Biliary tree

Images were obtained. Features of the cystic duct, including its junction with the common bile duct could be discriminated. Extraductal features, including adjacent ducts and vessels could also be seen, Figure 6.12.

6.4.3.2.2 Gall bladder

Images were obtained. Figure 6.14A, B & C show images obtained with the standard body coil. Figure 6.15 shows representative images obtained with the microcoil. Resolution was substantially better in the microcoil images. SNR ratios were eight-fold greater in the microcoil images (260 vs 30).

6.4.3.2.3 Pancreas

Images were successfully obtained. Figure 6.17 shows images obtained with the standard body coil. Figure 6.18 shows representative images obtained with the microcoil. Resolution was substantially better in the microcoil images. SNR ratios were ten-fold greater in the microcoil images (300 vs 30).

6.5 Duodenoscope coil testing

Again, ^{1}H MR imaging was carried out at 1.5 T using a GE Signa Excite clinical scanner at St Mary's Hospital, London. The system body coil was used for excitation, and the test coil was connected to an auxiliary coil input for signal reception. As the purpose-designed duodenoscope with integrated detector coil and transmission was not complete, the detector coil was rigged on a number of different dummy scope tips to permit evaluation. The right hand side of Figure 6.19 shows such an arrangement.

6.5.1 Sensitivity testing

6.5.1.1 Method

Initial evaluations were used to determine imaging sensitivity. To provide an internal fiducial, a 1.4 mm ID catheter was filled with Spenco and passed up the centreline of the coil. To provide an external signal source, the coil was mounted on a cuboid phantom filled with a solution containing 3.37 g/L $NiCl_2.6H_2O$ and 2.4 g/L NaCl, with T1 of 500-800ms and T2 of 100-200ms. Imaging was carried at the magnet isocentre, with

the coil at a variable angle with respect to the magnet bore. A gradient echo sequence was used, with a repetition time TR of 51 ms, an echo time TE of 2.692ms, a flip angle of 30^{o}, a slice thickness of 1.2 mm, a slice separation of 2.2 mm, and a FOV of 120 mm. 4 excitations were used to improve SNR. Figure 6.20 shows the general arrangement; with the coil parallel to the MR scanner bore.

6.5.1.2 Results

A representative axial image acquired in shown in Figure 6.20. The Spenco fiducial is visible as a small bright circle, with a SNR of ≈ 85. It remained visible up to an angle of about 60^{o} with respect to the bore, when the SNR fell to ≈ 40. At 80^{o}, the SNR fell rapidly, to ≈10, although the fiducial was still visible within the image.

6.5.2 Uniformity testing

6.5.2.1 Method

The scope tip coil was insulated and immersed in a tank fabricated from welded polyether ether ketone (PEEK) sheet. Water was used to fill the tank and the coil's internal lumen. The tank was placed on the scanner trolley, with the coil parallel to the magnet bore. Imaging was carried out with 2D gradient echo sequence with TR = 34ms, TE of 13ms, a slice thickness of 2 mm, a slice separation of 3 mm and a FOV of 80mm. Eight excitations were used to improve SNR. The left hand side of Figure 6.21 shows the experimental arrangement. Axial slices were obtained.

6.5.2.2 Results

A representative axial slice image is shown in the right of Figure 6.21. The surrounding water can be seen, together with a small central spot defining the fiducial. Four brighter regions indicate field concentrations near the four coil conductors. This image suggests that the coil response is highly uniform, with little angular variation and a gradual radial decay in sensitivity.

6.5.3 Porcine liver imaging

6.5.3.1 Method

The specimen was a butchered porcine liver, with the biliary ductal system and gall bladder intact, with a section of attached duodenum. The duodenum was opened with a scalpel to allow the tip coil to be located next to the ampulla of Vater. Imaging was carried out with a T2-weighted fast recovery fast spin echo sequence with TR of 33ms, TE of 15ms, a flip angle of 10^0 , a slice thickness of 1.2 mm and a FOV of 80 mm. 4 excitations were used to improve SNR. The left side of Figure 6.22 shows the experimental arrangement.

6.5.3.2 Results

A representative axial image is shown in the right side of Figure 6.22. The central dark region again corresponds to the tube the coil is mounted on, while the surrounding bright tissue represents the duodenum. The biliary duct may be seen on the left-hand side of the duodenum. Excellent image detail is obtained up to a radius of at least 30 mm.

6.5.4 Motion artefact simulation

To define the degree of image degradation caused by expected and unavoidable (respiratory) motion artefact, a simulator was designed and built.

6.5.4.1 Method

Motion artefacts were generated using a hydraulically driven stage consisting of a suspended table capable of carrying an in-vitro liver specimen and RF detector, which was remotely driven by a reciprocating syringe pump as shown in Figure 6.23 and 6.24. The table of the motion generator consisted of a large movable stage supported on multiple balanced flexures. The table was driven using a large syringe arranged as a hydraulic ram. All table components were non-magnetic. The excursion of movement was ± 2cm with a load-bearing capacity of around 1 kg. This table was located on the patient trolley within the scanner bore. The hydraulic drive system was located in the scanner control room, remote from the magnet bore. It consisted of a linear actuator

(Portescap 57DBM10B2U-L, 88N, RS 340-6489 or RS 510-1120), which drove a similar syringe between variable limits. At maximum speed, full travel between limits could be achieved in around 3s, a reasonable simulation of respiratory motion. The actuator and driver were linked together using 10 m of water-filled flexible hydraulic line, which ran between the scanner control room and the magnet room. The hydraulic line was passed through the waveguide port into the magnet room. A porcine duodenal specimen was placed in the sample tray and an duodenoscope tip RF microcoil coil was placed inside the duodenum. The RF coil was arranged parallel to the magnet bore.

Imaging was carried out using a standard 3D GRE biliary sequence. Slices were obtained in blocks of 28, with a FOV of 80mm, a slice thickness of 1.2 mm, a spacing between slices of 1.2mm, an echo time TE of 15ms, a repetition time TR of 33ms, and a flip angle of 10^{o}. Multiple excitations (NEX of 4) were used to improve SNR, and the total acquisition time was around 5min 40s. This time corresponded to a very large number of table motion cycles.

6.5.4.2 Results

Images were obtained with and without activation of the motion generator. The images on the left of Figure 4.25 show coronal slice images obtained with the table stationary. The duodenum is running vertically, and bright tissue with good detail and high contrast may be seen one either side of a central dark band indicating the position of the RF detector. The images on right of Figure 4.25 show exactly corresponding images obtained with the table moving. All anatomical detail has now been completely destroyed, although the approximate location of the RF detector may still be identified. These results confirm the need for a tracking algorithm for *in-vivo* work, which is another part of the wider Wellcome project that was introduced in 6.1.3.

6.5.5 Comparison with EUS

Endoscopic ultrasound (EUS) is the current gold standard for the local staging of distal CC and pancreatic cancer. To demonstrate any advantage of our MR system, we compared the images of a butchered porcine specimen obtained with those that can be produced at EUS.

6.5.5.1 Method

The imaged specimen comprised the distal 25cm of stomach, the duodenum, biliary tree and liver. The cystic duct was transected at the gallbladder neck and then cannulated. The ductal system was distended down to the ampulla with warm agar gel, which then cooled and solidified. Before imaging, the specimen was allowed to come to room temperature and was immersed in tap water in a large water bath.

Clinical governance and infection control rules mean that we cannot perform EUS studies on non-human material using clinical endoscopes. Therefore, a dedicated research EUS system, which will never be used in human subjects, was obtained on loan from the manufacturer's headquarters in Tokyo, Japan (Olympus KeyMed, Southend, UK). An Olympus GF-UM2000 EUS endoscope was used. This is a mechanical radial ultrasonic system that incorporates a HyperBand ultrasound transducer in its tip. This is able to scan at between 5 and 20 MHz with a field of view that can be set between 20 and 120mm. An Olympus EU-ME1 Universal Ultrasound Centre was used as the EUS image processor. Video image processing, air-water supply and suction were not connected to the endoscope, as these were not required for the conduct of this experiment. As in clinical practice, a water filled balloon was placed around the scope tip US transducer.

To confirm correct operation of the system, preliminary imaging of air voids and volumes of water were undertaken. The porcine specimens were then imaged. Using direct vision and palpation of the specimens, the EUS tip was placed at different points in the duodenum (left side of Figure 6.26). In most images, the scope tip balloon was inflated to 1.5cm diameter with tap water. B-mode EUS images were obtained at a

variety of US frequencies and range settings. Images obtained were compared to those from earlier MR microcoil imaging of an identical specimen.

6.5.5.2 Results

Multiple images were obtained within the duodenum with visualisation of the distended bile duct seen. A representative image is presented in the right of Figure 6.26, where EUS settings were 20MHz, 6cm FOV. Comparative imaging from the prior MR microcoil duodenal imaging appeared superior (Figure 6.27).

6.5.6 Integration of coil imaging with duodenoscope

To simulate the planned purpose-built endoscope, the scope tip coil was applied to the end of the previously constructed MR compatible duodenoscope. This does not have the integrated induction coil and transmission line, but otherwise is highly similar to the design outlined in 6.2.2.

The flexible light guide was made over-length, to allow the placement of a tungsten filament light source (SN LB-24, Welch Allyn, Skaneateles Falls, NY, USA) in the scanner control room. An eyepiece and then a MR-compatible video camera (WAT-231S, Watec, Orangeburg, NY, USA) were used for viewing the endoscopic image. The umbilical was also made over-length, to allow connection to air and suction supplies on the wall of the magnet room rather than the standard, non-MR compatible endoscopy service stack. Water was delivered from a bottle (MD-431, Olympus KeyMed).

The duodenoscope was evaluated cautiously in the magnet room, and no magnetic forces were observed. I obtained optical images of porcine specimens using the eyepiece and also using the video camera. The latter were successfully transmitted to a PC in the scanner control room and then projected on a screen in the magnet room. Figure 6.28 shows the instrument in use, and an optical image of a porcine ampulla obtained using the video camera.

Susceptibility effects and decoupling were observed by placing the instrument on a cuboid phantom, with the shaft parallel to the magnet axis, and obtaining MR images using the system body coil. No significant image disturbance was observed, suggesting that any perturbations caused by the duodenoscope and coil to the local magnetic field are minimal. MR imaging was then performed using the instrument itself. A 2D Axial T2 sequence was used, with TR of 68ms and TE of 3.44ms, a flip angle of 30^{o}, a 3 mm slice thickness and 3.3 mm slice separation. 6 excitations were used to improve SNR. Figure 6.29 shows the arrangement and an axial image in which the surrounding phantom and the internal fiducial may both clearly be seen.

6.6 Discussion

Two new RF detector microcoil devices have been developed and demonstrated. These devices are designed for ^{1}H MR imaging at 1.5T magnetic field strength. The first is a catheter-mounted system for insertion into the biliary tree via a duodenoscope, with the intention of producing high resolution imaging of the bile duct and any strictures therein. The second device is a RF detector coil system for integration into the tip of an MR compatible duodenoscope. This scope tip coil is intended to enhance the MR imaging of adjacent hepatopancreatobiliary tissues, and afford data that can be integrated into other areas of the Wellcome project, including automated catheter tracking and robotic scope control. The design of an MR compatible duodenoscope with integration of the scope tip detector device has been completed and it is now being manufactured.

6.6.1 Catheter mounted microcoil

6.6.1.1 Catheter microcoil performance

The catheter produced has an appropriate physical profile and handling characteristics for use through a duodenoscope. It is constructed of non-toxic materials and its electrical design is intrinsically safe, with no direct connection to any power source and built in decoupling to prevent large induced currents when used within the MR scanner.

We have demonstrated that this device can reliably produce images of a resolution of at less than 1mm, and field uniformity is excellent. The device is water resistant. The catheter microcoil can produce high quality images of *ex vivo* tissue, including biliary and pancreatic tissue. We have demonstrated interpretable anatomical detail with sub-millimetre resolution. The images and signal-to-noise ratios demonstrated are superior to those obtained using a standard body coil. Imaging has been successful, even with the catheter microcoil passed through a duodenoscope biopsy channel within the scanner.

A catheter-mounted microcoil has the potential to enhance clinical MR imaging of the bile duct, as well as a number of exciting research applications. A major advantage of the proposed technique is that it could provide cross sectional imaging of the diseased bile duct and adjacent tissue. This is not possible with techniques such as ERCP or choledochoscopy. The precise extent of invasion has shown to be the major predictor of survival after surgery for CC, and is poorly characterised by current imaging methods. In a clinical context, this new microcoil technology should provide very high-resolution images of any hilar or distal biliary lesion, and delineation of invasion into adjacent tissues.

6.6.1.2 Experimental techniques, catheter microcoil

The imaging protocols used to date have required long (>5 minutes) acquisition times, and this would not be acceptable in a clinical study, unless a mechanism for motion artefact correction is determined. In early experiments, prototype coils would fail during some scans. Measurements taken during a test sequence demonstrated large currents induced in the microcoil during the scanner RF transmission pulses, beyond the tolerance of the coils. Further progress was hampered by the high microcoil attrition rate caused. A decoupling system was therefore integrated into the microcoil design. This breaks the microcoil circuit during RF transmission, reactivating it between such pulses. This allows detection of lower energy RF signal during proton relaxation but prevents damaging current induction during scanner RF transmission.

The imaging of *ex vivo* tissue is useful in demonstrating the proof-of-concept of the microcoil and refining it. However, the study of *ex vivo* tissue is an imperfect model of clinical imaging. Devitalised tissue deteriorates rapidly, affecting imaging characteristics and the differentiating properties of adjacent tissues. The lack of blood flow through the *ex vivo* models prevented the use of contrast agents, which are increasingly important to tissue contrast in clinical MR scans. Respiratory motion artefact is common in MR imaging and degrades images substantially, as demonstrated in the motion simulation testing of the scope tip coil. We have not yet directly assessed the effect it may have on the catheter microcoil images, but would anticipate substantial image degradation. A variety of strategies can be used to mitigate respiratory artefact including gated imaged acquisition and, experimentally, field shift coils. These approaches are being investigated by other subgroups in the Wellcome project.

6.6.1.3 Further work warranted, catheter microcoil

My studies of the catheter-mounted microcoil demonstrate the potential of this new approach to imaging of the biliary tree. It is also clear that applications within the rest GI tract, and beyond, should be possible. It is now necessary to demonstrate that comparable, or better, images can be produced *in vivo*. Introduction of the catheter via an endoscope into a sedated human subject in the MRI scanner, with contemporaneous MR imaging would be a substantial first experiment. This would be unlikely to gain necessary approvals. However, after appropriate safety assessments and research ethics approval, we plan a number of incremental studies to further confirm the proof-of-concept and patient safety.

We will integrate a microcoil detector into a standard nasogastric (NG) tube. This could then be passed, without sedation or endoscopy, into a patient's oesophagus. The FOV of the catheter should permit oesophageal MR imaging. This would constitute an initial human study, with only minimal risk to the volunteer.

Introduction of the catheter into the biliary tree via existing clinical access could be achieved via the placement of a balloon anchored nasobiliary drainage catheter. The Cholangioscopy Access Balloon™ (Cook Medical, Bloomington, IL, USA) can be placed at ERCP and retained in the bile duct by an air filled balloon. For patient comfort, the access balloon catheter is then converted from a *per oral* route to a *per nasal* route. Once placed, this arrangement is entirely tolerable for a conscious patient.[486, 487] Once recovered, the patient could undergo MR scanning with the microcoil passed down the nasobiliary catheter that is already *in situ* – without the need for further sedation or an endoscopy within the MRI suite. Alternatively, a microcoil catheter could be passed down the lumen of a percutaneous biliary drainage system, as has been previously reported.[471, 472]

Although simulations are possible, investigation of microcoil tracking and management of motion artefact will only be possible once *in vivo* studies can proceed. Furthermore, the administration of IV contrast to enhance tissue resolution would then also be feasible. The acquisition of high resolution MR spectroscopy data should be possible using the microcoil.[488] This remains to be confirmed in *in vitro* experiments, but may offer an exciting further research application of this microcoil.

Animal studies could be considered for some of these steps. However, if patient safety can be assured then a research ethics committee may consider approving some initial human experiments without an animal phase.

6.6.2 Duodenoscope with tip microcoil

6.6.2.1 Scope tip coil performance

A duodenoscope tip RF detector coil has been produced and tested, but has yet to be fully integrated into an MR compatible duodenoscope. The coil systems have been fully developed and tested on dummy scope tip rigs. The design of the scope tip coil was informed by experience with the prototype catheter based devices, and the literature. Part of the coil system will fit over the end of the modified duodenoscope and will be

disposable. The secondary coil is retained within the scope tip itself. The external coil is non-toxic and has intrinsic electrical safety, with no direct connection between the coil and any external systems. The coil arrangement permits high quality images to be obtained over a wide field of view. Sensitivity, uniformity and resolution testing have confirmed good performance. Simulation of respiratory motion artefact caused complete loss of tissue resolution. Images of porcine specimens obtained with this MR coil compared favourably with those obtained by a commercial clinical EUS system. Interpretable duodenal detail was demonstrated, with a wide field of view (diameter >3cm).

The successful development of this coil, and its pending integration to the duodenoscope, are a critical part of the wider Wellcome funded project. This will permit progress with endoscope tip tracking within the magnet bore and the robotised control systems. If clinical imaging proves possible, the scope tip coil may offer improved resolution upper GI and hepatopancreatobiliary imaging, and has the theoretical potential to supersede EUS as the most sensitive staging test for cancers in these organs.

6.6.2.2 Integration with MR compatible duodenoscope

The duodenoscope, the design of which was considered in 6.2.2, is being constructed from MR-compatible materials, and will combine a coherent fibre bundle for optical imaging, an irrigation channel and a side-opening biopsy channel for the passage of catheter tools with the saddle coil for MR signal reception and an internal fiducial marker. The MR scope tip receiver coil will be magnetically coupled to an internal pickup coil for intrinsic safety, and the system will be tuned, matched and actively decoupled using the minimum of internal components.

A simulation of the designed instrument, with a scope tip coil attached to an existing MR duodenoscope, has been evaluated in a conventional MR imaging suite, and allows optical and MR images to be obtained simultaneously without loss of functionality in either system.

6.6.2.3 Experimental techniques, scope tip coil

As previously stated, the scope tip coil was never fully integrated into a duodenoscope during these experiments. Work has therefore focussed on rigs where the proposed coil system is built into dummy scope tips. This is an entirely reasonable approach, as getting the coil arrangements to work in isolation is likely to prove the most difficult part of the process. An external firm is building the duodenoscope itself and we could not await completion before commencing testing of the coils. Indeed, results from the initial testing of the scope tip coil arrangements have informed the duodenoscope design extensively, ensuring that the delivered device should function as required and expected. In obtaining the images to compare the EUS and MR systems, we used a butchered porcine model. As already considered in 6.5.1.2, *ex vivo* tissue is not necessarily an accurate model for normal tissue imaging. This may have caused relative disadvantage to the EUS system. However, it is likely that any deterioration in the tissue would disadvantage both techniques equally. The motion artefact generator worked as planned. It may not have represented a high-fidelity simulation of respiratory movement artefact, but our experiment is unlikely to have overestimated the impact of this type of movement. It is clear that motion detection and correction is going to prove critical to the successful *in vivo* application of this device.

6.6.2.4 Further work warranted, scope tip coil

The high quality of the images suggests that the duodenoscope tip coil will indeed provide useful *in vivo* performance. Once the integrated scope is in hand it will be further tested, but not on any non-human material as this would preclude its future use in human studies. Once the completed instrument has undergone safety testing on the bench top and in the MR scanner room, and ethical approval has been obtained, *in vivo* human imaging trials are planned.

It has been confirmed that approval from the medicines and healthcare products regulatory authority (MHRA) is not required for the further investigation of either device in humans as they are not commercially exploited and remain the subject of academic

research at present. Ethical approval will be sought for incremental clinical studies, as explained in 6.5.1.3. With the scope tip device, straightforward upper GI endoscopy could be performed first, on a patient or healthy volunteer. This could be undertaken within the MR scanner, but without any attempt to cannulate the bile duct. MR scanning of extra-luminal structures could be undertaken, with and without contrast, with minimal risk beyond that of a standard upper GI endoscopy.

Integration with other elements of the broader project will see the duodenoscope used in simulated procedures within the MR scanner, with MR tracking and robotic control systems in place.

6.6.3 Conclusion

Two new devices have been developed in collaboration between the Departments of Electrical & Electronic Engineering and Medicine, Imperial College London. In a variety of *in vivo* and *ex vivo* experiments, both devices have been shown to perform safely and produce high-resolution images.

The work, in a multi-disciplinary research team of physicians, engineers and MR physicists, has offered me substantial insight into the different techniques and strategies used in developing a new device from scratch. This has contrasted and complimented the more traditional biomedical approaches of my other research studies.

The work presented in this Chapter represents my contribution to two much larger projects, funded by the ESPRC and Wellcome Trust. Both of these projects have fulfilled their stated aims and have led to further successful project grants.

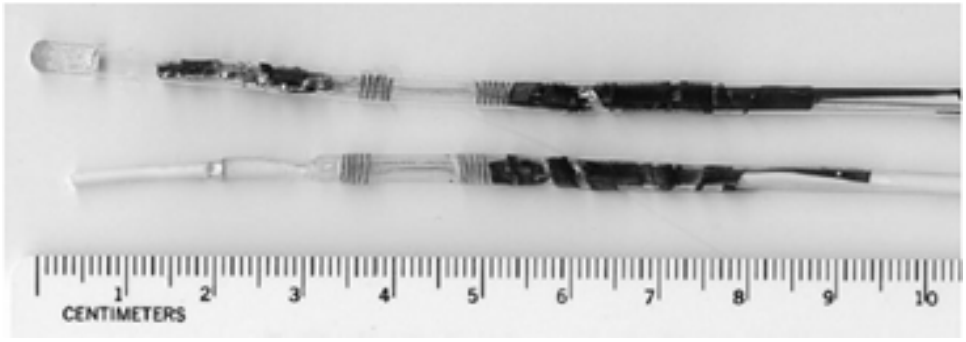
Significant further work, with translation of these devices in to human studies is planned. This work may lead to development of a clinically applicable system.

6.7 Chapter 6 Figures

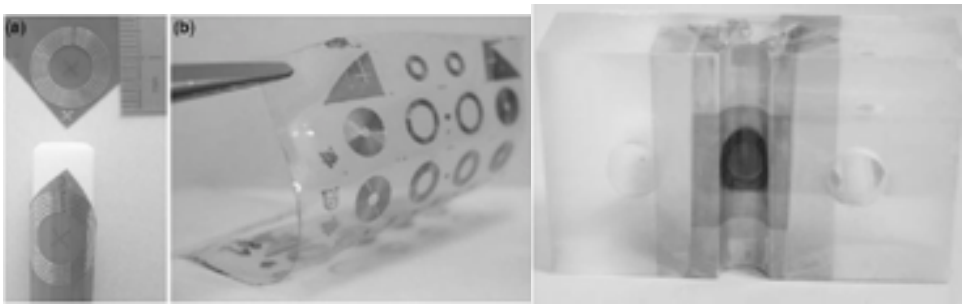
Figure 6.1: State of the art in RF detector microcoils for MR

Legend: Previous attempts to construct MRI microcoil detectors. All have demonstrated some function, but none are suitable for per-endoscope biliary imaging. (A) Hillenbrand 2004[489], (B) Woytasik, 2007[490] (C) Uelzen, 2006[491], (D) Zuehlsdorff, 2004[492]

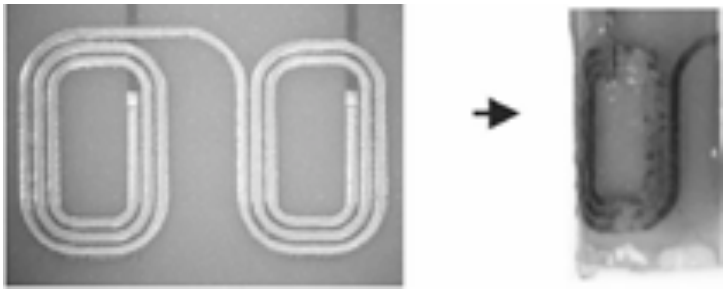
(A)



(B)



(C)



(D)

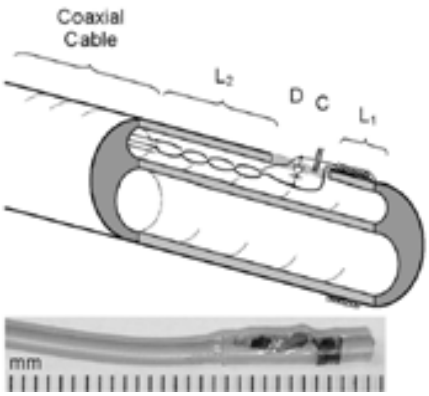


Figure 6.2: Prototype catheter microcoil

Legend: This microcoil is mounted on a flexible biliary catheter

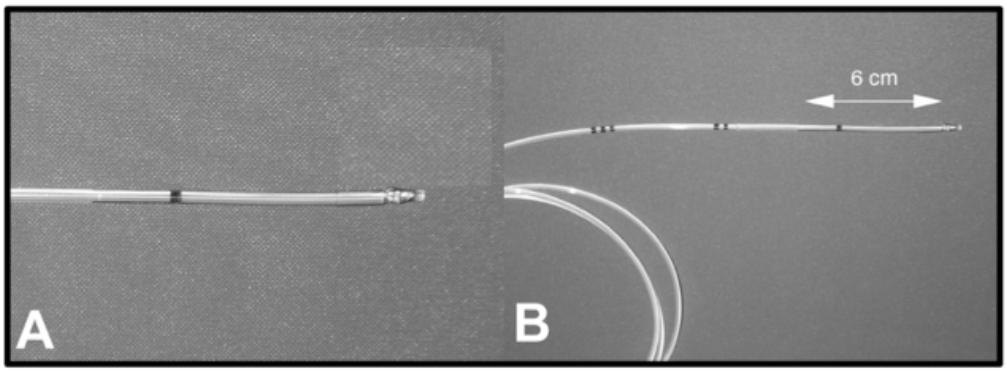


Figure 6.3: Diagram of catheter microcoil design

Legend: Thin film RF detection system: a) plan view and b) integration on catheter. Note preservation of both catheter lumens and only modest increase in diameter. C_M and C_T – integrated capacitors, L_C – coil. Diagram courtesy of Prof RRA Syms.

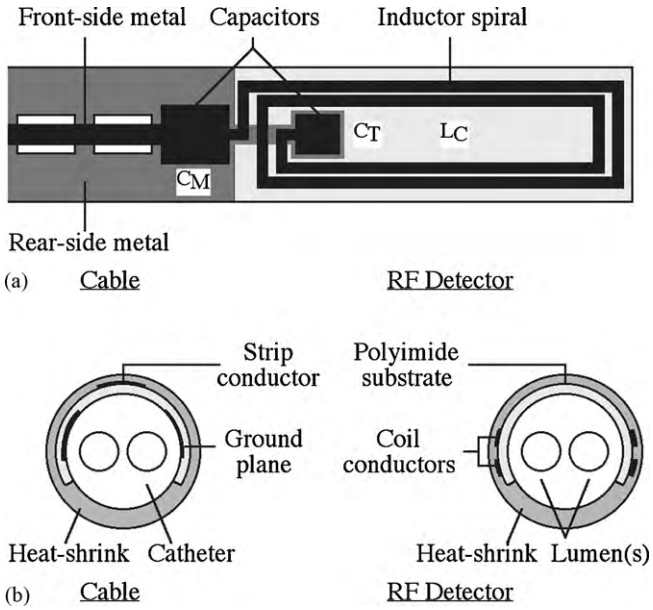


Figure 6.4: Diagram showing modified scope tip design

Legend: The upper image shows the scope tip viewed *en face* to the optics and biopsy channel terminus. The lower is a side view of the bridge deflection system. Diagram courtesy of Prof RRA Syms.

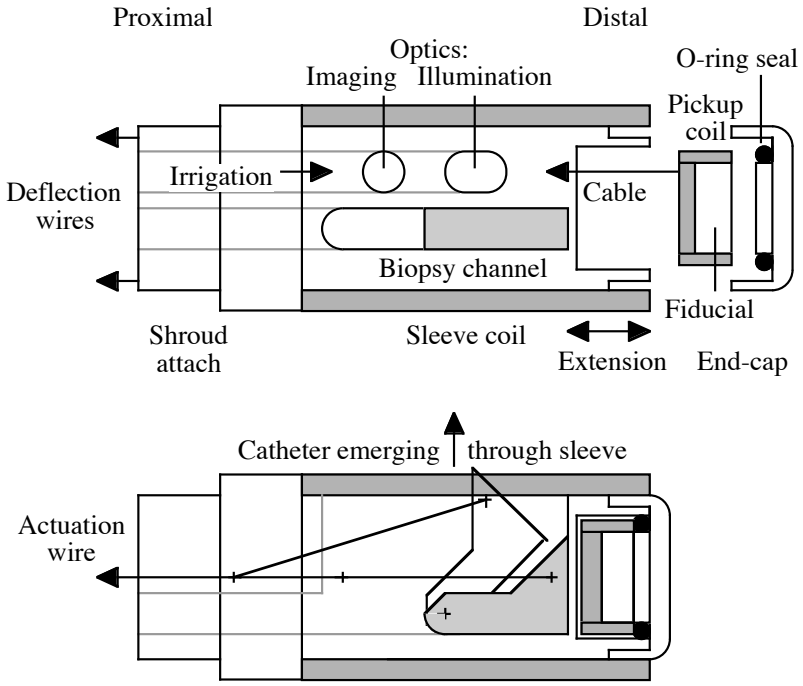


Figure 6.5: PCB layouts of scope tip coils

Legend: Printed circuit board (PCB) layouts of (A) External scope tip saddle coil (B) internal 'pick-up coil'. (C) is a 3D rendering of the coil once applied to the scope tip. C_{1A}, C_{1B} – capacitors of external coil, C2 – capacitor of internal 'pick-up' coil. L_1 – external detector coil, L_2 – internal 'pick-up' coil, PIN – PN diode. Diagram courtesy of Prof RRA Syms.

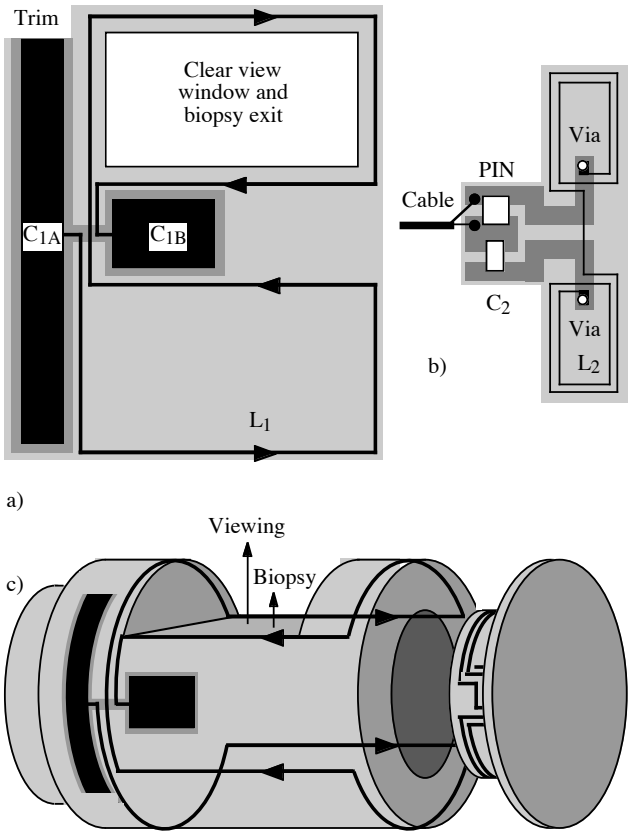


Figure 6.6: Arrangement of catheter microcoil for resolution testing

Legend: (A) The coil is shown orientated in the axis of the bore (indicated by arrow B) of the MR scanner (B) Cross-section of cuvette containing nylon bolt and copper sulphate solution

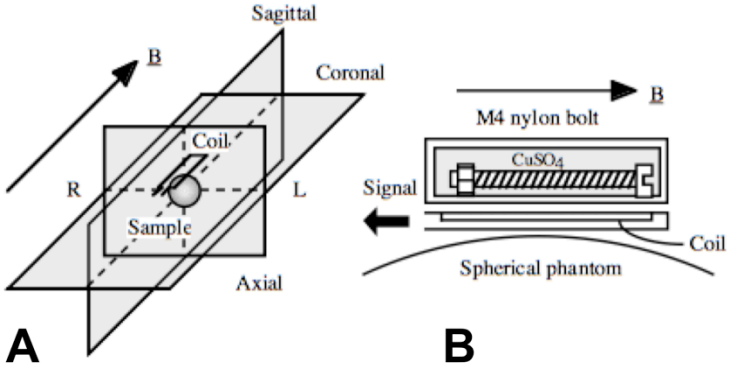


Figure 6.7: Image acquired during resolution testing with catheter microcoil

Legend: Thread pitch of bolt known to be 0.7mm

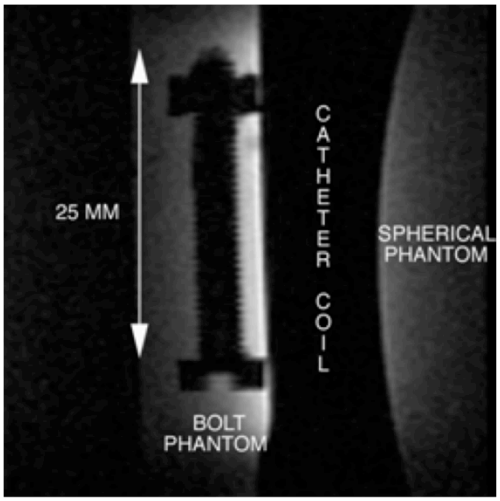


Figure 6.8: Uniformity testing set-up

Legend: Lines through coil represent orientation of axial slices acquired. B is the orientation of the magnet bore.

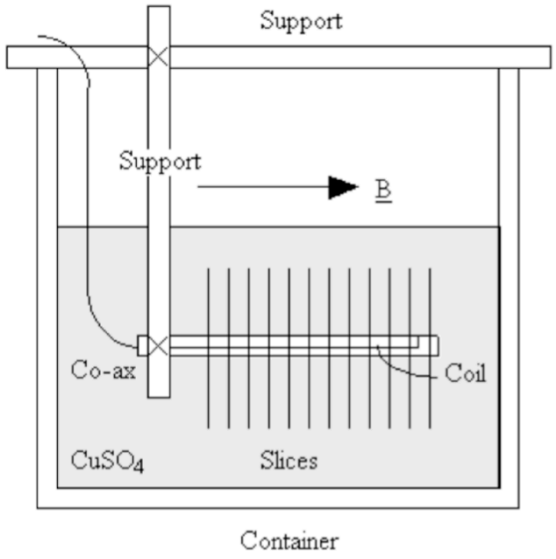


Figure 6.9: Saggital scout image of microcoil

Legend: The microcoil can be seen in saggital view with the coaxial wire extended out to the top left extreme of the field of view.

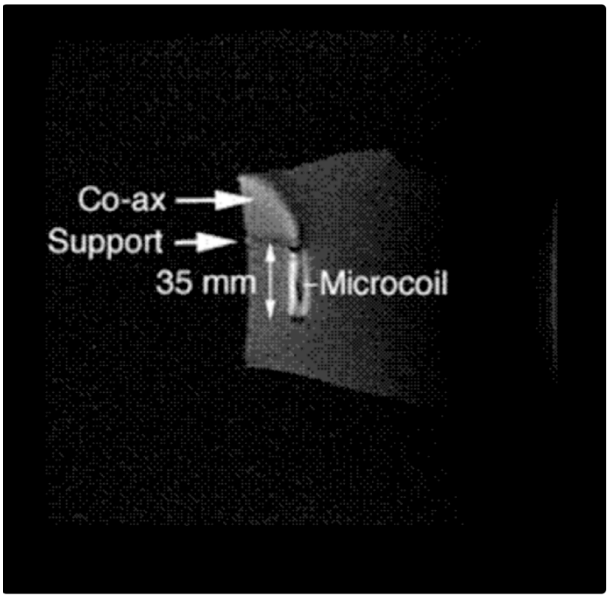


Figure 6.10: Axial images along length of microcoil

Legend: Intensise signal was seen around the whole of the probe for almost its entire length, confirming good signal uniformity

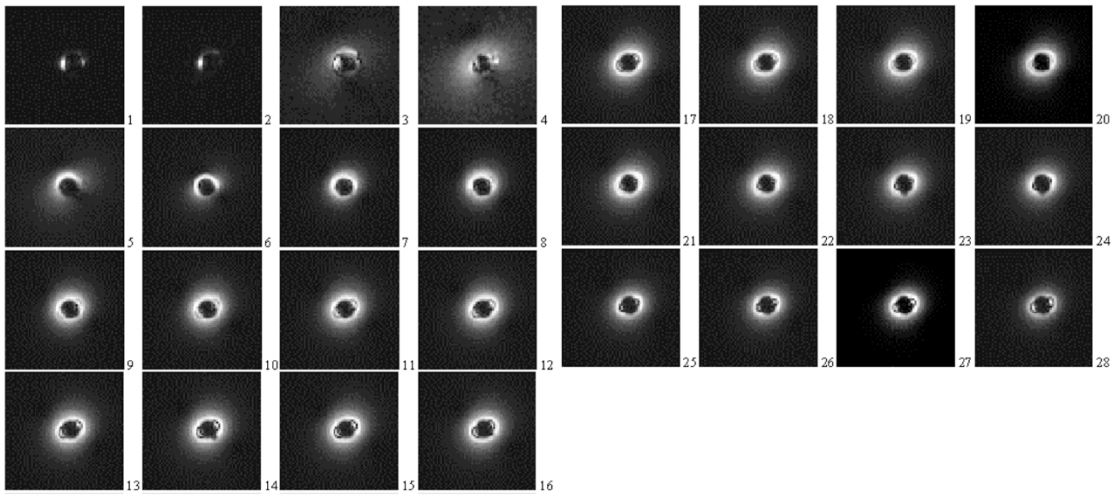


Figure 6.11: Experimental set up, porcine liver imaging

Legend: The cystic duct has been incised at is junction with the common bile duct and the catheter mounted coil inserted, retrogradely, up the cystic duct

Figure 6.12: MR image porcine liver

Legend: A representative axial slice through the catheter's axis

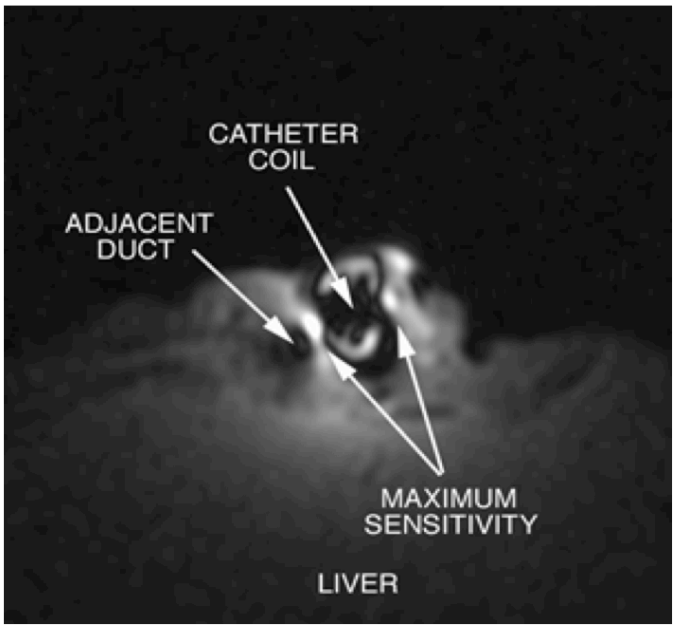


Figure 6.13: Arrangement of microcoil catheter on surface of liver specimen

Legend: The catheter has simply been taped to the specimen, in the approximate axis of the gall bladder.

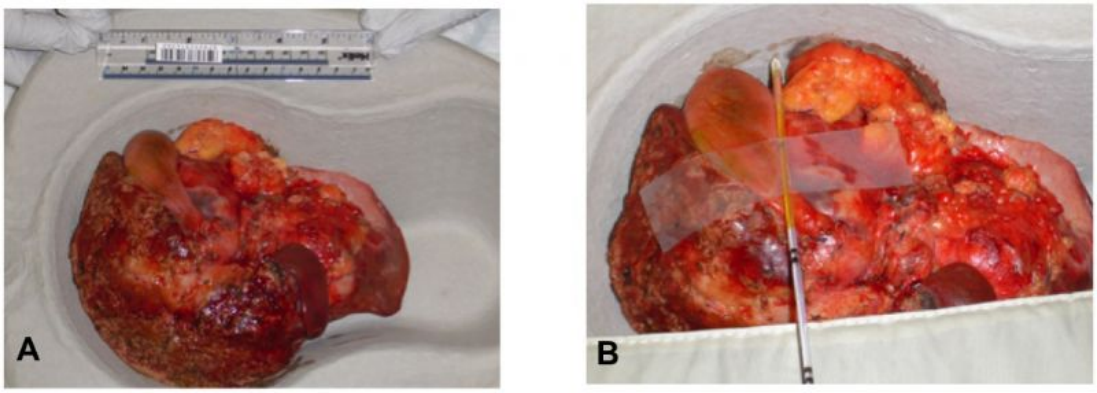


Figure 6.14: MR images of liver specimen, obtained with standard body coil

Legend: (A) the whole specimen, with microcoil seen as small, intense signal next to GB. (B) and (C) progressive magnification of body coil image of GB. Cf. MR microcoil images of same FOV in Figure 6.15.

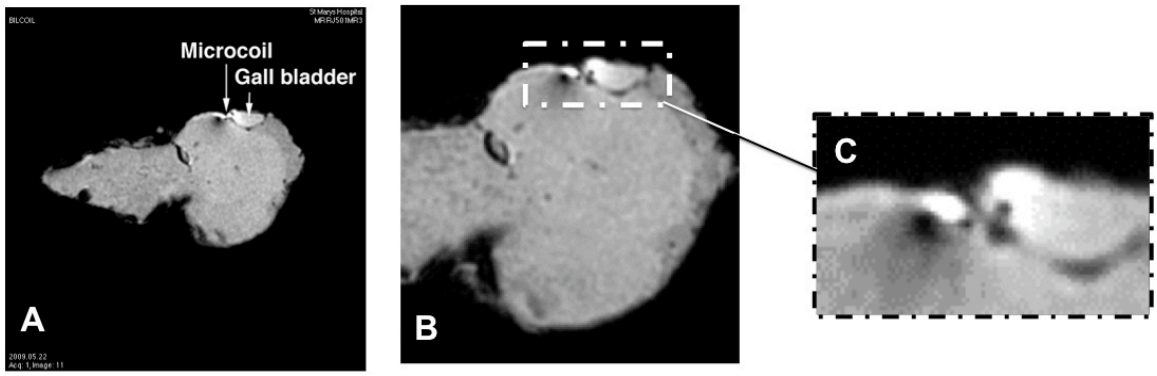


Figure 6.15: MR images of liver specimen, microcoil

Legend: Microcoil acquired axial images showing high resolution of GB fossa structures

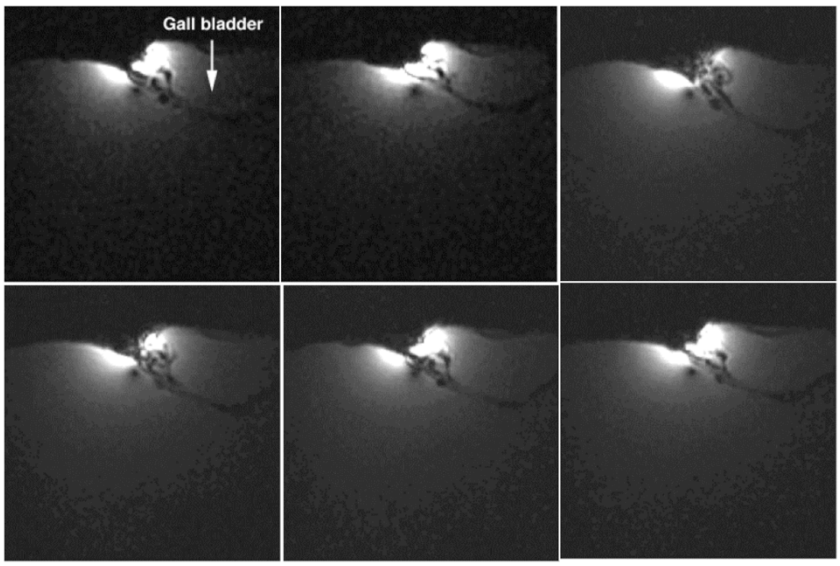


Figure 6.16: Arrangement of microcoil and pancreatic specimen

Legend: The pancreas was arranged with a copper sulphate filled cuvette to act as a marker, as the volume of tissue was a relatively small target in the scanner bore.

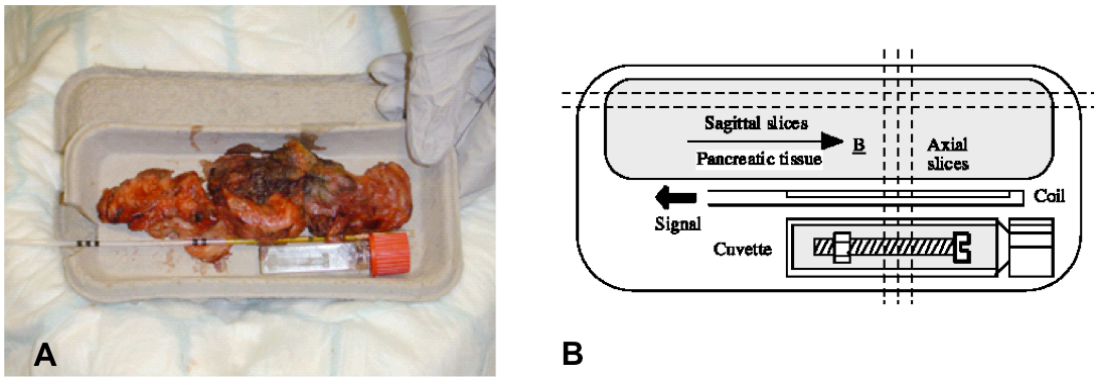


Figure 6.17: MR images of pancreas, obtained with standard body coil

Legend: Axial images through pancreatic specimen and copper sulphate filled cuvette using the body coil.

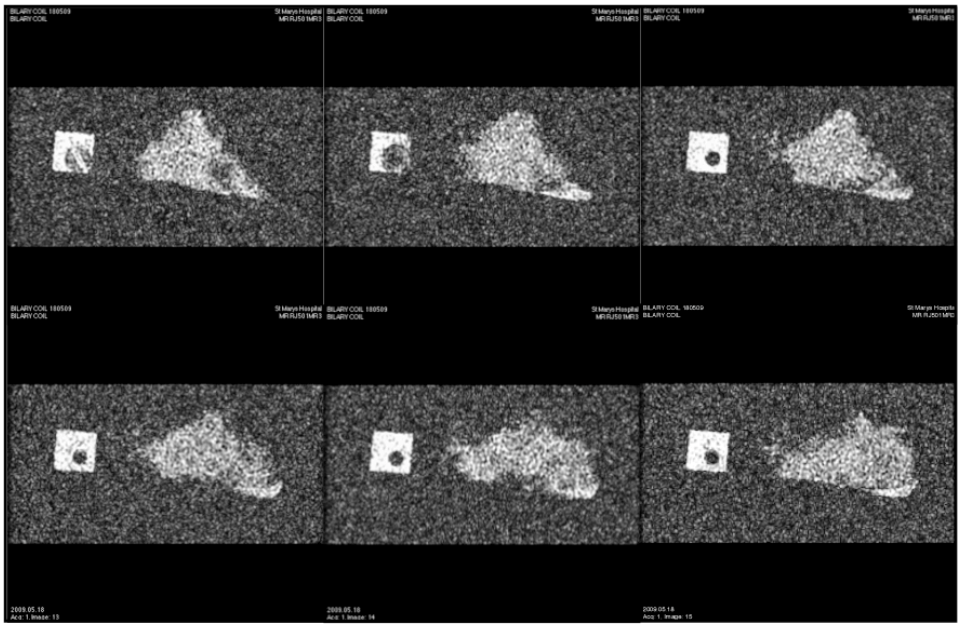


Figure 6.18: MR images of pancreas, obtained with microcoil

Legend: Microcoil images in the same orientation as those in 6.17. Resolution of structures approximate to the coil is substantially better. Note the screw head seen in some images.

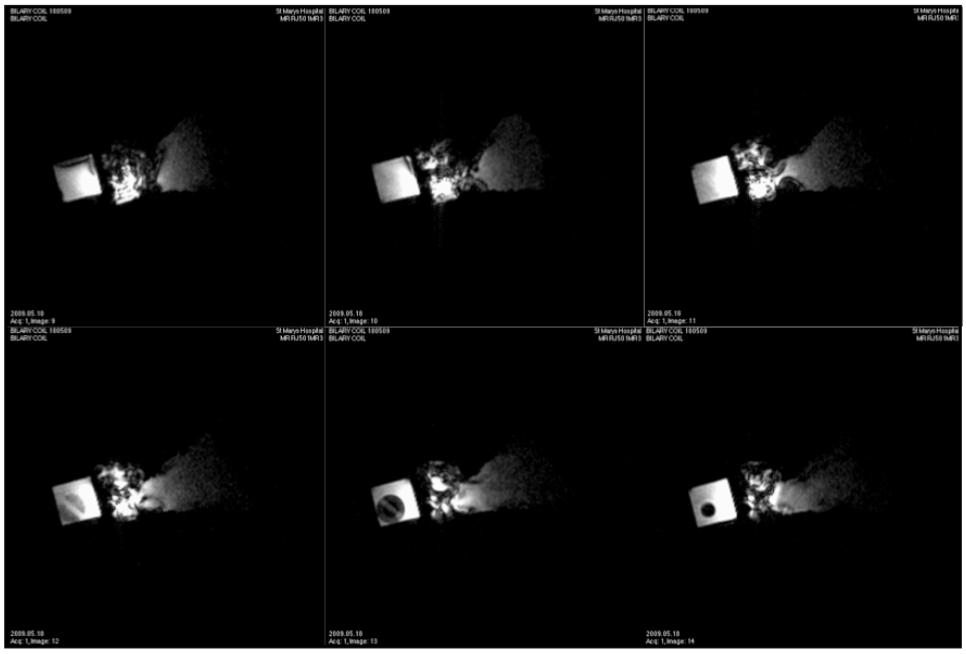


Figure 6.19: Thin film saddle coils before application to magrel

Legend: The PCB arrangement of the cope tip microcoils is shown. The right hand side of the image shows the coils applied to a dummy scope tip.

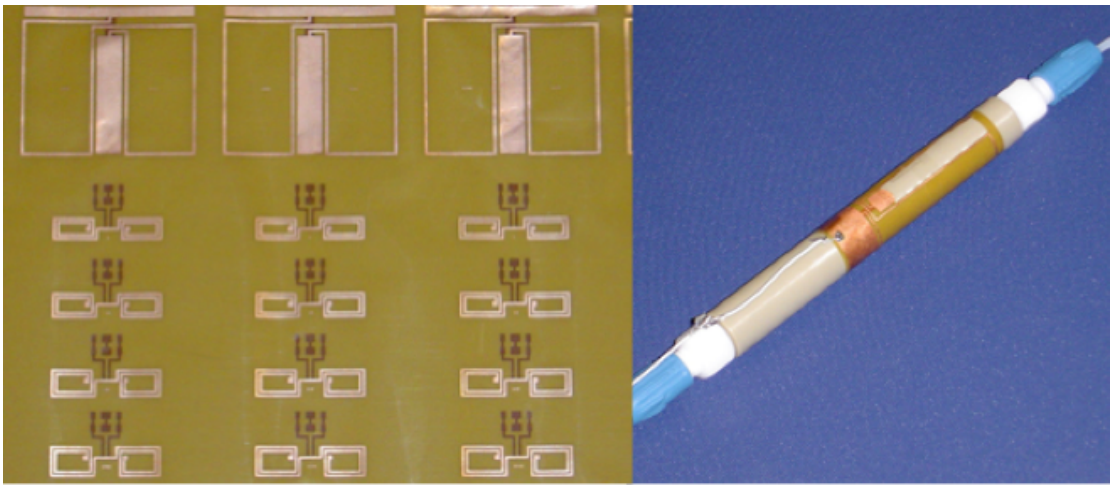


Figure 6.20: Sensitivity testing of scope tip coil

Legend: The left of the figure shows the experiment arrangement. The right of the figure shows the resulting axial MR image. Note the very high intensity spenco fiducial within the bore of the coil.

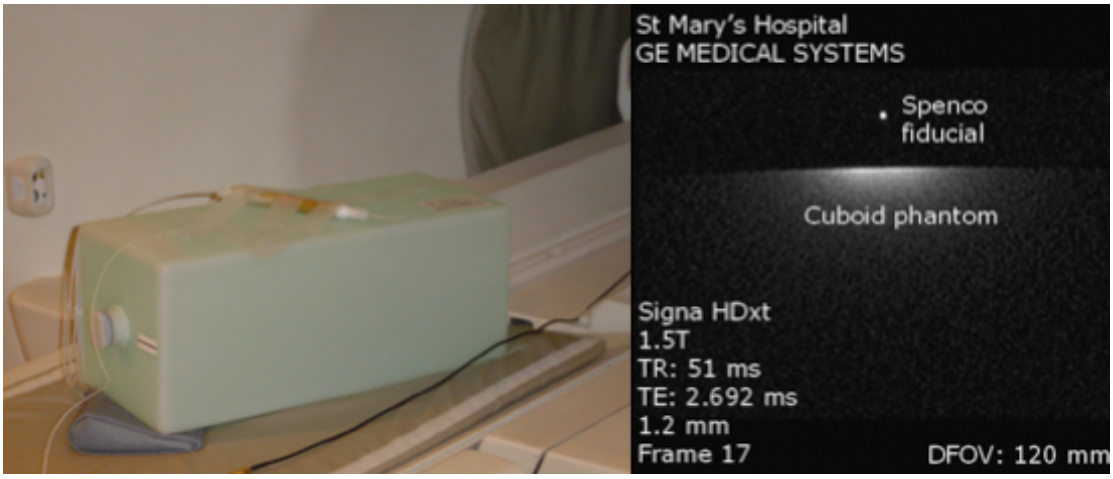


Figure 6.21: Uniformity testing of scope tip coil

Legend: Experimental arrangement is shown in the left of the figure, with the coil mounted in the PEEK tank, filled with water. The right of the figure shows the resulting axial MR image, with excellent uniformity. The tiny water fiducial can be seen in the centre of the coil, as can intense signal associated with the conductors.

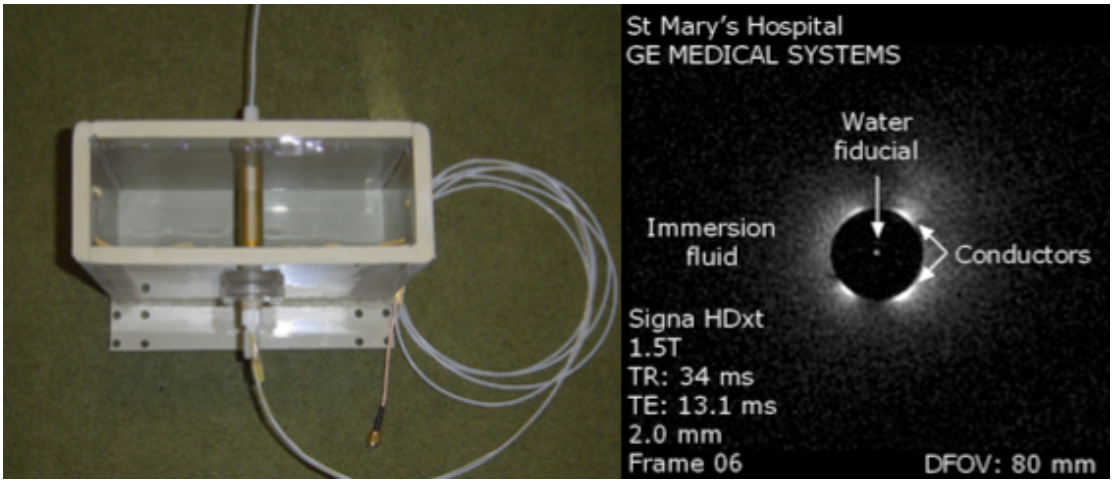


Figure 6.22: Porcine liver/duodenal imaging with scope tip coil

Legend: The left of the figure illustrates the positioning of the (covered) detector coil within the porcine duodenum. The right of the figure is a representative axial MR image showing good resolution of duodenal wall features.

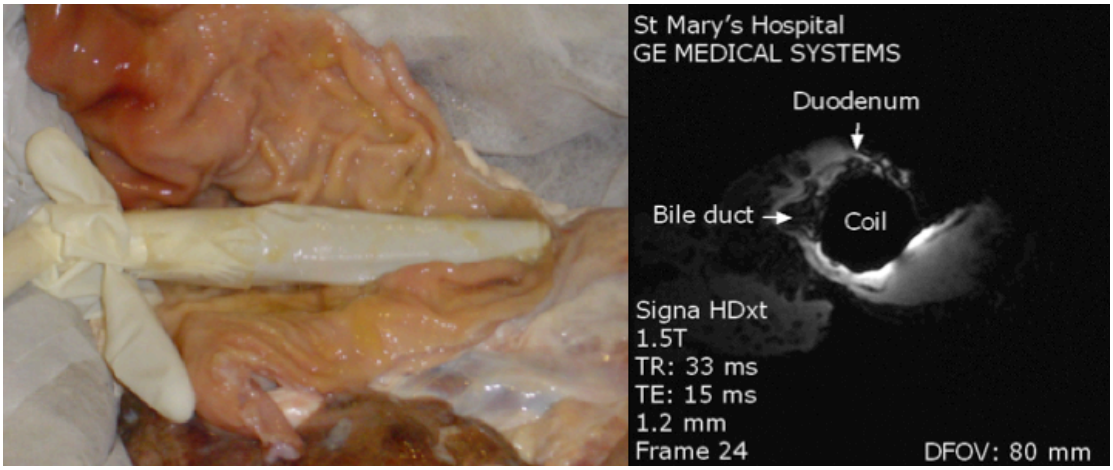


Figure 6.23 Motion artefact generator

Legend: Schematic shows arrangement of the motion artefact generator located in separate areas of the MR suite. Diagram courtesy of Prof RRA Syms.

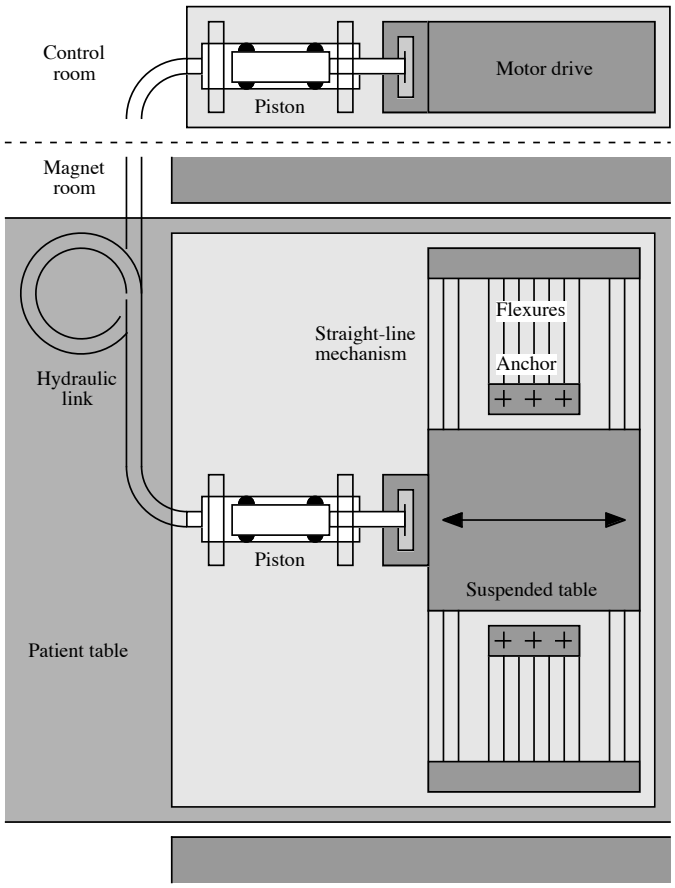


Figure 6.24: Photos of motion artefact experimental set up

Legend: The test rig can be seen before and then after loading with the porcine specimen and placement in the MR bore.

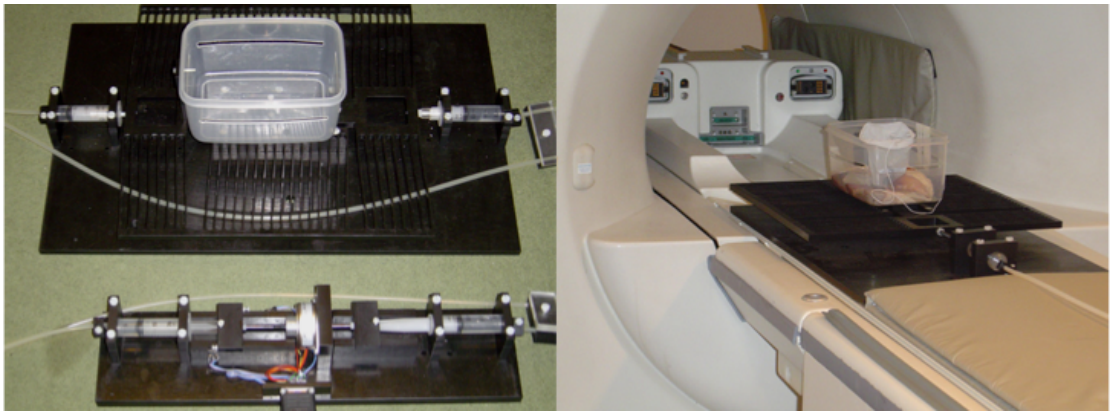


Figure 6.25: Coronal images before and during motion simulations

Legend: Images on the left are before activation of the motion generator, degradation of images once motion was generated can be seen in images on the right

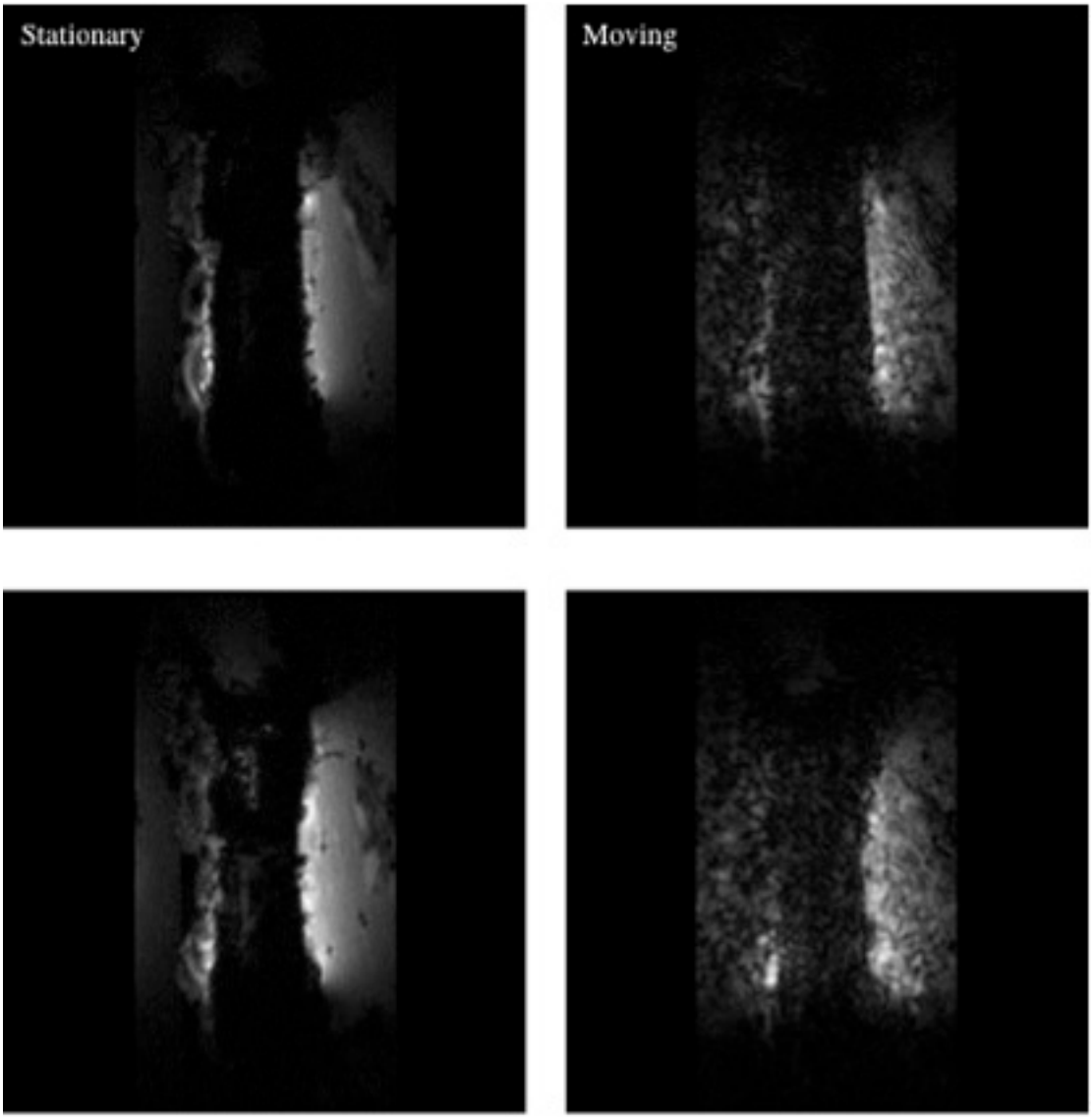


Figure 6.26: EUS imaging of ex vivo porcine duodenum

Legend: The arrangement of the EUS scope in the duodenum can be seen on the left. A representative axial EUS image is provided on the right. EUS settings were 20MHz, 6cm FOV

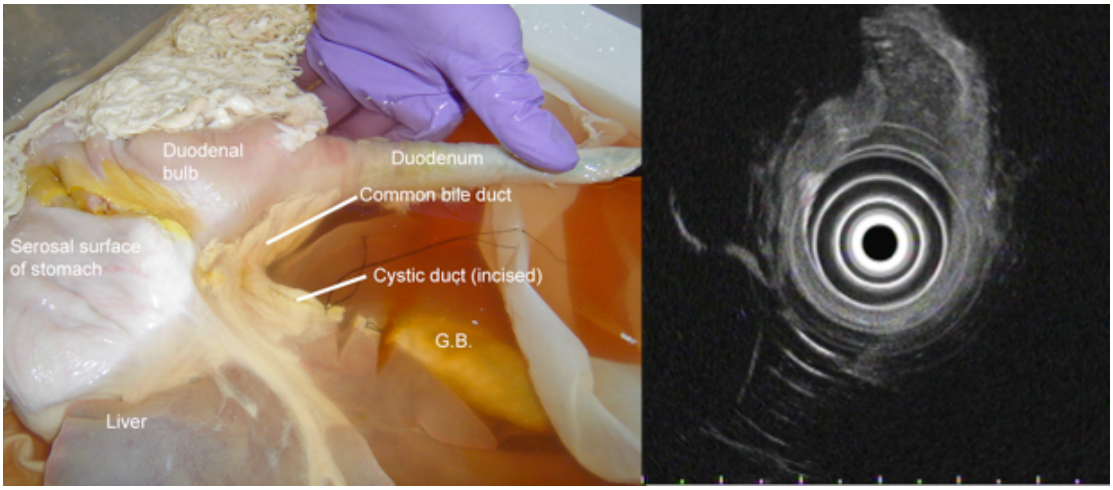


Figure 6.27: MR and EUS views of porcine duodenum

Legend: Recapitulation of comparable MR microcoil (left) and EUS (right) images obtained from porcine duodenum scanning

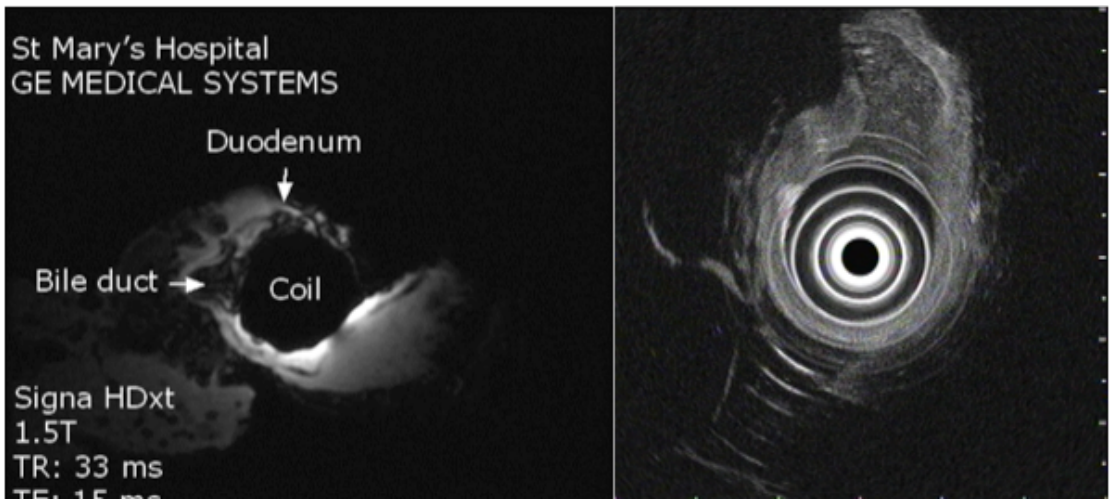


Figure 6.28: Integration of endoscopic and MR imaging

Legend: The experimental set up is illustrated in the left side of the figure. A representative endoscopic video capture image is shown on the right.

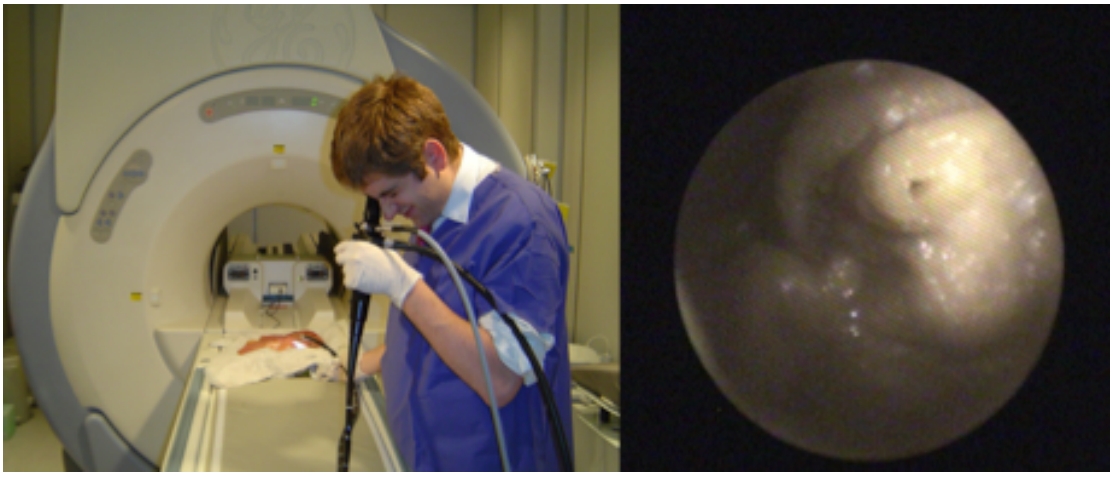
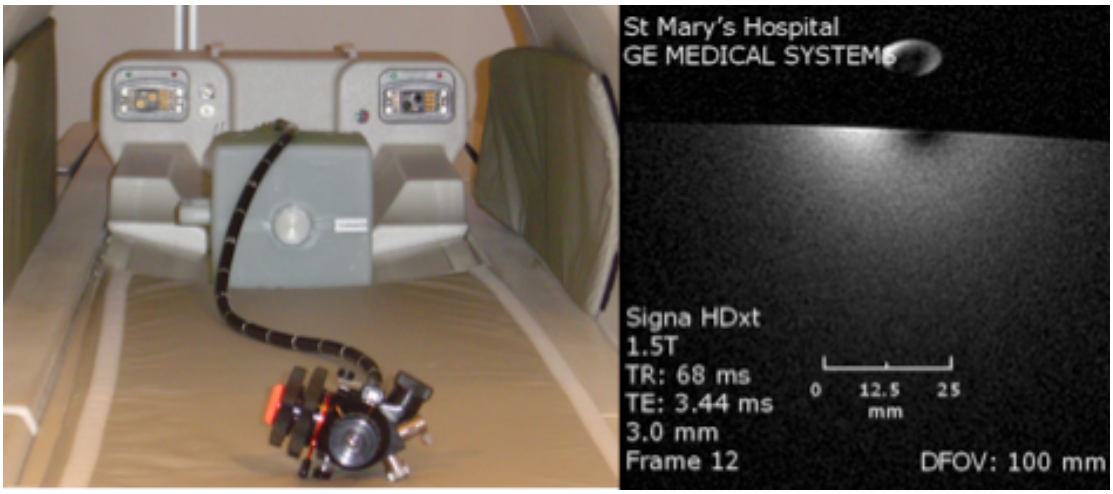


Figure 6.29: MR imaging with coil running through MR compatible duodenoscope

Legend: The experimental set up is shown on the left, minimal artefactual distortion is seen in the representative axial image, shown on the right.



7. Summary and conclusion

In pursuit of my stated overarching aim - 'to improve understanding of the aetiology of cholangiocarcinoma and enhance its diagnosis and staging' – I have undertaken experiments in three different fields.

I have performed two genetic studies. The first investigated germline polymorphisms in five genes encoding biliary transporter proteins. Polymorphisms in three of these genes, *ABCB11*, A*BCB4* and *ATP8B1* were nominally associated with increased risk of CC. The positive findings in one gene, *ATP8B1*, persisted after Bonferroni correction for multiple testing. I believe that the positive findings in these three transporters warrant further investigation in a validation cohort. A collaboration between the Cholangiocarcinoma Group at Imperial College London and counterparts at the Mayo Clinic, MA, USA has been established. A multicentre, multinational genome wide association study for CC risk factors is planned. Validation of the biliary transporter associations presented in this thesis is to be attempted in their substantial Biobank of DNA samples from patients with sporadic CC.

The second genetic study investigated polymorphisms in the NKG2D receptor as a risk factor for sporadic CC. Genetic polymorphisms in this his receptor have previously been implicated in PSC related CC.[72] The study presented here was more than adequately powered to detect the same effect seen in the prior study of PSC related CC, but the association was not replicated. This is an important negative finding and may point to important differences in pathogenesis between sporadic CC and PSC related CC.

Three protein studies have been undertaken. The first utilised a GeLC-MS/MS proteomic approach to investigate and identify differentially expressed proteins in blood plasma from patients with CC and patients with benign, gallstone disease. Six proteins of interest were identified. Interestingly, several of these have previously been associated with malignancy. Of particular interest, the protein LRG1 has previously

been identified as elevated in the plasma of CC patients.[383] Several of the identified proteins warrant further investigation in validation cohorts, and this is being pursued.

The second protein study employed a further proteomic technique, SELDI-TOF MS. This technique is high throughput and so large cohorts were investigated: 99 CC patients, 64 PSC and 107 healthy controls. To ensure avoidance of Type I error, separate discovery and validation studies were undertaken. Seven significant *m/z* peaks were found, all were validated in the second phase of the study. These seven peaks are likely to represent five unique proteins. Three of those five unique proteins are close *m/z* matches with those identified in a previous SELDI-TOF MS study of plasma from CC patients.[360] The strongest of the five unique peaks appears to a novel finding in studies of hepatobiliary malignancy: *m/z* 4631. Given the strength of the findings, and partial replication of a previous high quality study, all five of these proteins warrant identification and further investigation. Again, this work is underway.

The third protein study investigated a plasma protein, NGAL, as a biomarker for CC. This protein has previously been shown to elevated in bile samples from CC patients.[364] 97 CC patients, 64 PSC patients and 82 healthy controls were investigated. NGAL was not of utility in discriminating CC from PSC subjects in these cohorts. However, NGAL was elevated in both CC and PSC groups and showed reasonable utility in discriminating both disease groups from healthy controls. The difference between the groups was not attributable to sepsis, renal impairment, cholestasis or other factors investigated.

In the final study presented in this thesis, the development of two new RF microcoil detector devices is described. One is catheter based and intended for insertion into the biliary tree via a duodenoscope at ERCP. The second is to be integrated into an MR compatible duodenoscope. Both devices act as detector coils for an attached MR system. They permit high-resolution radial imaging of tissues to which they are closely apposed. Results from *in vitro* and *ex vivo* testing of these devices are presented. Very

high-resolution imaging has been demonstrated, with favourable performance when compared to contemporaneous EUS imaging. These devices were developed as part of a wider project, investigating endoscopy within the MR scanner. Work is on going, and clinical studies are planned to confirm that the very promising *ex vivo* imaging can be replicated *in vivo*.

The diversity of the studies I have undertaken and present in this thesis has required an understanding of a wide variety of analytical techniques, investigative strategies and scientific concepts. The overarching aim has been to enhance the diagnosis of cholangiocarcinoma. In the course of the research period collaborations have been formed with a number of other UK and international institutions, which remain active. Biobank ethical approval has been put in place and a UK multicentre sample collection CC biomarker programme has received UKCRN support. This will offer the opportunity for validation studies of the positive findings in the genetic and proteomic studies presented in this thesis.

Cholangiocarcinoma remains a relatively poorly understood disease, is an increasingly common diagnosis and carries a high mortality. Definitive diagnosis generally requires a multitude of investigations and treatment remains largely palliative. The complexity of diagnosis and management of patients with CC requires highly skilled surgeons, hepatologists, pathologists, radiologists, endoscopists and oncologists to collaborate as a multidisciplinary team. The clinical priority is to accurately identify the minority of patients who will benefit from prompt, curative resection and to optimise the palliative care of the majority who will not.

International and national scientific collaborations are crucial if large studies are to be conducted into this uncommon disease, and new therapeutic and diagnostic techniques are to emerge.

List of related publications to date

Papers

Dixon PH, Wadsworth CA, Williamson C et al. A comprehensive analysis of common genetic variation around six candidate loci for Intrahepatic Cholestasis of Pregnancy. Hepatol. Submitted December 2012.

Wadsworth CA, Dixon P, Taylor-Robinson SD et al. Common genetic variation in natural killer cell receptor protein G2D and susceptibility to sporadic cholangiocarcinoma. J Hepatol. Submitted May 2012.

Wadsworth CA, Taylor-Robinson SD, Khan SA. Risk factors and diagnosis of cholangiocarcinoma. Hepatology International. In press.

Andresen K, Boberg KM, Vedeld HM, Honne H, Hektoen M, Wadsworth CA, Clausen OP, Karlsen TM, Foss A, Mathisen O, Schrumpf E, Lothe RA, Lind GE. Novel target genes and a valid biomarker panel identified for cholangiocarcinoma. Epigenetics. 2012 Nov;7(11):1249-57.

North OJ, Ristic M, Wadsworth CA, Young IR, Taylor-Robinson SD. Design and Evaluation of Endoscope Remote Actuator for MRI-Guided Endoscopic Retrograde Cholangio-Pancreatography. Journal of Institute of Electrical and Electronic Engineers. In press.

Wadsworth CA, Dixon PH, Wong JH, Chapman M, S.C. M, Sharif A, Spalding DR, Pereira SP, Thomas HC, Taylor-Robinson SD, Whittaker J, Williamson C, Khan SA. Genetic Factors in the Pathogenesis of Cholangiocarcinoma. Dig Dis. 2011;29(1):93-7

Zabron AA, Horneffer-van der Sluis V, Wadsworth CA, Laird F, Gierula M, Thillainayagam AV, Vlavianos P, Westaby D, Taylor-Robinson SD, Edwards RJ, Khan SA. Elevated Levels of Neutrophil Gelatinase-Associated Lipocalin in Bile From

Patients With Malignant Pancreatobiliary Disease. Am J Gastroenterol. 2011 Sep;106(9):1711-7

Syms RRA, Young IR, Ahmad MM, Rea M, Wadsworth CA, Taylor-Robinson SD. Thin-Film Detector System for Internal Magnetic Resonance Imaging Sensors and Actuators A: Physical. [doi: 10.1016/j.sna.2010.05.017]. 2010;163(1):15-24.

Ahmad MM, Syms RRA, Young IR, Mathew B, Casperz W, Taylor-Robinson SD, Wadsworth CA et al. Catheter-based flexible microcoil RF detectors for internal magnetic resonance imaging. J Micromech Microeng. 2009 Jul;19(7):-.

Abstracts

Wadsworth CA, Horneffer-van der Sluis VM, Zabron A, Boberg KM, Pereira SP, Taylor-Robinson SD, Edwards R, Khan SA. Increased levels of neutrophil gelatinase associated lipocalin (NGAL) in the plasma of cholangiocarcinoma patients. **Poster of distinction:** BSG/DDF Meeting, Liverpool, 2012. Abstract: Gut 2012;61:A350-A351 doi:10.1136/gutjnl-2012-302514d.133 **Poster presentation:** EASL International Liver Congress 2012, Barcelona. Abstract: J Hepatol. Volume 56, Supplement 2, Page S130, April 2012

Latif A, Zabron A, Horneffer Van-der-Sluis VM, Wadsworth CA, Westaby D, Vlavianos P, Taylor-Robinson SD, Edwards RJ, Khan SA. Biliary neutrophil gelatinase-associated lipocalin and its interaction with matrix metalloproteinase 9 in patients with malignant jaundice. **Poster presentation:** BSG/DDF Meeting, Liverpool 2012. Abstract: Volume 56, Gut 2012;61:A350-A351.

Wadsworth CA, Dixon PH, Wong JH, Chapman M, S.C. M, Sharif A, Spalding DR, Pereira SP, Thomas HC, Taylor-Robinson SD, Whittaker J, Williamson C, Khan SA. Genetic variation in natural killer cell receptor protein G2D does not modify susceptibility to sporadic cholangiocarcinoma. **Poster of distinction:** BSG Annual Meeting, Birmingham, March 2011. Abstract: Gut. 2011;60(Suppl 1)A117. **Poster**

presentation: EASL International Liver Congress 2011, Berlin. Abstract: Journal of Hepatology. 2011;54:S108-S9.

Zabron A, Horneffer-van der Sluis V, Wadsworth CA, Gierula M, Thillainayagam AV, Vlavianos P, Westaby D, Taylor-Robinson SD, Edwards R, Khan SA. Neutrophil gelatinase-associated lipocalin is elevated in bile from patients with malignant pancreatobiliary disease. **Oral presentation:** BSG Annual Meeting, Birmingham, March 2011. Abstract: Gut. 2011;60(Suppl 1):A7.

Wadsworth CA, Dixon PH, Wong JH, Chapman MH, McKay SC, Sharif A, Spalding DR, Pereira SP, Thomas H, Taylor-Robinson SD, Whittaker J, Williamson C, Khan SA. Genetic variation in biliary transporters as a susceptibility factor for cholangiocarcinoma. **Oral presentation:** XXI International Bile Acid Meeting, Freiberg, Oct 2010. **Poster presentation:** EASL International Liver Congress 2011, Berlin. Abstract: Journal of Hepatology. 2011;54:S109. **Poster presentation:** BASL meeting, Edinburgh, September 2010. Abstract: Gut, September 2010, Volume 59, SII;A27

Dixon PH, Wadsworth CA, Chambers J, Donnelly J, Cooley SM, Jarvis S, Kubitz R, Lammert F, Marschall HU, Glantz A, Khan SA, Whittaker J, Geary M, Williamson C. The role of common genetic variation around six candidate loci for susceptibility to intrahepatic cholestasis of pregnancy. **Poster presentation:** XXI International Bile Acid Meeting, Freiberg, Oct 2010. **Poster presentation:** AASLD meeting, Boston, October 2010. Abstract: Hepatology. 2010;52(4) (suppl)):1.

Zabron A, Horneffer-Van der Sluis V, Gierula, M, Wadsworth C, et al Proteomic analysis of bile in pancreatic adenocarcinoma. **Oral Presentation:** UEGW Meeting, Barcelona, October 2010. Abstract: Gut 2010; 59 (Suppl III) A33

Zabron AA, Horneffer-Van der Sluis VM, Geirula M, Wadsworth CA, Laird F, Thillainayagam A, Vlavianos P, Westaby D, Taylor-Robinson S, Edwards RJ, Khan SA. Proteomic analysis of bile in benign and malignant pancreatobiliary disease. **Poster**

presentation: Annual Scientific Meeting of the BSG, Liverpool, March 2010. Abstract: Gut 59 SI:A82-A83 (Apr 2010)

Syms RRA, Young IR, Ahmad MM, Wadsworth C, Fitzpatrick J. Periodic interconnects for internal MRI. **Oral presentation:** 4th International Congress on Advanced Electromagnetic Materials in Microwaves and Optics, 13-16 Sept., Karlsruhe, Germany. **Abstract:** Proceedings of the 4th International Congress on Advanced Electromagnetic Materials in Microwaves and Optics pp 89-91 (2010)

Wadsworth CA, Khan S.A., Taylor-Robinson S.D. et al. MR imaging of the human biliary tree using a flexible catheter-mounted radio-frequency detector microcoil. **Poster presentation:** ISMRM 2010, Stockholm, May 1-7, paper 2617 (2010). Abstract: Proc. ISMRM Conf., Sweden, May 1-7 pp.2617 (2010)

Laird F, Edwards RJ, Taylor-Robinson SD, Zabron A, Wadsworth CA, et al. Proteomic analysis of bile in benign and malignant pancreatobiliary disease. **Poster presentation:** UEGW meeting, London, October 2009. Abstract: Gut 2009; 58 (Suppl II) P1411

Wadsworth CA, Young IR, Rea M, Ahmad MM, Khorsandi SE et al. A Catheter-Mounted Magnetic Resonance Detector Coil For Biliary Imaging: First Ex Vivo Human Hepatobiliary Images. **Poster presentation:** AASLD Annual Meeting, Boston, 2009. Abstract: Hepatology. 2009;50(S4):740A. **Poster presentation:** BASL Annual Meeting, London, 2009. Abstract: Proceedings of the BASL Annual Scientific Meeting 2009

Wadsworth CA, Young IR, Khan SA, Sundblom L, Ahmad MM, Taylor-Robinson SD, Syms RRA. Imaging Of An Ex Vivo Pancreatic Adenocarcinoma With A Catheter-Mounted Magnetic Resonance Detector Microcoil. **Poster presentation:** UEGW meeting, London, October 2009. Abstract: Endoscopy 2009; 41 (Suppl 1) P1389

Wadsworth C, Syms R, Khan S, Dickinson R, et al. A Catheter-Mounted Magnetic Resonance Detector Coil for Biliary Imaging: First in vitro biliary images. **Poster**

presentation: EASL International Liver Congress, Copenhagen, 2009. Abstract: Journal of Hepatology, Volume 50, Supplement 1, April 2009, Pages S302-S303

Poster presentation: BSG Scientific Meeting, Glasgow, 2009. Abstract: Gut 2009;58 (Suppl I) :A147

References

1. Durand-Fardel M. Recherches anat-path. sur la vesicule et les canaux biliaires. Archives générales de Médecine (Arch gen de Med). 1840;viii:167.

2. de Groen PC, Gores GJ, LaRusso NF, et al. Biliary tract cancers. N Engl J Med. 1999 Oct 28;341(18):1368-78.

3. West J, Wood H, Logan RF, et al. Trends in the incidence of primary liver and biliary tract cancers in England and Wales 1971-2001. Br J Cancer. 2006 Jun 5;94(11):1751-8.

4. Henson DE, Albores-Saavedra J, Corle D. Carcinoma of the extrahepatic bile ducts. Histologic types, stage of disease, grade, and survival rates. Cancer. 1992 Sep 15;70(6):1498-501.

5. Shaib Y, El-Serag HB. The epidemiology of cholangiocarcinoma. Semin Liver Dis. 2004 May;24(2):115-25.

6. McLean L, Patel T. Racial and ethnic variations in the epidemiology of intrahepatic cholangiocarcinoma in the United States. Liver Int. 2006 Nov;26(9):1047-53.

7. Taylor-Robinson SD, Toledano MB, Arora S, et al. Increase in mortality rates from intrahepatic cholangiocarcinoma in England and Wales 1968-1998. Gut. 2001 Jun;48(6):816-20.

8. Khan SA, Taylor-Robinson SD, Toledano MB, et al. Changing international trends in mortality rates for liver, biliary and pancreatic tumours. J Hepatol. 2002 Dec;37(6):806-13.

9. Khan SA, Thomas HC, Davidson BR, et al. Cholangiocarcinoma. Lancet. 2005 Oct 8;366(9493):1303-14.

10. Khan SA, Emadossadaty S, Ladep NG, et al. Rising trends in cholangiocarcinoma: is the ICD classification system misleading us? J Hepatol. 2012 Apr;56(4):848-54.

11. Khan SA, Toledano MB, Taylor-Robinson SD. Epidemiology, risk factors, and pathogenesis of cholangiocarcinoma. HPB (Oxford). 2008;10(2):77-82.

12. Shaib YH, Davila JA, Henderson L, et al. Endoscopic and surgical therapy for intrahepatic cholangiocarcinoma in the united states: a population-based study. J Clin Gastroenterol. 2007 Nov-Dec;41(10):911-7.

13. Ortner MA. Photodynamic therapy for cholangiocarcinoma: overview and new developments. Curr Opin Gastroenterol. 2009 Sep;25(5):472-6.

14. 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 Oct;5(10):1221-8.

15. Shaib YH, El-Serag HB, Davila JA, et al. Risk factors of intrahepatic cholangiocarcinoma in the United States: a case-control study. Gastroenterology. 2005 Mar;128(3):620-6.

16. Welzel TM, Mellemkjaer L, Gloria G, et al. Risk factors for intrahepatic cholangiocarcinoma in a low-risk population: a nationwide case-control study. Int J Cancer. 2007 Feb 1;120(3):638-41.

17. LaRusso NF, Shneider BL, Black D, et al. Primary sclerosing cholangitis: summary of a workshop. Hepatology. 2006 Sep;44(3):746-64.

18. Mairiang E, Haswell-Elkins MR, Mairiang P, et al. Reversal of biliary tract abnormalities associated with Opisthorchis viverrini infection following praziquantel treatment. Trans R Soc Trop Med Hyg. 1993 Mar-Apr;87(2):194-7.

19. Maclean J. Liver, lung, and intestinal fluke infections. 2nd ed. Guerrant R, editor. Philadelphia: Chirchill Livingstone; 2006.

20. Kubo S, Kinoshita H, Hirohashi K, et al. Hepatolithiasis associated with cholangiocarcinoma. World J Surg. 1995 Jul-Aug;19(4):637-41.

21. Lesurtel M, Regimbeau JM, Farges O, et al. Intrahepatic cholangiocarcinoma and hepatolithiasis: an unusual association in Western countries. Eur J Gastroenterol Hepatol. 2002 Sep;14(9):1025-7.

22. Tyson GL, El-Serag HB. Risk factors for cholangiocarcinoma. Hepatology. 2011 Jul;54(1):173-84.

23. Zhou YM, Yin ZF, Yang JM, et al. Risk factors for intrahepatic cholangiocarcinoma: a case-control study in China. World J Gastroenterol. 2008 Jan 28;14(4):632-5.

24. Lee TY, Lee SS, Jung SW, et al. Hepatitis B virus infection and intrahepatic cholangiocarcinoma in Korea: a case-control study. Am J Gastroenterol. 2008 Jul;103(7):1716-20.

25. Blechacz BR, Gores GJ. Cholangiocarcinoma. Clin Liver Dis. 2008 Feb;12(1):131-50, ix.

26. Soreide K, Korner H, Havnen J, et al. Bile duct cysts in adults. Br J Surg. 2004 Dec;91(12):1538-48.

27. Lipsett PA, Pitt HA, Colombani PM, et al. Choledochal cyst disease. A changing pattern of presentation. Ann Surg. 1994 Nov;220(5):644-52.

28. Scott J, Shousha S, Thomas HC, et al. Bile duct carcinoma: a late complication of congenital hepatic fibrosis. Case report and review of literature. Am J Gastroenterol. 1980 Feb;73(2):113-9.

29. Kobayashi M, Ikeda K, Saitoh S, et al. Incidence of primary cholangiocellular carcinoma of the liver in Japanese patients with hepatitis C virus-related cirrhosis. Cancer. 2000 Jun 1;88(11):2471-7.

30. 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 Jan;49(1):116-23.

31. Shin HR, Lee CU, Park HJ, et al. Hepatitis B and C virus, Clonorchis sinensis for the risk of liver cancer: a case-control study in Pusan, Korea. Int J Epidemiol. 1996 Oct;25(5):933-40.

32. Yamamoto S, Kubo S, Hai S, et al. Hepatitis C virus infection as a likely etiology of intrahepatic cholangiocarcinoma. Cancer Sci. 2004 Jul;95(7):592-5.

33. Shaib YH, El-Serag HB, Nooka AK, et al. Risk factors for intrahepatic and extrahepatic cholangiocarcinoma: a hospital-based case-control study. Am J Gastroenterol. 2007 May;102(5):1016-21.

34. Donato F, Gelatti U, Tagger A, et al. Intrahepatic cholangiocarcinoma and hepatitis C and B virus infection, alcohol intake, and hepatolithiasis: a case-control study in Italy. Cancer Causes Control. 2001 Dec;12(10):959-64.

35. Torbenson M, Yeh MM, Abraham SC. Bile duct dysplasia in the setting of chronic hepatitis C and alcohol cirrhosis. Am J Surg Pathol. 2007 Sep;31(9):1410-3.

36. Sorensen HT, Friis S, Olsen JH, et al. Risk of liver and other types of cancer in patients with cirrhosis: a nationwide cohort study in Denmark. Hepatology. 1998 Oct;28(4):921-5.

37. Khan SA, Carmichael PL, Taylor-Robinson SD, et al. DNA adducts, detected by 32P postlabelling, in human cholangiocarcinoma. Gut. 2003 Apr;52(4):586-91.

38. Sahani D, Prasad SR, Tannabe KK, et al. Thorotrast-induced cholangiocarcinoma: case report. Abdom Imaging. 2003 Jan-Feb;28(1):72-4.

39. Abrahamsson A, Krapivner S, Gustafsson U, et al. Common polymorphisms in the CYP7A1 gene do not contribute to variation in rates of bile acid synthesis and plasma LDL cholesterol concentration. Atherosclerosis. 2005 Sep;182(1):37-45.

40. Tao LY, He XD, Qu Q, et al. Risk factors for intrahepatic and extrahepatic cholangiocarcinoma: a case-control study in China. Liver Int. 2009 Oct 14.

41. Grainge MJ, West J, Solaymani-Dodaran M, et al. The antecedents of biliary cancer: a primary care case-control study in the United Kingdom. Br J Cancer. 2009 Jan 13;100(1):178-80.

42. Hardell L, Bengtsson NO, Jonsson U, et al. Aetiological aspects on primary liver cancer with special regard to alcohol, organic solvents and acute intermittent porphyria--an epidemiological investigation. Br J Cancer. 1984 Sep;50(3):389-97.

43. Mecklin JP, Jarvinen HJ, Virolainen M. The association between cholangiocarcinoma and hereditary nonpolyposis colorectal carcinoma. Cancer. 1992 Mar 1;69(5):1112-4.

44. Lee SS, Kim MH, Lee SK, et al. Clinicopathologic review of 58 patients with biliary papillomatosis. Cancer. 2004 Feb 15;100(4):783-93.

45. Hildebrandt F. Genetic kidney diseases. Lancet. 2010 Apr 10;375(9722):1287-95.

46. Gunay-Aygun M. Liver and kidney disease in ciliopathies. Am J Med Genet C Semin Med Genet. 2009 Nov 15;151C(4):296-306.

47. Onuchic LF, Furu L, Nagasawa Y, et al. PKHD1, the polycystic kidney and hepatic disease 1 gene, encodes a novel large protein containing multiple immunoglobulin-like plexin-transcription-factor domains and parallel beta-helix 1 repeats. Am J Hum Genet. 2002 May;70(5):1305-17.

48. Jacquemin E. Role of multidrug resistance 3 deficiency in pediatric and adult liver disease: one gene for three diseases. Semin Liver Dis. 2001 Nov;21(4):551-62.

49. Thompson R, Strautnieks S. BSEP: function and role in progressive familial intrahepatic cholestasis. Semin Liver Dis. 2001 Nov;21(4):545-50.

50. Knisely AS, Strautnieks SS, Meier Y, et al. Hepatocellular carcinoma in ten children under five years of age with bile salt export pump deficiency. Hepatology. 2006 Aug;44(2):478-86.

51. Meier Y. Hepatocellular malignancy in ABCB11/BStP disease (progressive familial intrahepatic cholestasis, type 2): four patients. Hepatology. 2004;40S:471A.

52. Scheimann AO, Strautnieks SS, Knisely AS, et al. Mutations in bile salt export pump (ABCB11) in two children with progressive familial intrahepatic cholestasis and cholangiocarcinoma. J Pediatr. 2007 May;150(5):556-9.

53. Katsika D, Grjibovski A, Einarsson C, et al. Genetic and environmental influences on symptomatic gallstone disease: a Swedish study of 43,141 twin pairs. Hepatology. 2005 May;41(5):1138-43.

54. Grunhage F, Acalovschi M, Tirziu S, et al. Increased gallstone risk in humans conferred by common variant of hepatic ATP-binding cassette transporter for cholesterol. Hepatology. 2007 Sep;46(3):793-801.

55. Buch S, Schafmayer C, Volzke H, et al. A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. Nat Genet. 2007 Aug;39(8):995-9.

56. Renner O, Harsch S, Schaeffeler E, et al. A variant of the SLC10A2 gene encoding the apical sodium-dependent bile acid transporter is a risk factor for gallstone disease. PLoS One. 2009;4(10):e7321.

57. Kovacs P, Kress R, Rocha J, et al. Variation of the gene encoding the nuclear bile salt receptor FXR and gallstone susceptibility in mice and humans. J Hepatol. 2008 Jan;48(1):116-24.

58. Rosmorduc O, Hermelin B, Boelle PY, et al. ABCB4 gene mutation-associated cholelithiasis in adults. Gastroenterology. 2003 Aug;125(2):452-9.

59. Lucena JF, Herrero JI, Quiroga J, et al. A multidrug resistance 3 gene mutation causing cholelithiasis, cholestasis of pregnancy, and adulthood biliary cirrhosis. Gastroenterology. 2003 Apr;124(4):1037-42.

60. Lammert F, Wang DQ, Hillebrandt S, et al. Spontaneous cholecysto- and hepatolithiasis in Mdr2-/- mice: a model for low phospholipid-associated cholelithiasis. Hepatology. 2004 Jan;39(1):117-28.

61. van Mil SW, van der Woerd WL, van der Brugge G, et al. Benign recurrent intrahepatic cholestasis type 2 is caused by mutations in ABCB11. Gastroenterology. 2004 Aug;127(2):379-84.

62. Schrumpf E, Fausa O, Forre O, et al. HLA antigens and immunoregulatory T cells in ulcerative colitis associated with hepatobiliary disease. Scand J Gastroenterol. 1982 Mar;17(2):187-91.

63. Karlsen TH, Schrumpf E, Boberg KM. Genetic epidemiology of primary sclerosing cholangitis. World J Gastroenterol. 2007 Nov 7;13(41):5421-31.

64. Pauli-Magnus C, Kerb R, Fattinger K, et al. BSEP and MDR3 haplotype structure in healthy Caucasians, primary biliary cirrhosis and primary sclerosing cholangitis. Hepatology. 2004 Mar;39(3):779-91.

65. Karlsen TH, Franke A, Melum E, et al. Genome-wide association analysis in primary sclerosing cholangitis. Gastroenterology. 2010 Mar;138(3):1102-11.

66. Marahatta SB, Punyarit P, Bhudisawasdi V, et al. Polymorphism of glutathione S-transferase omega gene and risk of cancer. Cancer Lett. 2006 May 18;236(2):276-81.

67. Wiencke K, Louka AS, Spurkland A, et al. Association of matrix metalloproteinase-1 and -3 promoter polymorphisms with clinical subsets of Norwegian primary sclerosing cholangitis patients. J Hepatol. 2004 Aug;41(2):209-14.

68. Ko KH, Kim NK, Yim DJ, et al. Polymorphisms of 5,10-methylenetetrahydrofolate reductase (MTHFR C677T) and thymidylate synthase enhancer region (TSER) as a risk factor of cholangiocarcinoma in a Korean population. Anticancer Res. 2006 Nov-Dec;26(6B):4229-33.

69. Prawan A, Kukongviriyapan V, Tassaneeyakul W, et al. Association between genetic polymorphisms of CYP1A2, arylamine N-acetyltransferase 1 and 2 and susceptibility to cholangiocarcinoma. Eur J Cancer Prev. 2005 Jun;14(3):245-50.

70. Mihalache F, Hoblinger A, Grunhage F, et al. Heterozygosity for the alpha1-antitrypsin Z allele may confer genetic risk of cholangiocarcinoma. Aliment Pharmacol Ther. 2011 Feb;33(3):389-94.

71. Hoblinger A, Grunhage F, Sauerbruch T, et al. Association of the c.3972C>T variant of the multidrug resistance-associated protein 2 Gene (MRP2/ABCC2) with susceptibility to bile duct cancer. Digestion. 2009;80(1):36-9.

72. Melum E, Karlsen TH, Schrumpf E, et al. Cholangiocarcinoma in primary sclerosing cholangitis is associated with NKG2D polymorphisms. Hepatology. 2008 Jan;47(1):90-6.

73. Nomoto K, Tsuneyama K, Cheng C, et al. Intrahepatic cholangiocarcinoma arising in cirrhotic liver frequently expressed p63-positive basal/stem-cell phenotype. Pathol Res Pract. 2006;202(2):71-6.

74. Sell S, Dunsford HA. Evidence for the stem cell origin of hepatocellular carcinoma and cholangiocarcinoma. Am J Pathol. 1989 Jun;134(6):1347-63.

75. Zen Y, Adsay NV, Bardadin K, et al. Biliary intraepithelial neoplasia: an international interobserver agreement study and proposal for diagnostic criteria. Mod Pathol. 2007 Jun;20(6):701-9.

76. Kloppel G, Kosmahl M. Is the intraductal papillary mucinous neoplasia of the biliary tract a counterpart of pancreatic papillary mucinous neoplasm? J Hepatol. 2006 Feb;44(2):249-50.

77. Jaiswal M, LaRusso NF, Nishioka N, et al. Human Ogg1, a protein involved in the repair of 8-oxoguanine, is inhibited by nitric oxide. Cancer Res. 2001 Sep 1;61(17):6388-93.

78. Jaiswal M, LaRusso NF, Burgart LJ, et al. Inflammatory cytokines induce DNA damage and inhibit DNA repair in cholangiocarcinoma cells by a nitric oxide-dependent mechanism. Cancer Res. 2000 Jan 1;60(1):184-90.

79. Sirica AE, Nathanson MH, Gores GJ, et al. Pathobiology of biliary epithelia and cholangiocarcinoma: proceedings of the Henry M. and Lillian Stratton Basic Research Single-Topic Conference. Hepatology. 2008 Dec;48(6):2040-6.

80. Liu Z, Sakamoto T, Ezure T, et al. Interleukin-6, hepatocyte growth factor, and their receptors in biliary epithelial cells during a type I ductular reaction in mice: interactions between the periductal inflammatory and stromal cells and the biliary epithelium. Hepatology. 1998 Nov;28(5):1260-8.

81. Matsumoto K, Fujii H, Michalopoulos G, et al. Human biliary epithelial cells secrete and respond to cytokines and hepatocyte growth factors in vitro: interleukin-6, hepatocyte growth factor and epidermal growth factor promote DNA synthesis in vitro. Hepatology. 1994 Aug;20(2):376-82.

82. Okada K, Shimizu Y, Nambu S, et al. Interleukin-6 functions as an autocrine growth factor in a cholangiocarcinoma cell line. J Gastroenterol Hepatol. 1994 Sep-Oct;9(5):462-7.

83. 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 Dec;42(6):1329-38.

84. Isomoto H, Mott JL, Kobayashi S, et al. Sustained IL-6/STAT-3 signaling in cholangiocarcinoma cells due to SOCS-3 epigenetic silencing. Gastroenterology. 2007 Jan;132(1):384-96.

85. Rosen HR, Winkle PJ, Kendall BJ, et al. Biliary interleukin-6 and tumor necrosis factor-alpha in patients undergoing endoscopic retrograde cholangiopancreatography. Dig Dis Sci. 1997 Jun;42(6):1290-4.

86. Goydos JS, Brumfield AM, Frezza E, et al. Marked elevation of serum interleukin-6 in patients with cholangiocarcinoma: validation of utility as a clinical marker. Ann Surg. 1998 Mar;227(3):398-404.

87. Liu Y, Michalopoulos GK, Zarnegar R. Structural and functional characterization of the mouse hepatocyte growth factor gene promoter. J Biol Chem. 1994 Feb 11;269(6):4152-60.

88. Moghul A, Lin L, Beedle A, et al. Modulation of c-MET proto-oncogene (HGF receptor) mRNA abundance by cytokines and hormones: evidence for rapid decay of the 8 kb c-MET transcript. Oncogene. 1994 Jul;9(7):2045-52.

89. Maihofner C, Charalambous MP, Bhambra U, et al. Expression of cyclooxygenase-2 parallels expression of interleukin-1beta, interleukin-6 and NF-kappaB in human colorectal cancer. Carcinogenesis. 2003 Apr;24(4):665-71.

90. Yamagiwa Y, Meng F, Patel T. Interleukin-6 decreases senescence and increases telomerase activity in malignant human cholangiocytes. Life Sci. 2006 Apr 18;78(21):2494-502.

91. Aishima S, Fujita N, Mano Y, et al. Different roles of S100P Overexpression in Intrahepatic Cholangiocarcinoma: Carcinogenesis of Perihilar Type and Aggressive Behavior of Peripheral Type. Am J Surg Pathol. 2011 Apr;35(4):590-8.

92. Liu YF, Zhao R, Guo S, et al. Expression and clinical significance of hepatoma-derived growth factor as a prognostic factor in human hilar cholangiocarcinoma. Ann Surg Oncol. 2011 Mar;18(3):872-9.

93. Terada T, Nakanuma Y, Sirica AE. Immunohistochemical demonstration of MET overexpression in human intrahepatic cholangiocarcinoma and in hepatolithiasis. Hum Pathol. 1998 Feb;29(2):175-80.

94. Hida Y, Morita T, Fujita M, et al. Clinical significance of hepatocyte growth factor and c-Met expression in extrahepatic biliary tract cancers. Oncol Rep. 1999 Sep-Oct;6(5):1051-6.

95. Aishima SI, Taguchi KI, Sugimachi K, et al. c-erbB-2 and c-Met expression relates to cholangiocarcinogenesis and progression of intrahepatic cholangiocarcinoma. Histopathology. 2002 Mar;40(3):269-78.

96. Nakazawa K, Dobashi Y, Suzuki S, et al. Amplification and overexpression of c-erbB-2, epidermal growth factor receptor, and c-met in biliary tract cancers. J Pathol. 2005 Jul;206(3):356-65.

97. Miyamoto M, Ojima H, Iwasaki M, et al. Prognostic significance of overexpression of c-Met oncoprotein in cholangiocarcinoma. Br J Cancer. 2011 Jun 28;105(1):131-8.

98. Reznik TE, Sang Y, Ma Y, et al. Transcription-dependent epidermal growth factor receptor activation by hepatocyte growth factor. Mol Cancer Res. 2008 Jan;6(1):139-50.

99. Tsubouchi H. Sustained activation of epidermal growth factor receptor in cholangiocarcinoma: a potent therapeutic target? J Hepatol. 2004 Nov;41(5):859-61.

100. Yoon JH, Gwak GY, Lee HS, et al. Enhanced epidermal growth factor receptor activation in human cholangiocarcinoma cells. J Hepatol. 2004 Nov;41(5):808-14.

101. Han C, Wu T. Cyclooxygenase-2-derived prostaglandin E2 promotes human cholangiocarcinoma cell growth and invasion through EP1 receptor-mediated activation of the epidermal growth factor receptor and Akt. J Biol Chem. 2005 Jun 24;280(25):24053-63.

102. Nzeako UC, Guicciardi ME, Yoon JH, et al. COX-2 inhibits Fas-mediated apoptosis in cholangiocarcinoma cells. Hepatology. 2002 Mar;35(3):552-9.

103. Kinami Y, Ashida Y, Gotoda H, et al. Promoting effects of bile acid load on the occurrence of cholangiocarcinoma induced by diisopropanolnitrosamine in hamsters. Oncology. 1993;50(1):46-51.

104. Yoon JH, Higuchi H, Werneburg NW, et al. Bile acids induce cyclooxygenase-2 expression via the epidermal growth factor receptor in a human cholangiocarcinoma cell line. Gastroenterology. 2002 Apr;122(4):985-93.

105. Yoon JH, Werneburg NW, Higuchi H, et al. Bile acids inhibit Mcl-1 protein turnover via an epidermal growth factor receptor/Raf-1-dependent mechanism. Cancer Res. 2002 Nov 15;62(22):6500-5.

106. Werneburg NW, Yoon JH, Higuchi H, et al. Bile acids activate EGF receptor via a TGF-alpha-dependent mechanism in human cholangiocyte cell lines. Am J Physiol Gastrointest Liver Physiol. 2003 Jul;285(1):G31-6.

107. 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 Oct;34(4 Pt 1):638-50.

108. Endo K, Yoon BI, Pairojkul C, et al. ERBB-2 overexpression and cyclooxygenase-2 up-regulation in human cholangiocarcinoma and risk conditions. Hepatology. 2002 Aug;36(2):439-50.

109. Chariyalertsak S, Sirikulchayanonta V, Mayer D, et al. Aberrant cyclooxygenase isozyme expression in human intrahepatic cholangiocarcinoma. Gut. 2001 Jan;48(1):80-6.

110. Han C, Demetris AJ, Liu Y, et al. Transforming growth factor-beta (TGF-beta) activates cytosolic phospholipase A2alpha (cPLA2alpha)-mediated prostaglandin E2 (PGE)2/EP1 and peroxisome proliferator-activated receptor-gamma (PPAR-gamma)/Smad signaling pathways in human liver cancer cells. A novel mechanism for subversion of TGF-beta-induced mitoinhibition. J Biol Chem. 2004 Oct 22;279(43):44344-54.

111. Wu T. Cyclooxygenase-2 and prostaglandin signaling in cholangiocarcinoma. Biochim Biophys Acta. 2005 Jul 25;1755(2):135-50.

112. Wu T, Han C, Lunz JG, 3rd, et al. Involvement of 85-kd cytosolic phospholipase A(2) and cyclooxygenase-2 in the proliferation of human cholangiocarcinoma cells. Hepatology. 2002 Aug;36(2):363-73.

113. Lai GH, Zhang Z, Sirica AE. Celecoxib acts in a cyclooxygenase-2-independent manner and in synergy with emodin to suppress rat cholangiocarcinoma growth in vitro through a mechanism involving enhanced Akt inactivation and increased activation of caspases-9 and -3. Mol Cancer Ther. 2003 Mar;2(3):265-71.

114. Blobe GC, Schiemann WP, Lodish HF. Role of transforming growth factor beta in human disease. N Engl J Med. 2000 May 4;342(18):1350-8.

115. Kopp HG, Placke T, Salih HR. Platelet-derived transforming growth factor-beta down-regulates NKG2D thereby inhibiting natural killer cell antitumor reactivity. Cancer Res. 2009 Oct 1;69(19):7775-83.

116. Eisele G, Wischhusen J, Mittelbronn M, et al. TGF-beta and metalloproteinases differentially suppress NKG2D ligand surface expression on malignant glioma cells. Brain. 2006 Sep;129(Pt 9):2416-25.

117. Lee JC, Lee KM, Kim DW, et al. Elevated TGF-beta1 secretion and down-modulation of NKG2D underlies impaired NK cytotoxicity in cancer patients. J Immunol. 2004 Jun 15;172(12):7335-40.

118. Zhang K, Zhaos J, Liu X, et al. Activation of NF-B upregulates Snail and consequent repression of E-cadherin in cholangiocarcinoma cell invasion. Hepatogastroenterology. 2011 Jan-Feb;58(105):1-7.

119. Ishimura N, Bronk SF, Gores GJ. Inducible nitric oxide synthase up-regulates Notch-1 in mouse cholangiocytes: implications for carcinogenesis. Gastroenterology. 2005 May;128(5):1354-68.

120. Ishimura N, Bronk SF, Gores GJ. Inducible nitric oxide synthase upregulates cyclooxygenase-2 in mouse cholangiocytes promoting cell growth. Am J Physiol Gastrointest Liver Physiol. 2004 Jul;287(1):G88-95.

121. Lazaridis KN, Gores GJ. Cholangiocarcinoma. Gastroenterology. 2005 May;128(6):1655-67.

122. Gores GJ. Cholangiocarcinoma: current concepts and insights. Hepatology. 2003 May;37(5):961-9.

123. Taniai M, Higuchi H, Burgart LJ, et al. p16INK4a promoter mutations are frequent in primary sclerosing cholangitis (PSC) and PSC-associated cholangiocarcinoma. Gastroenterology. 2002 Oct;123(4):1090-8.

124. Hahn SA, Bartsch D, Schroers A, et al. Mutations of the DPC4/Smad4 gene in biliary tract carcinoma. Cancer Res. 1998 Mar 15;58(6):1124-6.

125. Kang YK, Kim WH, Lee HW, et al. Mutation of p53 and K-ras, and loss of heterozygosity of APC in intrahepatic cholangiocarcinoma. Lab Invest. 1999 Apr;79(4):477-83.

126. Blechacz B, Gores GJ. Cholangiocarcinoma: advances in pathogenesis, diagnosis, and treatment. Hepatology. 2008 Jul;48(1):308-21.

127. Kuroki T, Tajima Y, Kanematsu T. Hepatolithiasis and intrahepatic cholangiocarcinoma: carcinogenesis based on molecular mechanisms. J Hepatobiliary Pancreat Surg. 2005;12(6):463-6.

128. Tietz PS, Larusso NF. Cholangiocyte biology. Curr Opin Gastroenterol. 2006 May;22(3):279-87.

129. Isomoto H. Epigenetic alterations associated with cholangiocarcinoma (review). Oncol Rep. 2009 Aug;22(2):227-32.

130. Baylin SB, Ohm JE. Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? Nat Rev Cancer. 2006 Feb;6(2):107-16.

131. Lim JH. Cholangiocarcinoma: morphologic classification according to growth pattern and imaging findings. AJR Am J Roentgenol. 2003 Sep;181(3):819-27.

132. Lim JH. Cholangiocarcinoma: recent advances in imaging and intervention. Abdom Imaging. 2004 Sep-Oct;29(5):538-9.

133. Ariff B, Lloyd CR, Khan S, et al. Imaging of liver cancer. World J Gastroenterol. 2009 Mar 21;15(11):1289-300.

134. Nakanuma Y, Harada K, Ishikawa A, et al. Anatomic and molecular pathology of intrahepatic cholangiocarcinoma. J Hepatobiliary Pancreat Surg. 2003;10(4):265-81.

135. Shimonishi T, Sasaki M, Nakanuma Y. Precancerous lesions of intrahepatic cholangiocarcinoma. J Hepatobiliary Pancreat Surg. 2000;7(6):542-50.

136. Sibulesky L, Nguyen J, Patel T. Preneoplastic conditions underlying bile duct cancer. Langenbecks Arch Surg. 2012 Mar 6.

137. Chen TC, Nakanuma Y, Zen Y, et al. Intraductal papillary neoplasia of the liver associated with hepatolithiasis. Hepatology. 2001 Oct;34(4 Pt 1):651-8.

138. Nakanuma Y, Sasaki M, Ishikawa A, et al. Biliary papillary neoplasm of the liver. Histology and histopathology. 2002;17(3):851-61.

139. Zen Y, Sasaki M, Fujii T, et al. Different expression patterns of mucin core proteins and cytokeratins during intrahepatic cholangiocarcinogenesis from biliary intraepithelial neoplasia

and intraductal papillary neoplasm of the bile duct--an immunohistochemical study of 110 cases of hepatolithiasis. J Hepatol. 2006 Feb;44(2):350-8.

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

141. Silva MA, Hegab B, Hyde C, et al. Needle track seeding following biopsy of liver lesions in the diagnosis of hepatocellular cancer: a systematic review and meta-analysis. Gut. 2008 Nov;57(11):1592-6.

142. Kurzawinski TR, Deery A, Dooley JS, et al. A prospective study of biliary cytology in 100 patients with bile duct strictures. Hepatology. 1993 Dec;18(6):1399-403.

143. Vandervoort J, Soetikno RM, Montes H, et al. Accuracy and complication rate of brush cytology from bile duct versus pancreatic duct. Gastrointest Endosc. 1999 Mar;49(3 Pt 1):322-7.

144. Weber A, von Weyhern C, Fend F, et al. Endoscopic transpapillary brush cytology and forceps biopsy in patients with hilar cholangiocarcinoma. World J Gastroenterol. 2008 Feb 21;14(7):1097-101.

145. McGuire DE, Venu RP, Brown RD, et al. Brush cytology for pancreatic carcinoma: an analysis of factors influencing results. Gastrointest Endosc. 1996 Sep;44(3):300-4.

146. Rumalla A, Baron TH, Leontovich O, et al. Improved diagnostic yield of endoscopic biliary brush cytology by digital image analysis. Mayo Clin Proc. 2001 Jan;76(1):29-33.

147. Logrono R, Kurtycz DF, Molina CP, et al. Analysis of false-negative diagnoses on endoscopic brush cytology of biliary and pancreatic duct strictures: the experience at 2 university hospitals. Arch Pathol Lab Med. 2000 Mar;124(3):387-92.

148. Ponsioen CY, Vrouenraets SM, van Milligen de Wit AW, et al. Value of brush cytology for dominant strictures in primary sclerosing cholangitis. Endoscopy. 1999 May;31(4):305-9.

149. Stewart CJ, Burke GM. Value of p53 immunostaining in pancreatico-biliary brush cytology specimens. Diagn Cytopathol. 2000 Nov;23(5):308-13.

150. Saurin JC, Joly-Pharaboz MO, Pernas P, et al. Detection of Ki-ras gene point mutations in bile specimens for the differential diagnosis of malignant and benign biliary strictures. Gut. 2000 Sep;47(3):357-61.

151. Baron TH, Harewood GC, Rumalla A, et al. A prospective comparison of digital image analysis and routine cytology for the identification of malignancy in biliary tract strictures. Clin Gastroenterol Hepatol. 2004 Mar;2(3):214-9.

152. Kipp BR, Stadheim LM, Halling SA, et al. A comparison of routine cytology and fluorescence in situ hybridization for the detection of malignant bile duct strictures. Am J Gastroenterol. 2004 Sep;99(9):1675-81.

153. Halling KC, Kipp BR. Fluorescence in situ hybridization in diagnostic cytology. Hum Pathol. 2007 Aug;38(8):1137-44.

154. Glasbrenner B, Ardan M, Boeck W, et al. Prospective evaluation of brush cytology of biliary strictures during endoscopic retrograde cholangiopancreatography. Endoscopy. 1999 Nov;31(9):712-7.

155. Ghazale A, Chari ST, Zhang L, et al. Immunoglobulin G4-associated cholangitis: clinical profile and response to therapy. Gastroenterology. 2008 Mar;134(3):706-15.

156. Khan S. Biliary Tract and Gallbladder Cancer: Diagnosis & Therapy. Thomas C, editor: Demos Medical Publishing; 2008.

157. Anderson C. Treatment of cholangiocarcinoma: UpToDate; 2009.

158. Nakeeb A, Pitt HA, Sohn TA, et al. Cholangiocarcinoma. A spectrum of intrahepatic, perihilar, and distal tumors. Ann Surg. 1996 Oct;224(4):463-73; discussion 73-5.

159. Nehls O, Gregor M, Klump B. Serum and bile markers for cholangiocarcinoma. Semin Liver Dis. 2004 May;24(2):139-54.

160. Kim HJ, Kim MH, Myung SJ, et al. A new strategy for the application of CA19-9 in the differentiation of pancreaticobiliary cancer: analysis using a receiver operating characteristic curve. Am J Gastroenterol. 1999 Jul;94(7):1941-6.

161. Reddy SB, Patel T. Current approaches to the diagnosis and treatment of cholangiocarcinoma. Curr Gastroenterol Rep. 2006 Feb;8(1):30-7.

162. Levy C, Lymp J, Angulo P, et al. The value of serum CA 19-9 in predicting cholangiocarcinomas in patients with primary sclerosing cholangitis. Dig Dis Sci. 2005 Sep;50(9):1734-40.

163. Alvaro D. Serum and bile biomarkers for cholangiocarcinoma. Curr Opin Gastroenterol. 2009 May;25(3):279-84.

164. Oseini AM, Chaiteerakij R, Shire AM, et al. Utility of serum immunoglobulin G4 in distinguishing immunoglobulin G4-associated cholangitis from cholangiocarcinoma. Hepatology. 2011 Jun 14.

165. Gores GJ. Early detection and treatment of cholangiocarcinoma. Liver Transpl. 2000 Nov;6(6 Suppl 2):S30-4.

166. Helzberg JH, Petersen JM, Boyer JL. Improved survival with primary sclerosing cholangitis. A review of clinicopathologic features and comparison of symptomatic and asymptomatic patients. Gastroenterology. 1987 Jun;92(6):1869-75.

167. Van Laethem JL, Deviere J, Bourgeois N, et al. Cholangiographic findings in deteriorating primary sclerosing cholangitis. Endoscopy. 1995 Mar;27(3):223-8.

168. Yachimski P, Pratt DS. Cholangiocarcinoma: natural history, treatment, and strategies for surveillance in high-risk patients. J Clin Gastroenterol. 2008 Feb;42(2):178-90.

169. Bjornsson E, Lindqvist-Ottosson J, Asztely M, et al. Dominant strictures in patients with primary sclerosing cholangitis. Am J Gastroenterol. 2004 Mar;99(3):502-8.

170. Charatcharoenwitthaya P, Enders FB, Halling KC, et al. Utility of serum tumor markers, imaging, and biliary cytology for detecting cholangiocarcinoma in primary sclerosing cholangitis. Hepatology. 2008 Oct;48(4):1106-17.

171. Foley WD, Quiroz FA. The role of sonography in imaging of the biliary tract. Ultrasound Q. 2007 Jun;23(2):123-35.

172. Robledo R, Muro A, Prieto ML. Extrahepatic bile duct carcinoma: US characteristics and accuracy in demonstration of tumors. Radiology. 1996 Mar;198(3):869-73.

173. Robledo R, Prieto ML, Perez M, et al. Carcinoma of the hepaticopancreatic ampullar region: role of US. Radiology. 1988 Feb;166(2):409-12.

174. Neumaier CE, Bertolotto M, Perrone R, et al. Staging of hilar cholangiocarcinoma with ultrasound. J Clin Ultrasound. 1995 Mar-Apr;23(3):173-8.

175. Hann LE, Greatrex KV, Bach AM, et al. Cholangiocarcinoma at the hepatic hilus: sonographic findings. AJR Am J Roentgenol. 1997 Apr;168(4):985-9.

176. Cosgrove DO, Eckersley R. Contrast-enhanced ultrasound: Basic Physics and technology overview. In: Lencioni R, editor. Enhancing the role of ultrasound contrast agents. Pisa: Springer; 2006. p. 3-14.

177. Piscaglia F, Bolondi L. The safety of Sonovue in abdominal applications: retrospective analysis of 23188 investigations. Ultrasound Med Biol. 2006 Sep;32(9):1369-75.

178. Chung YE, Kim MJ, Park YN, et al. Varying appearances of cholangiocarcinoma: radiologic-pathologic correlation. Radiographics. 2009 May-Jun;29(3):683-700.

179. Slattery JM, Sahani DV. What is the current state-of-the-art imaging for detection and staging of cholangiocarcinoma? Oncologist. 2006 Sep;11(8):913-22.

180. Tillich M, Mischinger HJ, Preisegger KH, et al. Multiphasic helical CT in diagnosis and staging of hilar cholangiocarcinoma. AJR Am J Roentgenol. 1998 Sep;171(3):651-8.

181. Aloia TA, Charnsangavej C, Faria S, et al. High-resolution computed tomography accurately predicts resectability in hilar cholangiocarcinoma. Am J Surg. 2007 Jun;193(6):702-6.

182. Petrowsky H, Wildbrett P, Husarik DB, et al. Impact of integrated positron emission tomography and computed tomography on staging and management of gallbladder cancer and cholangiocarcinoma. J Hepatol. 2006 Jul;45(1):43-50.

183. Anderson CD, Rice MH, Pinson CW, et al. Fluorodeoxyglucose PET imaging in the evaluation of gallbladder carcinoma and cholangiocarcinoma. J Gastrointest Surg. 2004 Jan;8(1):90-7.

184. Corvera CU, Blumgart LH, Akhurst T, et al. 18F-fluorodeoxyglucose positron emission tomography influences management decisions in patients with biliary cancer. J Am Coll Surg. 2008 Jan;206(1):57-65.

185. Fritscher-Ravens A, Bohuslavizki KH, Broering DC, et al. FDG PET in the diagnosis of hilar cholangiocarcinoma. Nucl Med Commun. 2001 Dec;22(12):1277-85.

186. Guthrie JA, Ward J, Robinson PJ. Hilar cholangiocarcinomas: T2-weighted spin-echo and gadolinium-enhanced FLASH MR imaging. Radiology. 1996 Nov;201(2):347-51.

187. Ros PR, Buck JL, Goodman ZD, et al. Intrahepatic cholangiocarcinoma: radiologic-pathologic correlation. Radiology. 1988 Jun;167(3):689-93.

188. Lopera JE, Soto JA, Munera F. Malignant hilar and perihilar biliary obstruction: use of MR cholangiography to define the extent of biliary ductal involvement and plan percutaneous interventions. Radiology. 2001 Jul;220(1):90-6.

189. Ba-Ssalamah A, Uffmann M, Saini S, et al. Clinical value of MRI liver-specific contrast agents: a tailored examination for a confident non-invasive diagnosis of focal liver lesions. Eur Radiol. 2009 Feb;19(2):342-57.

190. Romagnuolo J, Bardou M, Rahme E, et al. Magnetic resonance cholangiopancreatography: a meta-analysis of test performance in suspected biliary disease. Ann Intern Med. 2003 Oct 7;139(7):547-57.

191. Masselli G, Gualdi G. Hilar cholangiocarcinoma: MRI/MRCP in staging and treatment planning. Abdom Imaging. 2008 Jul-Aug;33(4):444-51.

192. Kim SY, Lee SS, Byun JH, et al. Malignant hepatic tumors: short-term reproducibility of apparent diffusion coefficients with breath-hold and respiratory-triggered diffusion-weighted MR imaging. Radiology. 2010 Jun;255(3):815-23.

193. Colagrande S, Belli G, Politi LS, et al. The influence of diffusion- and relaxation-related factors on signal intensity: an introductive guide to magnetic resonance diffusion-weighted imaging studies. J Comput Assist Tomogr. 2008 May-Jun;32(3):463-74.

194. Cui XY, Chen HW, Cai S, et al. Diffusion-weighted MR imaging for detection of extrahepatic cholangiocarcinoma. Eur J Radiol. 2012 Jan 27.

195. Escalante-Glorsky S. Endoscopic methods for the diagnosis of pancreatobiliary neoplasms: UpToDate; 2009.

196. Andriulli A, Loperfido S, Napolitano G, et al. Incidence rates of post-ERCP complications: a systematic survey of prospective studies. Am J Gastroenterol. 2007 Aug;102(8):1781-8.

197. Nimura Y, Kamiya J, Hayakawa N, et al. Cholangioscopic differentiation of biliary strictures and polyps. Endoscopy. 1989 Dec;21 Suppl 1:351-6.

198. Seo DW, Lee SK, Yoo KS, et al. Cholangioscopic findings in bile duct tumors. Gastrointest Endosc. 2000 Nov;52(5):630-4.

199. Tischendorf JJ, Kruger M, Trautwein C, et al. Cholangioscopic characterization of dominant bile duct stenoses in patients with primary sclerosing cholangitis. Endoscopy. 2006 Jul;38(7):665-9.

200. Siddique I, Galati J, Ankoma-Sey V, et al. The role of choledochoscopy in the diagnosis and management of biliary tract diseases. Gastrointest Endosc. 1999 Jul;50(1):67-73.

201. Awadallah NS, Chen YK, Piraka C, et al. Is there a role for cholangioscopy in patients with primary sclerosing cholangitis? Am J Gastroenterol. 2006 Feb;101(2):284-91.

202. Fukuda Y, Tsuyuguchi T, Sakai Y, et al. Diagnostic utility of peroral cholangioscopy for various bile-duct lesions. Gastrointest Endosc. 2005 Sep;62(3):374-82.

203. Riemann J. Cholangiokopie & Pankreatikoscopie. In: Fruemorgen P, editor. Gastroenterologische Endoskopie. Berlin, Heidelberg, New York: Springer; 1999. p. 151.

204. Meining A, Frimberger E, Becker V, et al. Detection of cholangiocarcinoma in vivo using miniprobe-based confocal fluorescence microscopy. Clin Gastroenterol Hepatol. 2008 Sep;6(9):1057-60.

205. Itoi T, Sofuni A, Itokawa F, et al. Peroral cholangioscopic diagnosis of biliary-tract diseases by using narrow-band imaging (with videos). Gastrointest Endosc. 2007 Oct;66(4):730-6.

206. Gleeson FC, Rajan E, Levy MJ, et al. EUS-guided FNA of regional lymph nodes in patients with unresectable hilar cholangiocarcinoma. Gastrointest Endosc. 2008 Mar;67(3):438-43.

207. Malhi H, Gores GJ. Review article: the modern diagnosis and therapy of cholangiocarcinoma. Aliment Pharmacol Ther. 2006 May 1;23(9):1287-96.

208. Tamada K, Ueno N, Tomiyama T, et al. Characterization of biliary strictures using intraductal ultrasonography: comparison with percutaneous cholangioscopic biopsy. Gastrointest Endosc. 1998 May;47(5):341-9.

209. Vazquez-Sequeiros E, Baron TH, Clain JE, et al. Evaluation of indeterminate bile duct strictures by intraductal US. Gastrointest Endosc. 2002 Sep;56(3):372-9.

210. Farrell RJ, Agarwal B, Brandwein SL, et al. Intraductal US is a useful adjunct to ERCP for distinguishing malignant from benign biliary strictures. Gastrointest Endosc. 2002 Nov;56(5):681-7.

211. Hyodo T, Hyodo N, Yamanaka T, et al. Contrast-enhanced intraductal ultrasonography for thickened bile duct wall. J Gastroenterol. 2001 Aug;36(8):557-9.

212. Tamada K, Ido K, Ueno N, et al. Preoperative staging of extrahepatic bile duct cancer with intraductal ultrasonography. Am J Gastroenterol. 1995 Feb;90(2):239-46.

213. Inui K, Nakazawa S, Yoshino J, et al. Ultrasound probes for biliary lesions. Endoscopy. 1998 Aug;30 Suppl 1:A120-3.

214. Wilson JM, Jungner YG. [Principles and practice of mass screening for disease]. Bol Oficina Sanit Panam. 1968 Oct;65(4):281-393.

215. Nichols JC, Gores GJ, LaRusso NF, et al. Diagnostic role of serum CA 19-9 for cholangiocarcinoma in patients with primary sclerosing cholangitis. Mayo Clin Proc. 1993 Sep;68(9):874-9.

216. Boberg KM, Jebsen P, Clausen OP, et al. Diagnostic benefit of biliary brush cytology in cholangiocarcinoma in primary sclerosing cholangitis. J Hepatol. 2006 Oct;45(4):568-74.

217. Jongsuksuntigul P, Imsomboon T. Opisthorchiasis control in Thailand. Acta Trop. 2003 Nov;88(3):229-32.

218. Sripa B, Bethony JM, Sithithaworn P, et al. Opisthorchiasis and Opisthorchis-associated cholangiocarcinoma in Thailand and Laos. Acta Trop. 2011 Sep;120 Suppl 1:S158-68.

219. Sripa B, Pairojkul C. Cholangiocarcinoma: lessons from Thailand. Curr Opin Gastroenterol. 2008 May;24(3):349-56.

220. Tung B. Clinical manifestations and diagnosis of primary sclerosing cholangitis: UpToDate; 2009.

221. Su CH, Tsay SH, Wu CC, et al. Factors influencing postoperative morbidity, mortality, and survival after resection for hilar cholangiocarcinoma. Ann Surg. 1996 Apr;223(4):384-94.

222. Burke EC, Jarnagin WR, Hochwald SN, et al. Hilar Cholangiocarcinoma: patterns of spread, the importance of hepatic resection for curative operation, and a presurgical clinical staging system. Ann Surg. 1998 Sep;228(3):385-94.

223. Pichlmayr R, Weimann A, Klempnauer J, et al. Surgical treatment in proximal bile duct cancer. A single-center experience. Ann Surg. 1996 Nov;224(5):628-38.

224. Klempnauer J, Ridder GJ, von Wasielewski R, et al. Resectional surgery of hilar cholangiocarcinoma: a multivariate analysis of prognostic factors. J Clin Oncol. 1997 Mar;15(3):947-54.

225. Blechacz B, Komuta M, Roskams T, et al. Clinical diagnosis and staging of cholangiocarcinoma. Nat Rev Gastroenterol Hepatol. 2011 Aug 2.

226. Okabayashi T, Yamamoto J, Kosuge T, et al. A new staging system for mass-forming intrahepatic cholangiocarcinoma: analysis of preoperative and postoperative variables. Cancer. 2001 Nov 1;92(9):2374-83.

227. Yamasaki S. Intrahepatic cholangiocarcinoma: macroscopic type and stage classification. Journal of Hepato-Biliary-Pancreatic Surgery. 2003;10(4):288-91.

228. Jarnagin WR, Fong Y, DeMatteo RP, et al. Staging, resectability, and outcome in 225 patients with hilar cholangiocarcinoma. Ann Surg. 2001 Oct;234(4):507-17; discussion 17-9.

229. Blechacz BR, Sanchez W, Gores GJ. A conceptual proposal for staging ductal cholangiocarcinoma. Curr Opin Gastroenterol. 2009 May;25(3):238-9.

230. Deoliveira ML, Schulick RD, Nimura Y, et al. New staging system and a registry for perihilar cholangiocarcinoma. Hepatology. 2011 Apr;53(4):1363-71.

231. Jang JY, Kim SW, Park DJ, et al. Actual long-term outcome of extrahepatic bile duct cancer after surgical resection. Ann Surg. 2005 Jan;241(1):77-84.

232. Zervos EE, Osborne D, Goldin SB, et al. Stage does not predict survival after resection of hilar cholangiocarcinomas promoting an aggressive operative approach. Am J Surg. 2005 Nov;190(5):810-5.

233. Hong SM, Pawlik TM, Cho H, et al. Depth of tumor invasion better predicts prognosis than the current American Joint Committee on Cancer T classification for distal bile duct carcinoma. Surgery. 2009 Aug;146(2):250-7.

234. Valverde A, Bonhomme N, Farges O, et al. Resection of intrahepatic cholangiocarcinoma: a Western experience. J Hepatobiliary Pancreat Surg. 1999;6(2):122-7.

235. Suzuki H, Isaji S, Pairojkul C, et al. Comparative clinicopathological study of resected intrahepatic cholangiocarcinoma in northeast Thailand and Japan. J Hepatobiliary Pancreat Surg. 2000;7(2):206-11.

236. Isaji S, Kawarada Y, Taoka H, et al. Clinicopathological features and outcome of hepatic resection for intrahepatic cholangiocarcinoma in Japan. J Hepatobiliary Pancreat Surg. 1999;6(2):108-16.

237. Ohtsuka M, Ito H, Kimura F, et al. Results of surgical treatment for intrahepatic cholangiocarcinoma and clinicopathological factors influencing survival. Br J Surg. 2002 Dec;89(12):1525-31.

238. Madariaga JR, Iwatsuki S, Todo S, et al. Liver resection for hilar and peripheral cholangiocarcinomas: a study of 62 cases. Ann Surg. 1998 Jan;227(1):70-9.

239. Lieser MJ, Barry MK, Rowland C, et al. Surgical management of intrahepatic cholangiocarcinoma: a 31-year experience. J Hepatobiliary Pancreat Surg. 1998;5(1):41-7.

240. Casavilla FA, Marsh JW, Iwatsuki S, et al. Hepatic resection and transplantation for peripheral cholangiocarcinoma. J Am Coll Surg. 1997 Nov;185(5):429-36.

241. Tsao JI, Nimura Y, Kamiya J, et al. Management of hilar cholangiocarcinoma: comparison of an American and a Japanese experience. Ann Surg. 2000 Aug;232(2):166-74.

242. Washburn WK, Lewis WD, Jenkins RL. Aggressive surgical resection for cholangiocarcinoma. Arch Surg. 1995 Mar;130(3):270-6.

243. Nagino M, Nimura Y, Kamiya J, et al. Segmental liver resections for hilar cholangiocarcinoma. Hepatogastroenterology. 1998 Jan-Feb;45(19):7-13.

244. Bismuth H, Nakache R, Diamond T. Management strategies in resection for hilar cholangiocarcinoma. Ann Surg. 1992 Jan;215(1):31-8.

245. Nagorney DM, Kendrick ML. Hepatic resection in the treatment of hilar cholangiocarcinoma. Adv Surg. 2006;40:159-71.

246. DeOliveira ML, Cunningham SC, Cameron JL, et al. Cholangiocarcinoma: thirty-one-year experience with 564 patients at a single institution. Ann Surg. 2007 May;245(5):755-62.

247. Neuhaus P, Jonas S, Bechstein WO, et al. Extended resections for hilar cholangiocarcinoma. Ann Surg. 1999 Dec;230(6):808-18; discussion 19.

248. Nathan H, Pawlik TM, Wolfgang CL, et al. Trends in survival after surgery for cholangiocarcinoma: a 30-year population-based SEER database analysis. J Gastrointest Surg. 2007 Nov;11(11):1488-96; discussion 96-7.

249. Tocchi A, Mazzoni G, Liotta G, et al. Late development of bile duct cancer in patients who had biliary-enteric drainage for benign disease: a follow-up study of more than 1,000 patients. Ann Surg. 2001 Aug;234(2):210-4.

250. Bettschart V, Clayton RA, Parks RW, et al. Cholangiocarcinoma arising after biliary-enteric drainage procedures for benign disease. Gut. 2002 Jul;51(1):128-9.

251. Kaya M, de Groen PC, Angulo P, et al. Treatment of cholangiocarcinoma complicating primary sclerosing cholangitis: the Mayo Clinic experience. Am J Gastroenterol. 2001 Apr;96(4):1164-9.

252. Rosen CB, Nagorney DM, Wiesner RH, et al. Cholangiocarcinoma complicating primary sclerosing cholangitis. Ann Surg. 1991 Jan;213(1):21-5.

253. Heimbach JK, Gores GJ, Nagorney DM, et al. Liver transplantation for perihilar cholangiocarcinoma after aggressive neoadjuvant therapy: a new paradigm for liver and biliary malignancies? Surgery. 2006 Sep;140(3):331-4.

254. Heimbach JK, Gores GJ, Haddock MG, et al. Liver transplantation for unresectable perihilar cholangiocarcinoma. Semin Liver Dis. 2004 May;24(2):201-7.

255. Heimbach JK, Gores GJ, Haddock MG, et al. Predictors of disease recurrence following neoadjuvant chemoradiotherapy and liver transplantation for unresectable perihilar cholangiocarcinoma. Transplantation. 2006 Dec 27;82(12):1703-7.

256. Heimbach JK, Haddock MG, Alberts SR, et al. Transplantation for hilar cholangiocarcinoma. Liver Transpl. 2004 Oct;10(10 Suppl 2):S65-8.

257. Rosen CB, Heimbach JK, Gores GJ. Surgery for cholangiocarcinoma: the role of liver transplantation. HPB (Oxford). 2008;10(3):186-9.

258. Rosen CB, Heimbach JK, Gores GJ. Liver transplantation for cholangiocarcinoma. Transpl Int. 2010 Jul;23(7):692-7.

259. Darwish Murad S, Kim WR, Harnois DM, et al. Efficacy of Neoadjuvant Chemoradiation, Followed by Liver Transplantation, for Perihilar Cholangiocarcinoma at 12 US Centers. Gastroenterology. 2012;143(1):88-98.e3.

260. McMasters KM, Tuttle TM, Leach SD, et al. Neoadjuvant chemoradiation for extrahepatic cholangiocarcinoma. Am J Surg. 1997 Dec;174(6):605-8; discussion 8-9.

261. Nelson JW, Ghafoori AP, Willett CG, et al. Concurrent chemoradiotherapy in resected extrahepatic cholangiocarcinoma. Int J Radiat Oncol Biol Phys. 2009 Jan 1;73(1):148-53.

262. Borghero Y, Crane CH, Szklaruk J, et al. Extrahepatic bile duct adenocarcinoma: patients at high-risk for local recurrence treated with surgery and adjuvant chemoradiation have an equivalent overall survival to patients with standard-risk treated with surgery alone. Ann Surg Oncol. 2008 Nov;15(11):3147-56.

263. Kim S, Kim SW, Bang YJ, et al. Role of postoperative radiotherapy in the management of extrahepatic bile duct cancer. Int J Radiat Oncol Biol Phys. 2002 Oct 1;54(2):414-9.

264. Klinkenbijl JH, Jeekel J, Sahmoud T, et al. Adjuvant radiotherapy and 5-fluorouracil after curative resection of cancer of the pancreas and periampullary region: phase III trial of the EORTC gastrointestinal tract cancer cooperative group. Ann Surg. 1999 Dec;230(6):776-82; discussion 82-4.

265. Gerhards MF, van Gulik TM, Gonzalez Gonzalez D, et al. Results of postoperative radiotherapy for resectable hilar cholangiocarcinoma. World J Surg. 2003 Feb;27(2):173-9.

266. Todoroki T, Ohara K, Kawamoto T, et al. Benefits of adjuvant radiotherapy after radical resection of locally advanced main hepatic duct carcinoma. Int J Radiat Oncol Biol Phys. 2000 Feb 1;46(3):581-7.

267. Glimelius B, Hoffman K, Sjoden PO, et al. Chemotherapy improves survival and quality of life in advanced pancreatic and biliary cancer. Ann Oncol. 1996 Aug;7(6):593-600.

268. Lai EC, Chu KM, Lo CY, et al. Choice of palliation for malignant hilar biliary obstruction. Am J Surg. 1992 Feb;163(2):208-12.

269. Stern N, Sturgess R. Endoscopic therapy in the management of malignant biliary obstruction. Eur J Surg Oncol. 2008 Mar;34(3):313-7.

270. Raju RP, Jaganmohan SR, Ross WA, et al. Optimum palliation of inoperable hilar cholangiocarcinoma: comparative assessment of the efficacy of plastic and self-expanding metal stents. Dig Dis Sci. 2011 May;56(5):1557-64.

271. Seehofer D, Kamphues C, Neuhaus P. Management of bile duct tumors. Expert Opin Pharmacother. 2008 Nov;9(16):2843-56.

272. Steel AW, Postgate AJ, Khorsandi S, et al. Endoscopically applied radiofrequency ablation appears to be safe in the treatment of malignant biliary obstruction. Gastrointest Endosc. 2011 Jan;73(1):149-53.

273. Zoepf T. Photodynamic therapy of cholangiocarcinoma. HPB (Oxford). 2008;10(3):161-3.

274. Zoepf T, Jakobs R, Arnold JC, et al. Palliation of nonresectable bile duct cancer: improved survival after photodynamic therapy. Am J Gastroenterol. 2005 Nov;100(11):2426-30.

275. Ortner ME, Caca K, Berr F, et al. Successful photodynamic therapy for nonresectable cholangiocarcinoma: a randomized prospective study. Gastroenterology. 2003 Nov;125(5):1355-63.

276. Pereira S, Hughes S, Roughton M, et al. Photostent-02; Porfimer Sodium photodynamic therapy plus stenting versus stenting alone in patients with advanced or metastatic cholangiocarcinomas and other biliary tract tumours: A multicentre, randomised phase III trial. European Society of Medical Oncology; Milan, Italy2010.

277. Westaby D. The use of endoscopic RFA in obstructed metal biliary stents, personal communication of unpublished data. In: Wadsworth C, editor. Letter ed. London2011.

278. Kallis Y. Case-control survival analysis after intraductal RFA for malignant disease, personal communication of unpublished data. [Case-control study]. In press 2012.

279. Ibrahim SM, Mulcahy MF, Lewandowski RJ, et al. Treatment of unresectable cholangiocarcinoma using yttrium-90 microspheres: results from a pilot study. Cancer. 2008 Oct 15;113(8):2119-28.

280. Smith AC, Dowsett JF, Russell RC, et al. Randomised trial of endoscopic stenting versus surgical bypass in malignant low bileduct obstruction. Lancet. 1994 Dec 17;344(8938):1655-60.

281. Whittington R, Neuberg D, Tester WJ, et al. Protracted intravenous fluorouracil infusion with radiation therapy in the management of localized pancreaticobiliary carcinoma: a phase I Eastern Cooperative Oncology Group Trial. J Clin Oncol. 1995 Jan;13(1):227-32.

282. Park JY, Park SW, Chung JB, et al. Concurrent chemoradiotherapy with doxifluridine and paclitaxel for extrahepatic bile duct cancer. Am J Clin Oncol. 2006 Jun;29(3):240-5.

283. Ohnishi H, Asada M, Shichijo Y, et al. External radiotherapy for biliary decompression of hilar cholangiocarcinoma. Hepatogastroenterology. 1995 Jul;42(3):265-8.

284. Alden ME, Mohiuddin M. The impact of radiation dose in combined external beam and intraluminal Ir-192 brachytherapy for bile duct cancer. Int J Radiat Oncol Biol Phys. 1994 Mar 1;28(4):945-51.

285. Van Beers BE. Diagnosis of cholangiocarcinoma. HPB (Oxford). 2008;10(2):87-93.

286. Eckel F, Brunner T, Jelic S. Biliary cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2011 Sep;22 Suppl 6:vi40-4.

287. Witzigmann H, Lang H, Lauer H. Guidelines for palliative surgery of cholangiocarcinoma. HPB (Oxford). 2008;10(3):154-60.

288. Huang WY, Gao YT, Rashid A, et al. Selected base excision repair gene polymorphisms and susceptibility to biliary tract cancer and biliary stones: a population-based case-control study in China. Carcinogenesis. 2008 Jan;29(1):100-5.

289. Sakoda LC, Gao YT, Chen BE, et al. Prostaglandin-endoperoxide synthase 2 (PTGS2) gene polymorphisms and risk of biliary tract cancer and gallstones: a population-based study in Shanghai, China. Carcinogenesis. 2006 Jun;27(6):1251-6.

290. Sainani NI, Catalano OA, Holalkere NS, et al. Cholangiocarcinoma: current and novel imaging techniques. Radiographics. 2008 Sep-Oct;28(5):1263-87.

291. Dougherty CP, Holtz SH, Reinert JC, et al. Dietary exposures to food contaminants across the United States. Environ Res. 2000 Oct;84(2):170-85.

292. Kjeller LO, Jones KC, Johnston AE, et al. Increases in the Polychlorinated Dibenzo-P-Dioxin and Dibenzo-P-Furan Content of Soils and Vegetation since the 1840s. Environ Sci Technol. 1991 Sep;25(9):1619-27.

293. Alcock RE, Johnston AE, Mcgrath SP, et al. Long-Term Changes in the Polychlorinated Biphenyl Content of United-Kingdom Soils. Environ Sci Technol. 1993 Sep;27(9):1918-23.

294. Trauner M, Boyer JL. Bile salt transporters: molecular characterization, function, and regulation. Physiol Rev. 2003 Apr;83(2):633-71.

295. Geier A, Fickert P, Trauner M. Mechanisms of disease: mechanisms and clinical implications of cholestasis in sepsis. Nat Clin Pract Gastroenterol Hepatol. 2006 Oct;3(10):574-85.

296. Wadsworth CA, Dixon PH, Wong JH, et al. Genetic factors in the pathogenesis of cholangiocarcinoma. Dig Dis. 2011;29(1):93-7.

297. Pauli-Magnus C, Stieger B, Meier Y, et al. Enterohepatic transport of bile salts and genetics of cholestasis. J Hepatol. 2005 Aug;43(2):342-57.

298. Strautnieks SS, Bull LN, Knisely AS, et al. A gene encoding a liver-specific ABC transporter is mutated in progressive familial intrahepatic cholestasis. Nat Genet. 1998 Nov;20(3):233-8.

299. Trauner M, Fickert P, Halilbasic E, et al. Lessons from the toxic bile concept for the pathogenesis and treatment of cholestatic liver diseases. Wien Med Wochenschr. 2008;158(19-20):542-8.

300. Ananthanarayanan M, Li Y. PFIC2 and ethnicity-specific bile salt export pump (BSEP, ABCB11) mutations: where do we go from here? Liver Int. 2010 Jul;30(6):777-9.

301. Gonzales E, Davit-Spraul A, Baussan C, et al. Liver diseases related to MDR3 (ABCB4) gene deficiency. Front Biosci. 2009;14:4242-56.

302. Delaunay JL, Durand-Schneider AM, Delautier D, et al. A missense mutation in ABCB4 gene involved in progressive familial intrahepatic cholestasis type 3 leads to a folding defect that can be rescued by low temperature. Hepatology. 2009 Apr;49(4):1218-27.

303. Nies AT, Schwab M, Keppler D. Interplay of conjugating enzymes with OATP uptake transporters and ABCC/MRP efflux pumps in the elimination of drugs. Expert Opin Drug Metab Toxicol. 2008 May;4(5):545-68.

304. van Mil SW, Klomp LW, Bull LN, et al. FIC1 disease: a spectrum of intrahepatic cholestatic disorders. Semin Liver Dis. 2001 Nov;21(4):535-44.

305. Cai SY, Gautam S, Nguyen T, et al. ATP8B1 deficiency disrupts the bile canalicular membrane bilayer structure in hepatocytes, but FXR expression and activity are maintained. Gastroenterology. 2009 Mar;136(3):1060-9.

306. Davit-Spraul A, Fabre M, Branchereau S, et al. ATP8B1 and ABCB11 analysis in 62 children with normal gamma-glutamyl transferase progressive familial intrahepatic cholestasis

(PFIC): phenotypic differences between PFIC1 and PFIC2 and natural history. Hepatology. 2010 May;51(5):1645-55.

307. Davit-Spraul A, Gonzales E, Baussan C, et al. Progressive familial intrahepatic cholestasis. Orphanet J Rare Dis. 2009;4:1.

308. Zhang Y, Edwards PA. FXR signaling in metabolic disease. FEBS Lett. 2008 Jan 9;582(1):10-8.

309. Ananthanarayanan M, Balasubramanian N, Makishima M, et al. Human bile salt export pump promoter is transactivated by the farnesoid X receptor/bile acid receptor. J Biol Chem. 2001 Aug 3;276(31):28857-65.

310. Huang L, Zhao A, Lew JL, et al. Farnesoid X receptor activates transcription of the phospholipid pump MDR3. J Biol Chem. 2003 Dec 19;278(51):51085-90.

311. Pezzullo J. Proportion Difference Power / Sample Size Calculation. 2009 [updated 10/30/2009; cited 2010 27th October]; Available from: http://statpages.org/proppowr.html.

312. Altman D. Practical Statistics for Medical Research. London: Chapman and Hall; 1991.

313. Machin D. Sample Size Tables for CLinical Studies. Oxford: Blackwell Science Limited; 1987.

314. van Belle G. Sample Size. Statistical Rules of Thumb. 2nd ed. Hoboken, NJ, USA: Wiley; 2008. p. 27-51.

315. Qiagen. QIAamp®DNA Mini and Blood Mini Handbook. 2010.

316. Qiagen. Purification of archive-quality DNA from clotted whole blood using Clotspin® Baskets and the Gentra® Puregene® Blood Kit. Qiagen; 2010.

317. Asquith S. Comparison of Taqman® and KASPar Fluorescent Assays for SNP genotyping Hoddesdon, UK: KBioscience Ltd; 2009; Available from: http://www.kbioscience.co.uk/chemistry/KASP_Taqmancomparison.pdf.

318. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics. 2007 Sep;81(3):559-75.

319. Schaid DJ, Rowland CM, Tines DE, et al. Score tests for association between traits and haplotypes when linkage phase is ambiguous. American Journal of Human Genetics. 2002 Feb;70(2):425-34.

320. Sinnwell J, Schaid D. Haplo.stats. 1.4.4 ed: The Comprehensive R Archive Network (CRAN); 2009.

321. Storey JD, Taylor JE, Siegmund D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. J Roy Stat Soc B. 2004;66:187-205.

322. NCBI. Database of Single Nucleotide Polymorphisms (dbSNP). 132 ed. Bethesda, MD: National Center for Biotechnology Information, National Library of Medicine; 2011.

323. Lang C, Meier Y, Stieger B, et al. Mutations and polymorphisms in the bile salt export pump and the multidrug resistance protein 3 associated with drug-induced liver injury. Pharmacogenet Genomics. 2007 Jan;17(1):47-60.

324. Dixon PH, van Mil SW, Chambers J, et al. Contribution of variant alleles of ABCB11 to susceptibility to intrahepatic cholestasis of pregnancy. Gut. 2009 Apr;58(4):537-44.

325. Strautnieks SS, Byrne JA, Pawlikowska L, et al. Severe bile salt export pump deficiency: 82 different ABCB11 mutations in 109 families. Gastroenterology. 2008 Apr;134(4):1203-14.

326. Hofmann AF. Bile acids as drugs: principles, mechanisms of action and formulations. Ital J Gastroenterol. 1995 Mar;27(2):106-13.

327. Hofmann AF. The continuing importance of bile acids in liver and intestinal disease. Arch Intern Med. 1999 Dec 13-27;159(22):2647-58.

328. Trauner M, Fickert P, Wagner M. MDR3 (ABCB4) defects: a paradigm for the genetics of adult cholestatic syndromes. Semin Liver Dis. 2007 Feb;27(1):77-98.

329. Lazaridis KN, Strazzabosco M, Larusso NF. The cholangiopathies: disorders of biliary epithelia. Gastroenterology. 2004 Nov;127(5):1565-77.

330. Paulusma CC, de Waart DR, Kunne C, et al. Activity of the bile salt export pump (ABCB11) is critically dependent on canalicular membrane cholesterol content. J Biol Chem. 2009 Apr 10;284(15):9947-54.

331. Verhulst PM, van der Velden LM, Oorschot V, et al. A flippase-independent function of ATP8B1, the protein affected in familial intrahepatic cholestasis type 1, is required for apical protein expression and microvillus formation in polarized epithelial cells. Hepatology. 2010 Jun;51(6):2049-60.

332. Vivier E, Romagne F. Good news, bad news for missing-self recognition by NK cells: autoimmune control but viral evasion. Immunity. 2007 May;26(5):549-51.

333. Moretta L, Biassoni R, Bottino C, et al. Human NK cells and their receptors. Microbes Infect. 2002 Dec;4(15):1539-44.

334. Moretta A, Bottino C, Mingari MC, et al. What is a natural killer cell? Nat Immunol. 2002 Jan;3(1):6-8.

335. Moretta L, Bottino C, Pende D, et al. Human natural killer cells: their origin, receptors and function. Eur J Immunol. 2002 May;32(5):1205-11.

336. Pende D, Cantoni C, Rivera P, et al. Role of NKG2D in tumor cell lysis mediated by human NK cells: cooperation with natural cytotoxicity receptors and capability of recognizing tumors of nonepithelial origin. Eur J Immunol. 2001 Apr;31(4):1076-86.

337. Eagle RA, Trowsdale J. Promiscuity and the single receptor: NKG2D. Nat Rev Immunol. 2007 Sep;7(9):737-44.

338. Gasser S, Orsulic S, Brown EJ, et al. The DNA damage pathway regulates innate immune system ligands of the NKG2D receptor. Nature. 2005 Aug 25;436(7054):1186-90.

339. Watzl C. The NKG2D receptor and its ligands-recognition beyond the "missing self"? Microbes Infect. 2003 Jan;5(1):31-7.

340. Benitez AC, Dai Z, Mann HH, et al. Expression, signaling proficiency, and stimulatory function of the NKG2D lymphocyte receptor in human cancer cells. Proc Natl Acad Sci U S A. 2011 Mar 8;108(10):4081-6.

341. Takeda K, Hayakawa Y, Smyth MJ, et al. Involvement of tumor necrosis factor-related apoptosis-inducing ligand in surveillance of tumor metastasis by liver natural killer cells. Nat Med. 2001 Jan;7(1):94-100.

342. Trapani JA, Smyth MJ. Functional significance of the perforin/granzyme cell death pathway. Nat Rev Immunol. 2002 Oct;2(10):735-47.

343. Hayakawa Y, Smyth MJ. NKG2D and cytotoxic effector function in tumor immune surveillance. Semin Immunol. 2006 Jun;18(3):176-85.

344. Bryceson YT, Ljunggren HG. Tumor cell recognition by the NK cell activating receptor NKG2D. Eur J Immunol. 2008 Nov;38(11):2957-61.

345. Ljunggren HG. Cancer immunosurveillance: NKG2D breaks cover. Immunity. 2008 Apr;28(4):492-4.

346. Hayashi T, Imai K, Morishita Y, et al. Identification of the NKG2D haplotypes associated with natural cytotoxic activity of peripheral blood lymphocytes and cancer immunosurveillance. Cancer Res. 2006 Jan 1;66(1):563-70.

347. Fleiss JL. Statistical Methods for Rates and Proportions. In: Fleiss JL, editor. Statistical Methods for Rates and Proportions. 2nd ed. New York: John Wiley & Sons; 1981.

348. Fevery J, Verslype C, Lai G, et al. Incidence, diagnosis, and therapy of cholangiocarcinoma in patients with primary sclerosing cholangitis. Dig Dis Sci. 2007 Nov;52(11):3123-35.

349. Graziadei IW, Wiesner RH, Marotta PJ, et al. Long-term results of patients undergoing liver transplantation for primary sclerosing cholangitis. Hepatology. 1999 Nov;30(5):1121-7.

350. Jesudian AB, Jacobson IM. Screening and diagnosis of cholangiocarcinoma in patients with primary sclerosing cholangitis. Rev Gastroenterol Disord. 2009 Spring;9(2):E41-7.

351. Lazaridis KN, Gores GJ. Primary sclerosing cholangitis and cholangiocarcinoma. Semin Liver Dis. 2006 Feb;26(1):42-51.

352. Guerra N, Tan YX, Joncker NT, et al. NKG2D-deficient mice are defective in tumor surveillance in models of spontaneous malignancy. Immunity. 2008 Apr;28(4):571-80.

353. 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 Jul;37(1):55-63.

354. 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 Nov;15(11):1181-90.

355. Lau SK, Weiss LM, Chu PG. Differential expression of MUC1, MUC2, and MUC5AC in carcinomas of various sites: an immunohistochemical study. Am J Clin Pathol. 2004 Jul;122(1):61-9.

356. Bonney GK, Craven RA, Prasad R, et al. Circulating markers of biliary malignancy: opportunities in proteomics? Lancet Oncol. 2008 Feb;9(2):149-58.

357. Obama K, Ura K, Li M, et al. Genome-wide analysis of gene expression in human intrahepatic cholangiocarcinoma. Hepatology. 2005 Jun;41(6):1339-48.

358. Obama K, Ura K, Satoh S, et al. Up-regulation of PSF2, a member of the GINS multiprotein complex, in intrahepatic cholangiocarcinoma. Oncol Rep. 2005 Sep;14(3):701-6.

359. Srisomsap C, Sawangareetrakul P, Subhasitanont P, et al. Proteomic analysis of cholangiocarcinoma cell line. Proteomics. 2004 Apr;4(4):1135-44.

360. Scarlett CJ, Saxby AJ, Nielsen A, et al. Proteomic profiling of cholangiocarcinoma: diagnostic potential of SELDI-TOF MS in malignant bile duct stricture. Hepatology. 2006 Sep;44(3):658-66.

361. Koopmann J, Thuluvath PJ, Zahurak ML, et al. Mac-2-binding protein is a diagnostic marker for biliary tract carcinoma. Cancer. 2004 Oct 1;101(7):1609-15.

362. Kristiansen TZ, Bunkenborg J, Gronborg M, et al. A proteomic analysis of human bile. Mol Cell Proteomics. 2004 Jul;3(7):715-28.

363. Farina A, Dumonceau JM, Frossard JL, et al. Proteomic analysis of human bile from malignant biliary stenosis induced by pancreatic cancer. J Proteome Res. 2009 Jan;8(1):159-69.

364. 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 Sep;106(9):1711-7.

365. Kashihara T, Ohki A, Kobayashi T, et al. Intrahepatic cholangiocarcinoma with increased serum CYFRA 21-1 level. J Gastroenterol. 1998 Jun;33(3):447-53.

366. Watanabe H, Enjoji M, Nakashima M, et al. Clinical significance of serum RCAS1 levels detected by monoclonal antibody 22-1-1 in patients with cholangiocellular carcinoma. J Hepatol. 2003 Oct;39(4):559-63.

367. Wongkham S, Sheehan JK, Boonla C, et al. Serum MUC5AC mucin as a potential marker for cholangiocarcinoma. Cancer Lett. 2003 May 30;195(1):93-9.

368. Tangkijvanich P, Thong-ngam D, Theamboonlers A, et al. Diagnostic role of serum interleukin 6 and CA 19-9 in patients with cholangiocarcinoma. Hepatogastroenterology. 2004 Jan-Feb;51(55):15-9.

369. Wilkins MR, Sanchez JC, Gooley AA, et al. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol Genet Eng Rev. 1996;13:19-50.

370. Griffiths WJ, Wang Y. Mass spectrometry: from proteomics to metabolomics and lipidomics. Chem Soc Rev. 2009 Jul;38(7):1882-96.

371. Abdul-Salam VB, Paul GA, Ali JO, et al. Identification of plasma protein biomarkers associated with idiopathic pulmonary arterial hypertension. Proteomics. 2006 Apr;6(7):2286-94.

372. Fang M, Boobis AR, Edwards RJ. Searching for novel biomarkers of centrally and peripehrally-acting neurotoxicants, using surface-enhanced laser desorption/ionisation-time-of-flight mass spectrometry (SELDI-TOF MS). Food Chem Toxicol. 2007 Nov;45(11):2126-37.

373. Matthews KW, Mueller-Ortiz SL, Wetsel RA. Carboxypeptidase N: a pleiotropic regulator of inflammation. Mol Immunol. 2004 Jan;40(11):785-93.

374. Mathews KP, Pan PM, Gardner NJ, et al. Familial carboxypeptidase N deficiency. Ann Intern Med. 1980 Sep;93(3):443-5.

375. UniProt Knowledgebase [database on the Internet]. UniProt Consortium. 2012 [cited 26/09/2012].

376. Gerhardt T, Milz S, Schepke M, et al. C-reactive protein is a prognostic indicator in patients with perihilar cholangiocarcinoma. World J Gastroenterol. 2006 Sep 14;12(34):5495-500.

377. Erlinger TP, Platz EA, Rifai N, et al. C-reactive protein and the risk of incident colorectal cancer. JAMA. 2004 Feb 4;291(5):585-90.

378. Pawlowicz Z, Zachara BA, Trafikowska U, et al. Blood selenium concentrations and glutathione peroxidase activities in patients with breast cancer and with advanced gastrointestinal cancer. Journal of trace elements and electrolytes in health and disease. 1991 Dec;5(4):275-7.

379. Wasowicz W, Gromadzinska J, Sklodowska M, et al. Selenium concentration and glutathione peroxidase activity in blood of children with cancer. Journal of trace elements and electrolytes in health and disease. 1994 Mar;8(1):53-7.

380. Agnani D, Camacho-Vanegas O, Camacho C, et al. Decreased levels of serum glutathione peroxidase 3 are associated with papillary serous ovarian cancer and disease progression. Journal of ovarian research. 2011;4:18.

381. Liu J, Hinkhouse MM, Sun W, et al. Redox regulation of pancreatic cancer cell growth: role of glutathione peroxidase in the suppression of the malignant phenotype. Human gene therapy. 2004 Mar;15(3):239-50.

382. Kakisaka T, Kondo T, Okano T, et al. Plasma proteomics of pancreatic cancer patients by multi-dimensional liquid chromatography and two-dimensional difference gel electrophoresis (2D-DIGE): up-regulation of leucine-rich alpha-2-glycoprotein in pancreatic cancer. J Chromatogr B Analyt Technol Biomed Life Sci. 2007 Jun 1;852(1-2):257-67.

383. Sandanayake NS, Sinclair J, Andreola F, et al. A combination of serum leucine-rich alpha-2-glycoprotein 1, CA19-9 and interleukin-6 differentiate biliary tract cancer from benign biliary strictures. Br J Cancer. 2011 Oct 25;105(9):1370-8.

384. Andersen JD, Boylan KL, Jemmerson R, et al. Leucine-rich alpha-2-glycoprotein-1 is upregulated in sera and tumors of ovarian cancer patients. Journal of ovarian research. 2010;3:21.

385. Ferrero S, Gillott DJ, Remorgida V, et al. Increased expression of one isoform of leucine-rich alpha-2-glycoprotein in peritoneal fluid of women with uterine leiomyomas. Archives of gynecology and obstetrics. 2009 Mar;279(3):365-71.

386. Kentsis A, Ahmed S, Kurek K, et al. Detection and diagnostic value of urine leucine-rich alpha-2-glycoprotein in children with suspected acute appendicitis. Ann Emerg Med. 2012 Jul;60(1):78-83 e1.

387. Serada S, Fujimoto M, Ogata A, et al. iTRAQ-based proteomic identification of leucine-rich alpha-2 glycoprotein as a novel inflammatory biomarker in autoimmune diseases. Annals of the rheumatic diseases. 2010 Apr;69(4):770-4.

388. Serada S, Fujimoto M, Terabe F, et al. Serum leucine-rich alpha-2 glycoprotein is a disease activity biomarker in ulcerative colitis. Inflamm Bowel Dis. 2012 Feb 28.

389. Kim BK, Lee JW, Park PJ, et al. The multiplex bead array approach to identifying serum biomarkers associated with breast cancer. Breast cancer research : BCR. 2009;11(2):R22.

390. Yu KH, Barry CG, Austin D, et al. Stable isotope dilution multidimensional liquid chromatography-tandem mass spectrometry for pancreatic cancer serum biomarker discovery. J Proteome Res. 2009 Mar;8(3):1565-76.

391. Yang MH, Tyan YC, Jong SB, et al. Identification of human hepatocellular carcinoma-related proteins by proteomic approaches. Analytical and bioanalytical chemistry. 2007 Jun;388(3):637-43.

392. Bozinovski S, Hutchinson A, Thompson M, et al. Serum amyloid a is a biomarker of acute exacerbations of chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine. 2008 Feb 1;177(3):269-78.

393. Haqqani AS, Hutchison JS, Ward R, et al. Biomarkers and diagnosis; protein biomarkers in serum of pediatric patients with severe traumatic brain injury identified by ICAT-LC-MS/MS. Journal of neurotrauma. 2007 Jan;24(1):54-74.

394. Malle E, Sodin-Semrl S, Kovacevic A. Serum amyloid A: an acute-phase protein involved in tumour pathogenesis. Cell Mol Life Sci. 2009 Jan;66(1):9-26.

395. Vargas I. A urine test for appendicitis? The Harvard News Gazette. 2009 23rd June 2009.

396. Abdul-Salam VB, Ramrakha P, Krishnan U, et al. Identification and assessment of plasma lysozyme as a putative biomarker of atherosclerosis. Arterioscler Thromb Vasc Biol. 2010 May;30(5):1027-33.

397. Child C, Turcotte J. Surgery and portal hypertension. In: Child C, editor. The liver and portal hypertension. Philadelphia: Saunders; 1964. p. 50-64.

398. Pugh RN, Murray-Lyon IM, Dawson JL, et al. Transection of the oesophagus for bleeding oesophageal varices. Br J Surg. 1973 Aug;60(8):646-9.

399. Liu L, Wang J, Liu B, et al. Serum levels of variants of transthyretin down-regulation in cholangiocarcinoma. J Cell Biochem. 2008 Jun 1;104(3):745-55.

400. Navaglia F, Fogar P, Basso D, et al. Pancreatic cancer biomarkers discovery by surface-enhanced laser desorption and ionization time-of-flight mass spectrometry. Clinical chemistry and laboratory medicine : CCLM / FESCC. 2009;47(6):713-23.

401. Zinkin NT, Grall F, Bhaskar K, et al. Serum proteomics and biomarkers in hepatocellular carcinoma and chronic liver disease. Clin Cancer Res. 2008 Jan 15;14(2):470-7.

402. Okabe H, Beppu T, Ueda M, et al. Identification of CXCL5/ENA-78 as a factor involved in the interaction between cholangiocarcinoma cells and cancer-associated fibroblasts. Int J Cancer. 2012 Nov 15;131(10):2234-41.

403. 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 Nov;18(6):625-35.

404. Kikkawa S, Sogawa K, Satoh M, et al. Identification of a Novel Biomarker for Biliary Tract Cancer Using Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. International journal of proteomics. 2012;2012:108609.

405. Qian JY, Mou SH, Liu CB. SELDI-TOF MS Combined with Magnetic Beads for Detecting Serum Protein Biomarkers and Establishment of a Boosting Decision Tree Model for Diagnosis of Pancreatic Cancer. Asian Pac J Cancer Prev. 2012;13(5):1911-5.

406. Issaq HJ, Veenstra TD, Conrads TP, et al. The SELDI-TOF MS approach to proteomics: protein profiling and biomarker identification. Biochem Biophys Res Commun. 2002 Apr 5;292(3):587-92.

407. Chakraborty S, Kaur S, Guha S, et al. The multifaceted roles of neutrophil gelatinase associated lipocalin (NGAL) in inflammation and cancer. Biochim Biophys Acta. 2012 Aug;1826(1):129-69.

408. Cowland JB, Sorensen OE, Sehested M, et al. Neutrophil gelatinase-associated lipocalin is up-regulated in human epithelial cells by IL-1 beta, but not by TNF-alpha. J Immunol. 2003 Dec 15;171(12):6630-9.

409. Cowland JB, Borregaard N. Molecular characterization and pattern of tissue expression of the gene for neutrophil gelatinase-associated lipocalin from humans. Genomics. 1997 Oct 1;45(1):17-23.

410. Goetz DH, Willie ST, Armen RS, et al. Ligand preference inferred from the structure of neutrophil gelatinase associated lipocalin. Biochemistry. 2000 Feb 29;39(8):1935-41.

411. Miethke M, Klotz O, Linne U, et al. Ferri-bacillibactin uptake and hydrolysis in Bacillus subtilis. Mol Microbiol. 2006 Sep;61(6):1413-27.

412. Miethke M, Skerra A. Neutrophil gelatinase-associated lipocalin expresses antimicrobial activity by interfering with L-norepinephrine-mediated bacterial iron acquisition. Antimicrob Agents Chemother. 2010 Apr;54(4):1580-9.

413. Aigner F, Maier HT, Schwelberger HG, et al. Lipocalin-2 regulates the inflammatory response during ischemia and reperfusion of the transplanted heart. Am J Transplant. 2007 Apr;7(4):779-88.

414. Flo TH, Smith KD, Sato S, et al. Lipocalin 2 mediates an innate immune response to bacterial infection by sequestrating iron. Nature. 2004 Dec 16;432(7019):917-21.

415. Bahmani P, Halabian R, Rouhbakhsh M, et al. Neutrophil gelatinase-associated lipocalin induces the expression of heme oxygenase-1 and superoxide dismutase 1, 2. Cell stress & chaperones. 2010 Jul;15(4):395-403.

416. Roudkenar MH, Halabian R, Ghasemipour Z, et al. Neutrophil gelatinase-associated lipocalin acts as a protective factor against H(2)O(2) toxicity. Archives of medical research. 2008 Aug;39(6):560-6.

417. Roudkenar MH, Kuwahara Y, Baba T, et al. Oxidative stress induced lipocalin 2 gene expression: addressing its expression under the harmful conditions. Journal of radiation research. 2007 Jan;48(1):39-44.

418. Roudkenar MH, Halabian R, Oodi A, et al. Upregulation of neutrophil gelatinase-associated lipocalin, NGAL/Lcn2, in beta-thalassemia patients. Archives of medical research. 2008 May;39(4):402-7.

419. te Boekhorst BC, Bovens SM, Hellings WE, et al. Molecular MRI of murine atherosclerotic plaque targeting NGAL: a protein associated with unstable human plaque characteristics. Cardiovascular research. 2011 Feb 15;89(3):680-8.

420. Viau A, El Karoui K, Laouari D, et al. Lipocalin 2 is essential for chronic kidney disease progression in mice and humans. J Clin Invest. 2010 Nov;120(11):4065-76.

421. Miharada K, Hiroyama T, Sudo K, et al. Lipocalin 2 functions as a negative regulator of red blood cell production in an autocrine fashion. FASEB J. 2005 Nov;19(13):1881-3.

422. Ding L, Hanawa H, Ota Y, et al. Lipocalin-2/neutrophil gelatinase-B associated lipocalin is strongly induced in hearts of rats with autoimmune myocarditis and in human myocarditis. Circulation journal : official journal of the Japanese Circulation Society. 2010 Mar;74(3):523-30.

423. Yndestad A, Landro L, Ueland T, et al. Increased systemic and myocardial expression of neutrophil gelatinase-associated lipocalin in clinical and experimental heart failure. European heart journal. 2009 May;30(10):1229-36.

424. Landro L, Damas JK, Flo TH, et al. Decreased serum lipocalin-2 levels in human immunodeficiency virus-infected patients: increase during highly active anti-retroviral therapy. Clinical and experimental immunology. 2008 Apr;152(1):57-63.

425. Seo SJ, Ahn JY, Hong CK, et al. Expression of neutrophil gelatinase-associated lipocalin in skin epidermis. J Invest Dermatol. 2006 Feb;126(2):510-2.

426. Anwaar I, Gottsater A, Hedblad B, et al. Endothelial derived vasoactive factors and leukocyte derived inflammatory mediators in subjects with asymptomatic atherosclerosis. Angiology. 1998 Dec;49(12):957-66.

427. Elneihoum AM, Falke P, Hedblad B, et al. Leukocyte activation in atherosclerosis: correlation with risk factors. Atherosclerosis. 1997 May;131(1):79-84.

428. Bu DX, Hemdahl AL, Gabrielsen A, et al. Induction of neutrophil gelatinase-associated lipocalin in vascular injury via activation of nuclear factor-kappaB. Am J Pathol. 2006 Dec;169(6):2245-53.

429. Sahinarslan A, Kocaman SA, Bas D, et al. Plasma neutrophil gelatinase-associated lipocalin levels in acute myocardial infarction and stable coronary artery disease. Coronary artery disease. 2011 Aug;22(5):333-8.

430. Mishra J, Dent C, Tarabishi R, et al. Neutrophil gelatinase-associated lipocalin (NGAL) as a biomarker for acute renal injury after cardiac surgery. Lancet. 2005 Apr 2-8;365(9466):1231-8.

431. D'Anna R, Baviera G, Corrado F, et al. First trimester serum neutrophil gelatinase-associated lipocalin in gestational diabetes. Diabetic medicine : a journal of the British Diabetic Association. 2009 Dec;26(12):1293-5.

432. D'Anna R, Baviera G, Giordano D, et al. Second trimester neutrophil gelatinase-associated lipocalin as a potential prediagnostic marker of preeclampsia. Acta obstetricia et gynecologica Scandinavica. 2008;87(12):1370-3.

433. Bykov I, Junnikkala S, Pekna M, et al. Effect of chronic ethanol consumption on the expression of complement components and acute-phase proteins in liver. Clin Immunol. 2007 Aug;124(2):213-20.

434. Ouchi Y, Kubota Y, Ito C. Serial analysis of gene expression in methamphetamine- and phencyclidine-treated rodent cerebral cortices: are there common mechanisms? Ann N Y Acad Sci. 2004 Oct;1025:57-61.

435. Adler M, Hoffmann D, Ellinger-Ziegelbauer H, et al. Assessment of candidate biomarkers of drug-induced hepatobiliary injury in preclinical toxicity studies. Toxicology letters. 2010 Jun 16;196(1):1-11.

436. Axelsson L, Bergenfeldt M, Ohlsson K. Studies of the release and turnover of a human neutrophil lipocalin. Scand J Clin Lab Invest. 1995 Nov;55(7):577-88.

437. Shapiro NI, Trzeciak S, Hollander JE, et al. A prospective, multicenter derivation of a biomarker panel to assess risk of organ dysfunction, shock, and death in emergency department patients with suspected sepsis. Critical care medicine. 2009 Jan;37(1):96-104.

438. Bolignano D, Della Torre A, Lacquaniti A, et al. Neutrophil gelatinase-associated lipocalin levels in patients with crohn disease undergoing treatment with infliximab. Journal of investigative medicine : the official publication of the American Federation for Clinical Research. 2010 Mar;58(3):569-71.

439. Reghunathan R, Jayapal M, Hsu LY, et al. Expression profile of immune response genes in patients with Severe Acute Respiratory Syndrome. BMC immunology. 2005;6:2.

440. Hvidberg V, Jacobsen C, Strong RK, et al. The endocytic receptor megalin binds the iron transporting neutrophil-gelatinase-associated lipocalin with high affinity and mediates its cellular uptake. FEBS Lett. 2005 Jan 31;579(3):773-7.

441. Kumpers P, Hafer C, Lukasz A, et al. Serum neutrophil gelatinase-associated lipocalin at inception of renal replacement therapy predicts survival in critically ill patients with acute kidney injury. Crit Care. 2010;14(1):R9.

442. Vaidya VS, Waikar SS, Ferguson MA, et al. Urinary biomarkers for sensitive and specific detection of acute kidney injury in humans. Clinical and translational science. 2008 Dec;1(3):200-8.

443. Cruz DN, de Cal M, Garzotto F, et al. Plasma neutrophil gelatinase-associated lipocalin is an early biomarker for acute kidney injury in an adult ICU population. Intensive care medicine. 2010 Mar;36(3):444-51.

444. Shapiro NI, Trzeciak S, Hollander JE, et al. The diagnostic accuracy of plasma neutrophil gelatinase-associated lipocalin in the prediction of acute kidney injury in emergency department patients with suspected sepsis. Ann Emerg Med. 2010 Jul;56(1):52-9 e1.

445. Jeong TD, Kim S, Lee W, et al. Neutrophil gelatinase-associated lipocalin as an early biomarker of acute kidney injury in liver transplantation. Clin Transplant. 2012 Mar 8.

446. Iannetti A, Pacifico F, Acquaviva R, et al. The neutrophil gelatinase-associated lipocalin (NGAL), a NF-kappaB-regulated gene, is a survival factor for thyroid neoplastic cells. Proc Natl Acad Sci U S A. 2008 Sep 16;105(37):14058-63.

447. Miyamoto T, Kashima H, Suzuki A, et al. Laser-captured microdissection-microarray analysis of the genes involved in endometrial carcinogenesis: stepwise up-regulation of lipocalin2 expression in normal and neoplastic endometria and its functional relevance. Hum Pathol. 2011 Sep;42(9):1265-74.

448. Tong Z, Wu X, Ovcharenko D, et al. Neutrophil gelatinase-associated lipocalin as a survival factor. Biochem J. 2005 Oct 15;391(Pt 2):441-8.

449. Seth P, Porter D, Lahti-Domenici J, et al. Cellular and molecular targets of estrogen in normal human breast tissue. Cancer Res. 2002 Aug 15;62(16):4540-4.

450. Sun Y, Yokoi K, Li H, et al. NGAL expression is elevated in both colorectal adenoma-carcinoma sequence and cancer progression and enhances tumorigenesis in xenograft mouse models. Clin Cancer Res. 2011 Jul 1;17(13):4331-40.

451. Lee HJ, Lee EK, Lee KJ, et al. Ectopic expression of neutrophil gelatinase-associated lipocalin suppresses the invasion and liver metastasis of colon cancer cells. Int J Cancer. 2006 May 15;118(10):2490-7.

452. Nielsen OH, Gionchetti P, Ainsworth M, et al. Rectal dialysate and fecal concentrations of neutrophil gelatinase-associated lipocalin, interleukin-8, and tumor necrosis factor-alpha in ulcerative colitis. Am J Gastroenterol. 1999 Oct;94(10):2923-8.

453. Furutani M, Arii S, Mizumoto M, et al. Identification of a neutrophil gelatinase-associated lipocalin mRNA in human pancreatic cancers using a modified signal sequence trap method. Cancer Lett. 1998 Jan 9;122(1-2):209-14.

454. Bartsch S, Tschesche H. Cloning and expression of human neutrophil lipocalin cDNA derived from bone marrow and ovarian cancer cells. FEBS Lett. 1995 Jan 9;357(3):255-9.

455. Fernandez CA, Yan L, Louis G, et al. 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 Aug 1;11(15):5390-5.

456. Wang HJ, He XJ, Ma YY, et al. Expressions of neutrophil gelatinase-associated lipocalin in gastric cancer: a potential biomarker for prognosis and an ancillary diagnostic test. Anat Rec (Hoboken). 2010 Nov;293(11):1855-63.

457. Porta C, Paglino C, De Amici M, et al. Predictive value of baseline serum vascular endothelial growth factor and neutrophil gelatinase-associated lipocalin in advanced kidney cancer patients receiving sunitinib. Kidney Int. 2010 May;77(9):809-15.

458. Meyer K, Lee JS, Dyck PA, et al. Molecular profiling of hepatocellular carcinomas developing spontaneously in acyl-CoA oxidase deficient mice: comparison with liver tumors induced in wild-type mice by a peroxisome proliferator and a genotoxic carcinogen. Carcinogenesis. 2003 May;24(5):975-84.

459. Lee EK, Kim HJ, Lee KJ, et al. Inhibition of the proliferation and invasion of hepatocellular carcinoma cells by lipocalin 2 through blockade of JNK and PI3K/Akt signaling. Int J Oncol. 2011 Feb;38(2):325-33.

460. 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 May 6;98(9):1540-7.

461. Tong Z, Kunnumakkara AB, Wang H, et al. Neutrophil gelatinase-associated lipocalin: a novel suppressor of invasion and angiogenesis in pancreatic cancer. Cancer Res. 2008 Aug 1;68(15):6100-8.

462. Kjeldsen L, Cowland JB, Borregaard N. Human neutrophil gelatinase-associated lipocalin and homologous proteins in rat and mouse. Biochim Biophys Acta. 2000 Oct 18;1482(1-2):272-83.

463. Le Cabec V, Calafat J, Borregaard N. Sorting of the specific granule protein, NGAL, during granulocytic maturation of HL-60 cells. Blood. 1997 Mar 15;89(6):2113-21.

464. Wang W, Lin J, Guha S, et al. Target-specific agents imaging ectopic and orthotopic human pancreatic cancer xenografts. Pancreas. 2011 Jul;40(5):689-94.

465. Galamb O, Gyorffy B, Sipos F, et al. Inflammation, adenoma and cancer: objective classification of colon biopsy specimens with gene expression signature. Dis Markers. 2008;25(1):1-16.

466. McRobbie D. MRI - From Picture to Proton. 2nd ed. Cambridge: Cambridge University Press; 2007.

467. Bottomley PA, Atalar E, Lee RF, et al. Cardiovascular MRI probes for the outside in and for the inside out. MAGMA. 2000 Nov;11(1-2):49-51.

468. Artifon EL, Sakai P, Cunha JE, et al. Guidewire cannulation reduces risk of post-ERCP pancreatitis and facilitates bile duct cannulation. Am J Gastroenterol. 2007 Oct;102(10):2147-53.

469. Freeman ML, DiSario JA, Nelson DB, et al. Risk factors for post-ERCP pancreatitis: a prospective, multicenter study. Gastrointest Endosc. 2001 Oct;54(4):425-34.

470. Quick HH, Ladd ME, Zimmermann-Paul GG, et al. Single-loop coil concepts for intravascular magnetic resonance imaging. Magnet Reson Med. 1999 Apr;41(4):751-8.

471. Weiss CR, Georgiades C, Hofmann LV, et al. Intrabiliary MR imaging: Assessment of biliary obstruction with use of an intraluminal MR receiver coil. Journal of Vascular and Interventional Radiology. 2006 May;17(5):845-53.

472. Arepally A, Georgiades C, Hofmann LV, et al. Hilar cholangiocarcinoma: staging with intrabiliary MRI. AJR Am J Roentgenol. 2004 Oct;183(4):1071-4.

473. Ahmad MM, Syms RRA, Young IR, et al. Catheter-based flexible microcoil RF detectors for internal magnetic resonance imaging. J Micromech Microeng. 2009 Jul;19(7):-.

474. Inui K, Nakazawa S, Yoshino J, et al. Endoscopic MRI: preliminary results of a new technique for visualization and staging of gastrointestinal tumors. Endoscopy. 1995 Sep;27(7):480-5.

475. Inui K, Nakazawa S, Yoshino J, et al. Endoscopic MRI. Pancreas. 1998 Apr;16(3):413-7.

476. Feldman DR, Kulling DP, Hawes RH, et al. MR endoscopy: preliminary experience in human trials. Radiology. 1997 Mar;202(3):868-70.

477. Kulling D, Feldman DR, Kay CL, et al. Local staging of anal and distal colorectal tumors with the magnetic resonance endoscope. Gastrointest Endosc. 1998 Feb;47(2):172-8.

478. Kulling D, Feldman DR, Kay CL, et al. Local staging of esophageal cancer using endoscopic magnetic resonance imaging: prospective comparison with endoscopic ultrasound. Endoscopy. 1998 Nov;30(9):745-9.

479. de Souza N. Magnetic resonance imaging during upper GI endoscopy: Technical considerations and clinical feasibility. Minimally Invasive Therapy & Allied Technologies. 1995;4(5-6):5.

480. de Souza N. Combined MRI and fibreoptic colonoscopy: technical considerations and clinical feasibility. Minimally Invasive Therapy & Allied Technologies. 2000;9:5.

481. Dave UR, Williams AD, Wilson JA, et al. Esophageal cancer staging with endoscopic MR imaging: pilot study. Radiology. 2004 Jan;230(1):281-6.

482. Gilderdale DJ, Williams AD, Dave U, et al. An inductively-coupled, detachable receiver coil system for use with magnetic resonance compatible endoscopes. J Magn Reson Imaging. 2003 Jul;18(1):131-5.

483. Ginsberg DM, Melchner MJ. Optimum Geometry of Saddle Shaped Coils for Generating a Uniform Magnetic Field. Review of Scientific Instruments. 1970;41(1):122-&.

484. Coutts GA, Gilderdale DJ, Chui M, et al. Integrated and interactive position tracking and imaging of interventional tools and internal devices using small fiducial receiver coils. Magn Reson Med. 1998 Dec;40(6):908-13.

485. Firbank MJ, Coulthard A, Harrison RM, et al. A comparison of two methods for measuring the signal to noise ratio on MR images. Phys Med Biol. 1999 Dec;44(12):N261-4.

486. Moon JH, Ko BM, Choi HJ, et al. Intraductal balloon-guided direct peroral cholangioscopy with an ultraslim upper endoscope (with videos). Gastrointest Endosc. 2009 Aug;70(2):297-302.

487. Waxman I, Dillon T, Chmura K, et al. Feasibility of a novel system for intraductal balloon-anchored direct peroral cholangioscopy and endotherapy with an ultraslim endoscope (with videos). Gastrointest Endosc. 2010 Nov;72(5):1052-6.

488. Howe FA, Syms RR, Ahmad MM, et al. In vivo 31P magnetic resonance spectroscopy using a needle microcoil. Magn Reson Med. 2009 May;61(5):1238-41.

489. Hillenbrand CM, Elgort DR, Wong EY, et al. Active device tracking and high-resolution intravascular MRI using a novel catheter-based, opposed-solenoid phased array coil. Magn Reson Med. 2004 Apr;51(4):668-75.

490. Woytasik M, Ginefri JC, Raynaud JS, et al. Characterization of flexible RF microcoils dedicated to local MRI. Microsystem Technologies. 2007;13(11):1575-80.

491. Uelzen T. Mechanical and electrical properties of electroplated copper for MR-imaging coils. 2006.

492. Zuehlsdorff S, Umathum R, Volz S, et al. MR coil design for simultaneous tip tracking and curvature delineation of a catheter. Magnet Reson Med. 2004;52(1):214-8.

Appendices

Appendix 1: Sample collection and processing SOP

Appendix 2: Case record form (CRF)

Appendix 3: Results of all genotype testing, biliary transporter study

Imperial College London

Imperial College Healthcare

BILIARY MALIGNANCY PROTEOMICS AND METABONOMICS STUDY

Sample Collection SOP A: for subjects undergoing blood and urine collection only

SAMPLE COLLECTION

1. Confirm suitability, issue PIS, obtain **informed consent**. Copy of consent form to patient and notes

2. Complete CRF including demographic details, clinical details, dietary and drug history

3. A total of four tubes and **16mLs** of blood is required. Note the time of blood collection on the CRF
   - Take one **4mLs EDTA** tube (PURPLE TOP VACUTAINER) for proteomic and genomic studies
   - Take one **4mLs plain serum/no additive** tube (BRIGHT RED TOP VACUTAINER) for metabonomic studies using MR spectroscopy
   - Take one **4mLs** lithium heparin (GREEN TOP VACUTAINER) for metabonomic studies using MR spectroscopy
   - Take one **4mLs** SST tube (YELLOW TOP VACUTAINER) for standard biochemical analysis

4. Obtain **≥ 15mLs** of urine in a plain, additive free universal container

5. All samples (blood and urine) should be kept on ice with light excluded whilst processing is awaited

6. Samples should be left for a minimum of 30 minutes before processing, but should be processed as soon as possible after this. Samples should be processed within a maximum of 2.5 hours. Collection and processing times should be documented on the CRF.

SAMPLE PROCESSING

7. Blood:
   - Keep blood samples on ice or in +4 ^{o}C fridge whilst processing awaited
   - Spin all blood samples at +4^{o}C at **1000g for 10 minutes**.
   - Store 200µL aliquots of each supernatant in Eppendorf tubes.
   - Please aim for a minimum of 4 aliquots from each vacutainer tube.
   - Note the time of centrifuging the on the CRF.
   - Remove the buffy coat layer from the EDTA sample and store in a cryovial or Eppendorph tube.
   - Discard all remaining cell layers in the vacutainers.

8. Urine
   - Keep urine on ice or in +4^{o}C fridge whilst processing awaited
   - Spin the urine sample at +4^{o}C at **1000g for 10 minutes**.
   - Store **six** 2mL aliquots of urine supernatant in Eppendorph tubes
   - Discard any remaining urine, cellular debris etc remaining in the urine tube

Version 1.8

Imperial College London

Imperial College Healthcare

9. All samples should be stored at -80^{o}C as soon as possible. However, short term storage in a standard -20^{o}C freezer is acceptable, pending transfer to a -80^{o}C freezer

10. Allocate case number and record this on CRF

11. Labelling

    - A unique case/subject identifier should be allocated in the site master file
    - Each aliquot should be labelled with a code.
    - This code starts with the allocated case/subject number, followed by an alphabetic code for the type of sample, followed by the aliquot number.
    - The coding for sample type is as follows: **U** = Urine, **S** = red cap serum, **E** = EDTA plasma, **H** = lithium heparin plasma, **T** = SST serum **BC** = buffy coat
    - E.g. for the third aliquot of EDTA plasma from subject HC1234 the label would read: '***HC1234E3***'

12. Record sample details in sample registry

13. Archive consent form and CRF securely.

Please direct any queries about this protocol to Dr Chris Wadsworth or Dr Mo Shariff
Email: c.wadsworth@imperial.ac.uk or m.shariff@imperial.ac.uk
Office: 0208 383 8067 or 0208 383 5856
Mobile: 07879811042 or 07971569850

Version 1.8

Imperial College London

Imperial College Healthcare NHS
NHS Trust

BILIARY MALIGNANCY PROTEOMICS AND METABONOMICS STUDY - CRF

Please complete as fully as possible.

CASE NUMBER:

DEMOGRAPHICS

Patient initials:_____ Hospital number:_____ Sex: M / F DoB: / /

Ethnic origin:_____ Units of alcohol/day: _____ Cigarettes/day:_____

Occupation:_____ Rec drugs/betel nut:_____ LMP & cycle length:_____

Weight (kg):_____ Height (m):_____ BMI:_____

Exercise in last 24hrs: none / mild / moderate / hard

MEDICAL HISTORY

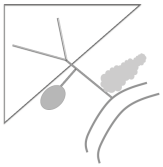
Principal current hepatobiliary diagnosis (circle all that apply):

CC / HCC / Liver Mets / PSC / Pancreatic Ca / Cirrhotic / Healthy control / Other……

Diagnostic certainty:

☐ Histologically proven ☐ MDT consensus, fully investigated ☐ Diagnosis likely, Ix pending

Location of disease, if known:



☐ Diffuse process ☐ Metastatic, extrahepatic disease

Previous therapy *(tick all that apply, with date of last session)*

☐ Resection / / ☐ Ablation / / ☐ Chemotherapy / / ☐ Radiotherapy
☐ Biliary stent / / ☐ TACE / / ☐ PDT / / ☐ Other…… / /

OTHER CURRENT MEDICAL PROBLEMS

List active diagnoses (esp malignant, inflammatory or infective processes):

☐ Diabetes ☐ Cholangitis ☐ Systemic sepsis, any source

DRUG HISTORY

Please list all current medications (including herbal remedies and paracetamol/acetaminophen):

VERSION 2.1

Imperial College London

Imperial College Healthcare NHS Trust

DIETARY HISTORY

Time last ate ANYTHING ___ : ___

Details of food and drink taken in the last 6 hours:

List of other food and drink taken in last 24 hrs:

Checklist of specific foods and drink taken in last 72hrs:

Food	✓	Details or comments	Food	✓	Details or comments
Vegetarian		*Vegan/non-vegan*	Nutrition drinks		
Meat		*Red/white*	Cheese		
Fish		*Type:*	Grapefruit		
Fizzy drinks			Berries		
Tea		*Normal/herbal*	Liquorice		
Milk			Walnuts		
Cherries			Vanilla		
Yoghurt					

SAMPLE COLLECTION

Pt. information sheet issued? Y / N

Informed consent obtained? Y / N

Samples taken:

- ☐ 2 x EDTA (Purple FBC tube)
- ☐ 1 x plain serum (Bright red cap tube)
- ☐ 15 ml urine ☐ MSU ☐ CSU
- ☐ Bile, uncontaminated
- ☐ Bile, with some saline
- ☐ Bile, contaminated with contrast

Samples placed on ice Y / N

Light excluded? Y / N

Samples processed by:

Samples stored in which freezer?:

Blood samples taken at: __:__

Plain serum sample centrifuged at: __:__

☐ **All sections completed esp. time last ate, time of blood sample, time of centrifuge**

OTHER INVESTIGATIONS

Most recent bloods on __ / __/ __ ALT: ALP: Bili: Alb: INR:
AFP: Ca 19-9: CeA: Ur: Cr:

Summary of ERCP findings *(if performed)*:

Summary of most recent axial imaging findings *(if performed)*:

Please direct any queries about this CRF to Dr Chris Wadsworth or Dr Mo Shariff
Email: c.wadsworth@imperial.ac.uk or m.shariff@imperial.ac.uk
Office: 0208 383 8067 or 0208 383 5856
Fax: 0207 402 2796
Mobile: 07879811042 or 07971569850

VERSION 2.1

Table 2.7 Results from allelic association testing, Cochran-Armitage trend testing, dominant and recessive modelling for all SNPs

Abbreviations: Chrom, Chromosome; SNP, single nucleotide polymorphism; A1, allele 1; A2, allele 2; Freq A, frequency of allele 1 in affected group; Freq UnA, frequency of allele 1 in unaffected group; OR, odds ratio; SE, standard error; L95 and U95, lower and upper limits of OR 95% confidence interval; allelic $\chi 2$, allelic association testing result; p, p-value; **bold text** - p-value <0.05.

Gene **Chrom**	**SNP**	**A1**	**Freq A**	**Freq UnA**	**A2**	**OR**	**SE**	**L95 OR**	**U95 OR**	**χ^2 TEST**	**AFF**	**U-FF**	**χ^2**	**DF**	**P**
ABCB11	rs497692	A	0.4908	0.4506	G	1.175	0.1423	0.8891	1.553	ALLELIC	160/166	228/278	1.288	1	0.2565
Chrom 2										TREND	160/166	228/278	1.149	1	0.2838
										GENO	43/74/46	60/108/85	1.361	2	0.5064
										DOM	117/46	168/85	1.328	1	0.2492
										REC	43/120	60/193	0.3779	1	0.5387
	rs3755157	T	0.1108	0.098	C	1.146	0.234	0.7247	1.814	ALLELIC	35/281	49/451	0.3414	1	0.559
										TREND	35/281	49/451	0.3268	1	0.5675
										GENO	4/27/127	2/45/203	-	-	-
										DOM	31/127	47/203	-	-	-
										REC	4/154	2/248	-	-	-
	rs6709087	G	0.2406	0.2273	A	1.077	0.1684	0.7745	1.499	ALLELIC	77/243	115/391	0.1959	1	0.6581
										TREND	77/243	115/391	0.1994	1	0.6552
										GENO	6/65/89	15/85/153	2.641	2	0.2669
										DOM	71/89	100/153	0.95	1	0.3297
										REC	6/154	15/238	0.9641	1	0.3261
	rs853773	T	0.4905	0.45	C	1.177	0.144	0.8873	1.56	ALLELIC	155/161	225/275	1.277	1	0.2585
										TREND	155/161	225/275	1.167	1	0.2799
										GENO	40/75/43	58/109/83	1.625	2	0.4438
										DOM	115/43	167/83	1.625	1	0.2025
										REC	40/118	58/192	0.2376	1	0.6259
	rs853772	C	0.5032	0.4643	A	1.169	0.1437	0.8818	1.549	ALLELIC	159/157	234/270	1.176	1	0.2781
										TREND	159/157	234/270	1.064	1	0.3023
Gene	**SNP**	**A1**	**Freq A**	**Freq UnA**	**A2**	**OR**	**SE**	**L95 OR**	**U95 OR**	**χ^2 TEST**	**AFF**	**U-FF**	**χ^2**	**DF**	**P**

Chrom															
										GENO	44/71/43	61/112/79	1.066	2	0.5869
										DOM	115/43	173/79	0.794	1	0.3729
										REC	44/114	61/191	0.676	1	0.411
	rs3821120	C	0.4099	0.4012	G	1.037	0.1451	0.7802	1.378	ALLELIC	132/190	203/303	0.06257	1	0.8025
										TREND	132/190	203/303	0.06181	1	0.8037
										GENO	25/82/54	44/115/94	1.185	2	0.553
										DOM	107/54	159/94	0.5594	1	0.4545
										REC	25/136	44/209	0.246	1	0.6199
	rs16823014	A	0.06329	0.076	G	0.8215	0.2861	0.4689	1.439	ALLELIC	20/296	38/462	0.4737	1	0.4913
										TREND	20/296	38/462	0.4577	1	0.4987
										GENO	1/18/139	2/34/214	-	-	-
										DOM	19/139	36/214	-	-	-
										REC	1/157	2/248	-	-	-
	rs4148797	G	0.3057	0.2874	A	1.092	0.1569	0.8028	1.485	ALLELIC	96/218	146/362	0.3139	1	0.5753
										TREND	96/218	146/362	0.2953	1	0.5868
										GENO	16/64/77	25/96/133	0.4405	2	0.8023
										DOM	80/77	121/133	0.4274	1	0.5133
										REC	16/141	25/229	0.01313	1	0.9088
	rs16856300	C	0.4267	0.3944	A	1.143	0.1482	0.8545	1.528	ALLELIC	128/172	198/304	0.8092	1	0.3683
										TREND	128/172	198/304	0.7869	1	0.375
										GENO	25/78/47	44/110/97	2.777	2	0.2494
										DOM	103/47	154/97	2.181	1	0.1397
										REC	25/125	44/207	0.04911	1	0.8246
	rs4148794	C	0.4399	0.3755	T	1.306	0.1458	0.9814	1.738	ALLELIC	139/177	190/316	3.359	1	0.06685
										TREND	139/177	190/316	3.102	1	0.07819
										GENO	31/77/50	43/104/106	4.35	2	0.1136
										DOM	108/50	147/106	4.341	1	**0.03722**
										REC	31/127	43/210	0.4537	1	0.5006
Gene	**SNP**	**A1**	**Freq A**	**Freq UnA**	**A2**	**OR**	**SE**	**L95 OR**	**U95 OR**	**χ^2 TEST**	**AFF**	**U-FF**	**χ^2**	**DF**	**P**

Chrom															
	rs17267869	C	0.1338	0.1587	T	0.8184	0.2058	0.5468	1.225	ALLELIC	42/272	80/424	0.9507	1	0.3295
										TREND	42/272	80/424	0.9713	1	0.3244
										GENO	0/42/115	8/64/180	-	-	-
										DOM	42/115	72/180	-	-	-
										REC	0/157	8/244	-	-	-
	rs3770585	A	0.4	0.49	G	0.6939	0.148	0.5192	0.9273	ALLELIC	120/180	245/255	6.122	1	**0.01335**
										TREND	120/180	245/255	6.201	1	**0.01277**
										GENO	23/74/53	59/127/64	6.202	2	**0.04501**
										DOM	97/53	186/64	4.292	1	**0.0383**
										REC	23/127	59/191	3.931	1	**0.0474**
	rs2287622	T	0.45	0.3563	C	1.478	0.1456	1.111	1.966	ALLELIC	144/176	181/327	7.229	1	**0.007173**
										TREND	144/176	181/327	6.801	1	**0.009112**
										GENO	31/82/47	39/103/112	8.991	2	**0.01116**
										DOM	113/47	142/112	8.991	1	**0.002713**
										REC	31/129	39/215	1.13	1	0.2879
	rs2058996	A	0.4903	0.424	G	1.307	0.1455	0.9823	1.738	ALLELIC	151/157	212/288	3.382	1	0.06591
										TREND	151/157	212/288	3.619	1	0.05713
										GENO	36/79/39	39/134/77	4.195	2	0.1228
										DOM	115/39	173/77	1.396	1	0.2374
										REC	36/118	39/211	3.812	1	0.05088
	rs3770596	A	0.4968	0.4221	T	1.351	0.1457	1.016	1.798	ALLELIC	155/157	206/282	4.285	1	**0.03846**
										TREND	155/157	206/282	4.096	1	**0.04297**
										GENO	35/85/36	51/104/89	8.406	2	**0.01495**
										DOM	120/36	155/89	7.951	1	**0.004805**
										REC	35/121	51/193	0.1327	1	0.7156
	rs13416802	C	0.1313	0.1146	G	1.167	0.2165	0.7634	1.784	ALLELIC	42/278	58/448	0.5092	1	0.4755
										TREND	42/278	58/448	0.5098	1	0.4752
										GENO	2/38/120	4/50/199	-	-	-
Gene	**SNP**	**A1**	**Freq A**	**Freq UnA**	**A2**	**OR**	**SE**	**L95 OR**	**U95 OR**	**χ^2 TEST**	**AFF**	**U-FF**	**χ^2**	**DF**	**P**

Chrom															
									DOM	40/120	54/199	-	-	-
									REC	2/158	4/249	-	-	-
	rs2287618	A	0.3449	0.2996	G	1.231	0.1532	0.9118	1.662	ALLELIC	109/207	151/353	1.843	1	0.1746
										TREND	109/207	151/353	1.713	1	0.1907
										GENO	18/73/67	30/91/131	4.339	2	0.1142
										DOM	91/67	121/131	3.569	1	0.05889
										REC	18/140	30/222	0.02466	1	0.8752
	rs7605199	G	0.4299	0.502	A	0.7483	0.1445	0.5638	0.9932	ALLELIC	135/179	255/253	4.038	1	**0.04449**
										TREND	135/179	255/253	4.019	1	**0.04499**
										GENO	28/79/50	65/125/64	4.151	2	0.1255
										DOM	107/50	190/64	2.141	1	0.1434
										REC	28/129	65/189	3.334	1	0.06786
	rs3815676	G	0.02454	0.02953	A	0.8268	0.4437	0.3466	1.973	ALLELIC	8/318	15/493	0.1842	1	0.6678
										TREND	8/318	15/493	0.1896	1	0.6633
										GENO	0/8/155	0/15/239	-	-	-
										DOM	8/155	15/239	-	-	-
										REC	0/163	0/254	-	-	-
	rs4148773	T	0.07669	0.0754	C	1.019	0.2679	0.6024	1.722	ALLELIC	25/301	38/466	0.004699	1	0.9454
										TREND	25/301	38/466	0.004602	1	0.9459
										GENO	1/23/139	2/34/216	-	-	-
										DOM	24/139	36/216	-	-	-
										REC	1/162	2/250	-	-	-
	rs3814382	T	0.3549	0.4104	C	0.7906	0.1474	0.5923	1.055	ALLELIC	115/209	206/296	2.545	1	0.1106
										TREND	115/209	206/296	2.555	1	0.1099
										GENO	19/77/66	43/120/88	2.768	2	0.2505
										DOM	96/66	163/88	1.359	1	0.2437
										REC	19/143	43/208	2.253	1	0.1334
	rs10930343	G	0.4815	0.4272	A	1.245	0.1429	0.9411	1.648	ALLELIC	156/168	217/291	2.36	1	0.1245
Gene	**SNP**	**A1**	**Freq A**	**Freq UnA**	**A2**	**OR**	**SE**	**L95 OR**	**U95 OR**	**χ^2 TEST**	**AFF**	**U-FF**	**χ^2**	**DF**	**P**

Chrom															
										TREND	156/168	217/291	2.306	1	0.1289
										GENO	34/88/40	52/113/89	5.408	2	0.06695
										DOM	122/40	165/89	4.951	1	**0.02608**
										REC	34/128	52/202	0.01601	1	0.8993
	rs7577650	A	0.4472	0.3972	G	1.228	0.1443	0.9252	1.629	ALLELIC	144/178	201/305	2.022	1	0.1551
										TREND	144/178	201/305	1.999	1	0.1574
										GENO	31/82/48	42/117/94	2.388	2	0.303
										DOM	113/48	159/94	2.353	1	0.1251
										REC	31/130	42/211	0.4771	1	0.4897
ABCB4	rs31652aa	G	0.1335	0.09717	A	1.432	0.2234	0.9242	2.219	ALLELIC	43/279	48/446	2.603	1	0.1067
Chrom 7										TREND	43/279	48/446	2.485	1	0.115
										GENO	5/33/123	2/44/201	-	-	-
										DOM	38/123	46/201	-	-	-
										REC	5/156	2/245	-	-	-
	rs2097937	G	0.2125	0.1574	A	1.445	0.1836	1.008	2.07	ALLELIC	68/252	79/423	4.045	1	**0.04431**
										TREND	68/252	79/423	3.924	1	**0.0476**
										GENO	11/46/103	4/71/176	-	-	-
										DOM	57/103	75/176	-	-	-
										REC	11/149	4/247	-	-	-
	rs31653aa	A	0.09938	0.08434	G	1.198	0.2464	0.7392	1.942	ALLELIC	32/290	42/456	0.5389	1	0.4629
										TREND	32/290	42/456	0.4994	1	0.4798
										GENO	4/24/133	2/38/209	-	-	-
										DOM	28/133	40/209	-	-	-
										REC	4/157	2/247	-	-	-
	rs2373593	C	0.1104	0.08333	A	1.366	0.2392	0.8545	2.182	ALLELIC	36/290	42/462	1.707	1	0.1914
										TREND	36/290	42/462	1.601	1	0.2057
										GENO	4/28/131	2/38/212	-	-	-
										DOM	32/131	40/212	-	-	-
Gene	**SNP**	**A1**	**Freq A**	**Freq UnA**	**A2**	**OR**	**SE**	**L95 OR**	**U95 OR**	**χ^2 TEST**	**AFF**	**U-FF**	**χ^2**	**DF**	**P**

Chrom															
										REC	4/159	2/250	-	-	-
	rs31666aa	C	0.1437	0.132	T	1.104	0.207	0.7358	1.656	ALLELIC	46/274	66/434	0.2284	1	0.6327
										TREND	46/274	66/434	0.2222	1	0.6374
										GENO	3/40/117	6/54/190	-	-	-
										DOM	43/117	60/190	-	-	-
										REC	3/157	6/244	-	-	-
	rs31668aa	G	0.06707	0.06324	A	1.065	0.2865	0.6074	1.867	ALLELIC	22/306	32/474	0.04826	1	0.8261
										TREND	22/306	32/474	0.04431	1	0.8333
										GENO	3/16/145	1/30/222	-	-	-
										DOM	19/145	31/222	-	-	-
										REC	3/161	1/252	-	-	-
	rs8187799	G	0.09451	0.0754	A	1.28	0.2532	0.7793	2.102	ALLELIC	31/297	38/466	0.9546	1	0.3286
										TREND	31/297	38/466	0.8942	1	0.3444
										GENO	2/27/135	3/32/217	-	-	-
										DOM	29/135	35/217	-	-	-
										REC	2/162	3/249	-	-	-
	rs31676aa	T	0.2075	0.2044	C	1.02	0.177	0.7208	1.442	ALLELIC	66/252	103/401	0.01209	1	0.9125
										TREND	66/252	103/401	0.01071	1	0.9176
										GENO	11/44/104	15/73/164	0.2028	2	0.9036
										DOM	55/104	88/164	0.004664	1	0.9456
										REC	11/148	15/237	0.1535	1	0.6952
	rs1149222	G	0.1962	0.2191	T	0.8699	0.1781	0.6136	1.233	ALLELIC	62/254	110/392	0.6135	1	0.4335
										TREND	62/254	110/392	0.5495	1	0.4585
										GENO	11/40/107	15/80/156	2.037	2	0.3612
										DOM	51/107	95/156	1.311	1	0.2523
										REC	11/147	15/236	0.1583	1	0.6907
	rs4148826	G	0.177	0.2028	A	0.8458	0.183	0.5908	1.211	ALLELIC	57/265	103/405	0.8389	1	0.3597
										TREND	57/265	103/405	0.7209	1	0.3959
Gene	**SNP**	**A1**	**Freq A**	**Freq UnA**	**A2**	**OR**	**SE**	**L95 OR**	**U95 OR**	**χ^2 TEST**	**AFF**	**U-FF**	**χ^2**	**DF**	**P**

Chrom															
										GENO	9/39/113	17/69/168	0.7571	2	0.6849
										DOM	48/113	86/168	0.7373	1	0.3905
										REC	9/152	17/237	0.2041	1	0.6514
	rs2109505	A	0.1687	0.2016	T	0.8038	0.1848	0.5596	1.155	ALLELIC	55/271	102/404	1.399	1	0.2369
										TREND	55/271	102/404	1.223	1	0.2688
										GENO	7/41/115	17/68/168	1.374	2	0.5032
										DOM	48/115	85/168	0.7846	1	0.3757
										REC	7/156	17/236	1.072	1	0.3004
	ABCB4_ind	:-	0.02516	0.01786	AGAAA	1.419	0.4913	0.5419	3.718	ALLELIC	8/310	9/495	0.513	1	0.4738
										TREND	8/310	9/495	0.5241	1	0.4691
										GENO	0/8/151	0/9/243	-	-	-
										DOM	8/151	9/243	-	-	-
										REC	0/159	0/252	-	-	-
	rs1202283	G	0.5127	0.4724	A	1.175	0.1437	0.8867	1.557	ALLELIC	161/153	240/268	1.261	1	0.2614
										TREND	161/153	240/268	1.168	1	0.2798
										GENO	46/69/42	60/120/74	1.637	2	0.4411
										DOM	115/42	180/74	0.2718	1	0.6021
										REC	46/111	60/194	1.634	1	0.2011
	rs4148812	C	0.3281	0.266	G	1.348	0.1563	0.9921	1.831	ALLELIC	105/215	133/367	3.656	1	0.05588
										TREND	105/215	133/367	3.594	1	0.05801
										GENO	20/65/75	16/101/133	4.905	2	0.08608
										DOM	85/75	117/133	1.562	1	0.2114
										REC	20/140	16/234	4.532	1	**0.03326**
	rs2302386	G	0.1419	0.1627	A	0.851	0.2057	0.5686	1.274	ALLELIC	42/254	82/422	0.6164	1	0.4324
										TREND	42/254	82/422	0.5789	1	0.4467
										GENO	4/34/110	9/64/179	-	-	-
										DOM	38/110	73/179	-	-	-
										REC	4/144	9/243	-	-	-
Gene	**SNP**	**A1**	**Freq A**	**Freq UnA**	**A2**	**OR**	**SE**	**L95 OR**	**U95 OR**	**χ^2 TEST**	**AFF**	**U-FF**	**χ^2**	**DF**	**P**

Chrom															
ABCC2	rs717620	T	0.1957	0.206	C	0.9375	0.1788	0.6604	1.331	ALLELIC	63/259	103/397	0.1301	1	0.7183
Chrom 10										TREND	63/259	103/397	0.1316	1	0.7167
										GENO	4/55/102	12/79/159	-	-	-
										DOM	59/102	91/159	-	-	-
										REC	4/157	12/238	-	-	-
	rs2756109	G	0.4395	0.4272	T	1.051	0.1448	0.7916	1.397	ALLELIC	138/176	217/291	0.1201	1	0.7289
										TREND	138/176	217/291	0.111	1	0.739
										GENO	33/72/52	52/113/89	0.1587	2	0.9237
										DOM	105/52	165/89	0.1584	1	0.6906
										REC	33/124	52/202	0.01768	1	0.8942
	rs4148391	A	0.1442	0.138	T	1.052	0.2041	0.7053	1.57	ALLELIC	47/279	69/431	0.06227	1	0.8029
										TREND	47/279	69/431	0.06373	1	0.8007
										GENO	3/41/119	4/61/185	-	-	-
										DOM	44/119	65/185	-	-	-
										REC	3/160	4/246	-	-	-
	rs2073337	G	0.375	0.3976	A	0.9089	0.1468	0.6816	1.212	ALLELIC	120/200	202/306	0.4234	1	0.5153
										TREND	120/200	202/306	0.4442	1	0.5051
										GENO	19/82/59	39/124/91	0.9946	2	0.6082
										DOM	101/59	163/91	0.04668	1	0.8289
										REC	19/141	39/215	0.9864	1	0.3206
	rs2002042	T	0.2438	0.2233	C	1.121	0.1677	0.8072	1.558	ALLELIC	79/245	113/393	0.4671	1	0.4943
										TREND	79/245	113/393	0.4884	1	0.4847
										GENO	6/67/89	13/87/153	2.256	2	0.3236
										DOM	73/89	100/153	1.245	1	0.2645
										REC	6/156	13/240	0.4653	1	0.4952
	rs7476245	A	0.07187	0.05976	G	1.218	0.2869	0.6944	2.138	ALLELIC	23/297	30/472	0.4754	1	0.4905
										TREND	23/297	30/472	0.5106	1	0.4749
										GENO	0/23/137	0/30/221	-	-	-
Gene	**SNP**	**A1**	**Freq A**	**Freq UnA**	**A2**	**OR**	**SE**	**L95 OR**	**U95 OR**	**χ^2 TEST**	**AFF**	**U-FF**	**χ^2**	**DF**	**P**

Chrom															
										DOM	23/137	30/221	-	-	-
										REC	0/160	0/251	-	-	-
	rs3740066	A	0.3281	0.3834	G	0.7854	0.1501	0.5852	1.054	ALLELIC	105/215	194/312	2.593	1	0.1073
										TREND	105/215	194/312	2.802	1	0.09413
										GENO	14/77/69	33/128/92	2.857	2	0.2396
										DOM	91/69	161/92	1.884	1	0.1699
										REC	14/146	33/220	1.792	1	0.1807
	rs3740065	C	0.1415	0.126	T	1.143	0.2099	0.7578	1.725	ALLELIC	45/273	63/437	0.408	1	0.523
										TREND	45/273	63/437	0.3696	1	0.5432
										GENO	5/35/119	7/49/194	0.4113	2	0.8141
										DOM	40/119	56/194	0.4113	1	0.5213
										REC	5/154	7/243	0.04054	1	0.8404
	rs3740063	C	0.4264	0.4605	T	0.8709	0.1432	0.6578	1.153	ALLELIC	139/187	233/273	0.9323	1	0.3343
										TREND	139/187	233/273	1.013	1	0.3143
										GENO	28/83/52	47/139/67	1.426	2	0.4902
										DOM	111/52	186/67	1.426	1	0.2325
										REC	28/135	47/206	0.1313	1	0.7171
ATP8B1	rs7241054	T	0.4658	0.504	C	0.8584	0.1429	0.6487	1.136	ALLELIC	150/172	254/250	1.143	1	0.285
Chrom 18										TREND	150/172	254/250	1.163	1	0.2808
										GENO	33/84/44	64/126/62	1.38	2	0.5016
										DOM	117/44	190/62	0.3826	1	0.5362
										REC	33/128	64/188	1.312	1	0.2519
	rs1968274	C	0.2044	0.2185	T	0.9189	0.1757	0.6512	1.297	ALLELIC	65/253	111/397	0.2319	1	0.6301
										TREND	65/253	111/397	0.2311	1	0.6307
										GENO	8/49/102	11/89/154	0.8216	2	0.6631
										DOM	57/102	100/154	0.5145	1	0.4732
										REC	8/151	11/243	0.1094	1	0.7408
	rs17685876	G	0.2236	0.25	A	0.864	0.1687	0.6207	1.203	ALLELIC	72/250	126/378	0.7512	1	0.3861
Gene	**SNP**	**A1**	**Freq A**	**Freq UnA**	**A2**	**OR**	**SE**	**L95 OR**	**U95 OR**	**χ^2 TEST**	**AFF**	**U-FF**	**χ^2**	**DF**	**P**

Chrom															
										TREND	72/250	126/378	0.8131	1	0.3672
										GENO	6/60/95	12/102/138	0.8131	2	0.6659
										DOM	66/95	114/138	0.7197	1	0.3962
										REC	6/155	12/240	0.2526	1	0.6153
	rs317826	A	0.3057	0.253	G	1.3	0.1596	0.9512	1.778	ALLELIC	96/218	128/378	2.717	1	0.09926
										TREND	96/218	128/378	2.88	1	0.08969
										GENO	14/68/75	12/104/137	3.537	2	0.1706
										DOM	82/75	116/137	1.579	1	0.2089
										REC	14/143	12/241	2.842	1	0.09183
	rs4308033	A	0.4359	0.3933	C	1.192	0.146	0.8954	1.587	ALLELIC	136/176	199/307	1.45	1	0.2286
										TREND	136/176	199/307	1.473	1	0.2248
										GENO	30/76/50	37/125/91	1.689	2	0.4298
										DOM	106/50	162/91	0.6555	1	0.4182
										REC	30/126	37/216	1.495	1	0.2215
	rs4306606	T	0.307	0.2579	C	1.274	0.1589	0.9333	1.74	ALLELIC	97/219	130/374	2.332	1	0.1268
										TREND	97/219	130/374	2.315	1	0.1281
										GENO	16/65/77	16/98/138	2.572	2	0.2764
										DOM	81/77	114/138	1.415	1	0.2343
										REC	16/142	16/236	1.926	1	0.1652
	rs317845	C	0.2437	0.2461	T	0.9876	0.166	0.7133	1.367	ALLELIC	78/242	125/383	0.005676	1	0.9399
										TREND	78/242	125/383	0.005979	1	0.9384
										GENO	10/58/92	11/103/140	1.279	2	0.5275
										DOM	68/92	114/140	0.2261	1	0.6345
										REC	10/150	11/243	0.751	1	0.3862
	rs317838	T	0.3187	0.354	C	0.8538	0.1521	0.6337	1.15	ALLELIC	102/218	177/323	1.08	1	0.2987
										TREND	102/218	177/323	1.175	1	0.2783
										GENO	14/74/72	26/125/99	1.237	2	0.5387
										DOM	88/72	151/99	1.17	1	0.2794
Gene	**SNP**	**A1**	**Freq A**	**Freq UnA**	**A2**	**OR**	**SE**	**L95 OR**	**U95 OR**	**χ^2 TEST**	**AFF**	**U-FF**	**χ^2**	**DF**	**P**

Chrom															
										REC	14/146	26/224	0.3017	1	0.5828
	rs317837	C	0.2357	0.25	T	0.925	0.1679	0.6657	1.285	ALLELIC	74/240	127/381	0.2157	1	0.6423
										TREND	74/240	127/381	0.2328	1	0.6294
										GENO	8/58/91	11/105/138	0.8251	2	0.6619
										DOM	66/91	116/138	0.5185	1	0.4715
										REC	8/149	11/243	0.1287	1	0.7198
	rs11659313	C	0.1437	0.1601	T	0.8809	0.2002	0.595	1.304	ALLELIC	46/274	81/425	0.4017	1	0.5262
										TREND	46/274	81/425	0.4	1	0.5271
										GENO	3/40/117	7/67/179	-	-	-
										DOM	43/117	74/179	-	-	-
										REC	3/157	7/246	-	-	-
	rs319454	A	0.3491	0.2811	G	1.371	0.1542	1.014	1.855	ALLELIC	111/207	140/358	4.205	1	**0.0403**
										TREND	111/207	140/358	4.139	1	**0.04192**
										GENO	21/69/69	19/102/128	4.505	2	0.1052
										DOM	90/69	121/128	2.493	1	0.1144
										REC	21/138	19/230	3.413	1	0.06468
	rs7236365	T	0.4245	0.4084	C	1.069	0.1453	0.8039	1.421	ALLELIC	135/183	205/297	0.2095	1	0.6472
										TREND	135/183	205/297	0.2149	1	0.643
										GENO	28/79/52	40/125/86	0.2349	2	0.8892
										DOM	107/52	165/86	0.1059	1	0.7449
										REC	28/131	40/211	0.1971	1	0.6571
	rs160993	A	0.1595	0.132	G	1.248	0.2008	0.8419	1.85	ALLELIC	52/274	66/434	1.22	1	0.2694
										TREND	52/274	66/434	1.183	1	0.2768
										GENO	5/42/116	5/56/189	1.198	2	0.5492
										DOM	47/116	61/189	1.005	1	0.3162
										REC	5/158	5/245	0.4759	1	0.4903
	rs319448	G	0.4625	0.3849	A	1.375	0.1447	1.035	1.826	ALLELIC	148/172	194/310	4.852	1	**0.02761**
										TREND	148/172	194/310	4.576	1	**0.03242**
Gene	**SNP**	**A1**	**Freq A**	**Freq UnA**	**A2**	**OR**	**SE**	**L95 OR**	**U95 OR**	**χ^2 TEST**	**AFF**	**U-FF**	**χ^2**	**DF**	**P**

Chrom															
										GENO	39/70/51	38/118/96	5.789	2	0.05533
										DOM	109/51	156/96	1.65	1	0.199
										REC	39/121	38/214	5.565	1	**0.01833**
	rs319440	C	0.4255	0.4558	T	0.8841	0.1442	0.6664	1.173	ALLELIC	137/185	227/271	0.7301	1	0.3929
										TREND	137/185	227/271	0.7214	1	0.3957
										GENO	32/73/56	50/127/72	1.723	2	0.4226
										DOM	105/56	177/72	1.567	1	0.2106
										REC	32/129	50/199	0.002557	1	0.9597
	rs17686300	A	0.2932	0.246	G	1.272	0.1602	0.9288	1.741	ALLELIC	95/229	123/377	2.252	1	0.1334
										TREND	95/229	123/377	2.193	1	0.1386
										GENO	16/63/83	15/93/142	2.595	2	0.2732
										DOM	79/83	108/142	1.228	1	0.2677
										REC	16/146	15/235	2.123	1	0.1451
	rs319457	T	0.1074	0.1076	C	0.9978	0.2297	0.6361	1.565	ALLELIC	35/291	54/448	8.89E-05	1	0.9925
										TREND	35/291	54/448	8.84E-05	1	0.9925
										GENO	0/35/128	5/44/202	-	-	-
										DOM	35/128	49/202	-	-	-
										REC	0/163	5/246	-	-	-
	rs319406	G	0.3639	0.386	A	0.9101	0.1487	0.68	1.218	ALLELIC	115/201	193/307	0.4016	1	0.5263
										TREND	115/201	193/307	0.4151	1	0.5194
										GENO	22/71/65	33/127/90	1.397	2	0.4974
										DOM	93/65	160/90	1.085	1	0.2975
										REC	22/136	33/217	0.04352	1	0.8348
	rs319409	A	0.4506	0.4587	C	0.9681	0.1428	0.7317	1.281	ALLELIC	146/178	233/275	0.05161	1	0.8203
										TREND	146/178	233/275	0.05177	1	0.82
										GENO	35/76/51	51/131/72	0.8721	2	0.6466
										DOM	111/51	182/72	0.4668	1	0.4945
										REC	35/127	51/203	0.1405	1	0.7078
Gene	**SNP**	**A1**	**Freq A**	**Freq UnA**	**A2**	**OR**	**SE**	**L95 OR**	**U95 OR**	**χ^2 TEST**	**AFF**	**U-FF**	**χ^2**	**DF**	**P**

Chrom															
	rs12456346	G	0.443	0.3735	A	1.334	0.1458	1.002	1.776	ALLELIC	140/176	189/317	3.916	1	**0.04782**
										TREND	140/176	189/317	3.686	1	0.05486
										GENO	32/76/50	40/109/104	3.964	2	0.1378
										DOM	108/50	149/104	3.716	1	0.0539
										REC	32/126	40/213	1.329	1	0.249
	rs9676158	T	0.07372	0.06073	C	1.231	0.2871	0.7012	2.161	ALLELIC	23/289	30/464	0.5252	1	0.4686
										TREND	23/289	30/464	0.5649	1	0.4523
										GENO	0/23/133	0/30/217	-	-	-
										DOM	23/133	30/217	-	-	-
										REC	0/156	0/247	-	-	-
NR1H4	rs4764980	G	0.4565	0.4961	A	0.8533	0.1428	0.645	1.129	ALLELIC	147/175	252/256	1.234	1	0.2666
Chrom 12										TREND	147/175	252/256	1.166	1	0.2803
										GENO	36/75/50	66/120/68	1.172	2	0.5566
										DOM	111/50	186/68	0.8888	1	0.3458
										REC	36/125	66/188	0.6981	1	0.4034
	rs56163822	T	0.03988	0.02988	G	1.348	0.3858	0.6331	2.872	ALLELIC	13/313	15/487	0.6045	1	0.4368
										TREND	13/313	15/487	0.5819	1	0.4456
										GENO	1/11/151	0/15/236	-	-	-
										DOM	12/151	15/236	-	-	-
										REC	1/162	0/251	-	-	-
	rs61755050	C	0.003049	0.004016	T	0.7584	1.227	0.06849	8.398	ALLELIC	1/327	2/496	0.05113	1	0.8211
										TREND	1/327	2/496	0.05132	1	0.8208
										GENO	0/1/163	0/2/247	-	-	-
										DOM	1/163	2/247	-	-	-
										REC	0/164	0/249	-	-	-
	rs1030454	G	0.1677	0.144	A	1.198	0.1972	0.8139	1.763	ALLELIC	53/263	72/428	0.8399	1	0.3594
										TREND	53/263	72/428	0.8491	1	0.3568
										GENO	5/43/110	4/64/182	-	-	-
Gene	**SNP**	**A1**	**Freq A**	**Freq UnA**	**A2**	**OR**	**SE**	**L95 OR**	**U95 OR**	**χ^2 TEST**	**AFF**	**U-FF**	**χ^2**	**DF**	**P**

Chrom															
										DOM	48/110	68/182	-	-	-
										REC	5/153	4/246	-	-	-
	rs35724aa	G	0.3926	0.3386	C	1.263	0.1472	0.9465	1.685	ALLELIC	128/198	172/336	2.519	1	0.1125
										TREND	128/198	172/336	2.37	1	0.1237
										GENO	31/66/66	29/114/111	4.671	2	0.09675
										DOM	97/66	143/111	0.4188	1	0.5175
										REC	31/132	29/225	4.657	1	**0.03093**