Urinary Metabolic Profiling of Liver Fluke-induced Cholangiocarcinoma – a Validation Study

Munirah Alsaleh
Imperial College London, St Mary’s Hospital Campus

Paiboon Sithithaworn
Khon Kaen University

Narong Khuntikeo
Khon Kaen University

Watcharin Loilome
Khon Kaen University

Puangrat Yongvanit
Khon Kaen University

Thomas Hughes
Imperial College London, St Mary’s Hospital Campus

Thomas O’Connor
Imperial College London, St Mary’s Hospital Campus

Ross H. Andrews
Imperial College London, St Mary’s Hospital Campus

Christopher A. Wadsworth
Imperial College London, St Mary’s Hospital Campus

Roger Williams
The Roger Williams Institute of Hepatology

Larry Koomson
Imperial College London, St Mary’s Hospital Campus

I. Jane Cox
The Roger Williams Institute of Hepatology

Elaine Holmes
Imperial College London, St Mary’s Hospital Campus

Simon D. Taylor-Robinson (✉️ s.taylor-robinson@imperial.ac.uk)
Imperial College London, St Mary’s Hospital Campus

Article

Keywords:
Abstract

Background

Global liquid chromatography mass spectrometry (LC-MS) profiling in a Thai population identified a urinary metabolic signature in *Opisthorchis viverrini*-induced cholangiocarcinoma (CCA), primarily characterised by disturbance in acylcarnitine, bile acid, steroid and purine metabolism. However, the detection of thousands of analytes by LC-MS in a biological sample in a single experiment potentially introduces false discovery errors. To verify these observed metabolic perturbations, a second validation dataset from the same population was profiled in similar fashion.

Methods

Reverse-phase ultra-performance liquid-chromatography mass spectrometry (RP-UPLC-MS) was utilised to acquire the global spectral profile of 98 spot urine samples (from 46 healthy volunteers and 52 CCA patients) recruited from Khon Kaen, northeast Thailand (the highest incidence of CCA globally).

Results

Metabolites were differentially expressed in the urinary profiles from CCA patients. High urinary elimination of bile acids was affected by the presence of obstructive jaundice. The urine metabolome associated with non-jaundiced CCA patients showed a distinctive pattern, similar but not identical to published studies. A panel of 10 metabolites achieved a diagnostic accuracy of 93.4% and AUC value of 98.8% (CI = 96.3%-100%) for the presence of CCA.

Conclusion

Global characterisation of CCA urinary metabolome identified several metabolites of biological interest in this validation study. Analyses of the diagnostic utility of the discriminant metabolites showed excellent diagnostic potential. Further larger scale studies are required to confirm these findings internationally, particularly in comparison to sporadic CCA, not associated with liver fluke infestation.

Introduction

Liver-fluke induced cholangiocarcinoma (CCA) is a serious health issue in the countries through which the Mekong River flows (Thailand, Myanmar, the Lao Peoples’ Democratic Republic, Cambodia and Vietnam) (1). It has been reported that there are more than 20,000 new cases of liver fluke-induced CCA detected each year just in Thailand, where infestation with the liver fluke, *Opisthorchis viverrini*, is common through the practice of eating raw, fermented or partially cooked river fish (2). However, easy, reliable screening techniques do not exist and, currently, labour-intensive stool examination for worm eggs and mass
ultrasound-screening of village populations are being undertaken in the absence of a reliable point-of-care test to detect CCA (3).

In a previous publication, we aimed to use metabonomic techniques to begin to address this situation and several metabolic derangements were found to be characteristic of cholangiocarcinoma (CCA) urinary metabotype (4). We previously studied 48 Thai subjects at high risk of infection, 41 with active *Opisthorichis viverrini* infection, 34 with periductal fibrosis and owing to the difficulty of sample collection, only 14 with CCA (4). Mass spectrometry (MS) is a key analytical platform for metabonomics; it offers sensitive and precise metabolite recovery from a given biospecimen. However, the simultaneous detection of thousands of analytes in a relatively small sample set can potentially introduce false positive findings, despite correction for multiple testing using Bonferroni or similar tests (5). As a result, rigorous biomarker testing is required prior to their clinical implementation. The biomarker discovery pipeline consists of a sequence of preclinical phases, starting with an exploratory phase in a small pilot dataset (typically 10 samples) then usually followed by a verification phase on a set of 10–50 samples. Quantifying key biomarkers (usually <10) on 100–500 samples is then recommended before the final clinical validation, ideally on 500–1000 biological samples (5).

To verify the preliminary urinary MS study on liver fluke-induced CCA in the literature (4), a larger sample set (n = 98) was collected over a 3-year period to validate the initial findings. A total of 52 urine specimens from patients with CCA were collected from the Specimen Bank of the Liver Fluke and Cholangiocarcinoma Research Center, Faculty of Medicine, Khon Kaen University. Healthy volunteers were selected from Khon Kaen as a control group for the study. To validate our previous work, the same metabolic profiling pipeline was implemented throughout from sample preparation to analytical procedures and statistical analyses (4).

An important aspect of this study was to promote the transitioning from biomarker discovery to biomarker validation. We therefore aimed to assess the reproducibility of the assay in detecting metabolic alterations associated with the urine metabolome in CCA patients.

**Methodology**

**Ethics**

Urine samples were obtained from the Specimen Bank of the Cholangiocarcinoma Research Institute, Faculty of Medicine, Khon Kaen University, Khon Kaen Province in Thailand. The study was approved by the Ethics Committee for Human Research, Khon Kaen University, (reference no. HE571283 and HE521209). Written, informed consent was obtained from each participant prior to recruitment. Ethical approval was also obtained from Imperial College London REC, London, UK (REC Reference 09/H0712/82). The study was conducted according to the principles set out in the 1975 Declaration of Helsinki.

**Sample collection**
Study samples were collected from the Isaan peoples, an ethnic community native to the North-Eastern region of Thailand. Raw, partially cooked and/or fermented fish dishes which are likely to contain the *O. viverrini* parasite are distinctive to their cultural cuisine. Patients with *O. viverrini*-induced CCA were recruited from the inpatient population in Srinagarind Hospital, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. Healthy Isaan volunteers were recruited from the staff of Srinagarind Hospital and Khon Kaen University staff and students. The study thus consisted of 98 spot urine samples, (n= 52) from participants with CCA and (n= 46) from healthy volunteers. CCA was diagnosed by CT or MRI and further confirmed by histology after surgical operation.

**Sample transport, preparation and MS analysis**

A courier service transported the samples (frozen on dry ice) from Khon Kaen University to St. Mary’s Hospital, Liver Unit, London, UK. They were kept stored at -80°C until liquid chromatography mass spectroscopy (LC-MS) analysis at Imperial College London.

**Chromatographic conditions**

LC-MS conditions, spectral pre-processing and metabolite annotation were as described in previous work (4).

The spectral profiles of samples were acquired using an ACQUITY ultra performance liquid chromatography (UPLC) system (Waters Ltd. Elstree, U.K.), coupled to a LCT Premier mass spectrometer (Waters MS Technologies, Ltd., Manchester, U.K.) (4). Reverse phase (RP)-UPLC-MS was performed with electrospray ionisation (ESI) in both positive and negative modes (4). The conditions were optimized using quality control (QC) samples in terms of peak shape, reproducibility and retention time (4).

**Tandem mass spectrometry**

Tandem mass spectrometry (MS/MS) analysis was performed using a quadrupole time-of-flight (TOF) Premier instrument (Waters MS Technologies, Manchester, UK) (4). Collision-induced dissociation (CID) experiments of the QC sample were performed for structural elucidation of detected ions in each ionisation mode (4). This was conducted subsequent to the original profiling run to save time and limit analytical variations in retention time and performance that can occur when returning to the instrument for CID analysis (4). Two complementary tandem mass spectroscopy (MS/MS) acquisition modes were used to ensure sufficient MS/MS coverage of ions of interest, data-dependent acquisition (DDA) and acquisition with no precursor ion selection, or data-independent acquisition (MS^E_) (4).

The DDA experiment was set to switch automatically from the MS to MS/MS mode using data-dependent criteria (4). It triggered MS/MS on the most abundant ions in each MS scan and provided fragments specifically attributed to the precursor ion (4).
In MS\(^E\) mode, eluting peaks were subjected to both high and low collision energies in the collision cell of the mass spectrometer, with no prior precursor ion selection (4).

**Metabolite assignment verification**

The molecular mass, retention time and fragmentation spectrum of the discriminant features were compared against on-line spectral libraries such as HMDB (www.hmdb.ca) and METLIN (https://metlin.scripps.edu). Metabolites were classified as either:

a. identified compounds confirmed with an authentic standard;

b. putatively annotated compounds (such as those based upon fragmentation pattern and/or spectral similarity with spectral databases);

c. putatively identified to match a certain chemical class (such as those based on spectral similarity to known compounds of a chemical class); or

d. as unknown compounds.

**Pre-processing**

The raw LC-MS data files were converted to CSV format by MassLynx version 4.1 application manager (Waters Corporation, Milford, U.S.A.) and then imported into R Project version 3.1.0 (The R Foundation for Statistical Computing, 2014) for pre-processing using XCMS package version 2.14. (Bioconductor). Computational scripts written in-house were applied to: 1) filter and identify peaks; 2) correct for retention time drift; 3) match peaks across samples; and 4) fill in missing peaks.

**Statistical analysis**

The spectral data matrix was imported to SIMCA-P+ version 13.0.2 (Umetrics, Umeå, Sweden) for multivariate statistical analysis and feature selection. Univariate significance tests and correlation analysis were performed in R Project version 3.1.0 (The R Foundation for Statistical Computing, 2014).

MetaboAnalyst v3.0 platform (http://www.metaboanalyst.ca/) was used to assess the clinical utility potential of the biomarkers (6). The receiver operating characteristic (ROC) curve and the area under the curve (AUC) were calculated for individual markers and for possible marker combinations. In the univariate mode, the discrimination potential of each feature was assessed by testing all possible cut-off values within all data points. The sensitivity and specificity of each marker was then calculated at the best cut-off value.

The multivariate testing of marker combinations produced ROC curves, based on the cross-validation performance of the Random Forest algorithm. Subsequently, the software produced a plot illustrating the order of the features, based on the frequency of being selected, which showed the stability of the rank of the metabolites. The confidence bounds of the AUCs were measured, based on re-sampling approach,
using bootstrapping for the univariate AUCs or Monte Carlo cross-validation for the multi-panel biomarker analysis (6).

**Results**

**Demographics, clinical data and cohort description**

A total of 100 urine samples were acquired using global LC-MS metabolic profiling. Two samples were found to be mis-identified upon receiving the subsequent clinical information. One participant had ampullary carcinoma, whereas, the medical condition of the second individual was not specified. Spectra from both individuals were excluded from the subsequent statistical modelling processes. Patients with CCA were notably older than the healthy group (mean age = 59 years vs. 31 years, respectively). Healthy volunteers were comprised mostly of women with approximately 74%, compared to 46.3% women in the CCA group.

The majority of CCA patients presented with tumours in the intrahepatic (44%) or perihilar (40%) bile ducts (Table 1). Only 15% of the CCA cohort had distal tumours. Approximately half of the cases (46.1%) presented with obstructive jaundice, particularly among patients with perihilar CCA (pCCA) and distal CCA (dCCA) tumours. None of the patients with intrahepatic tumours (iCCA) was jaundiced.

**Overview of MS results from study samples**

Good QC clustering trends in the middle of the PCA plots indicated instrument reproducibility, but CCA samples (in red) were scattered with several outlier samples outside the 95% confidence interval in both ESI modes (Fig. 1). Upon examining their raw spectral profiles, several features attributed to drug intake were present in high concentrations (Table 2).

Nine out of the top ten features that discriminated CCA patients from healthy controls using OPLS-DA were related to drug intake, predominantly antibiotics (Ceftriaxone, Metronidazole and Ofloxacin). Only one metabolite, hippurate, was endogenous and was found greater in the urine metabolome from the healthy group, Table 2. Features of drug and drug co-metabolites were excluded using correlation analysis between each feature and the metabolic data matrix.

**Cancer patients vs. healthy volunteers**

The PCA model of the urine metabolic profiles of healthy Thai participants compared to CCA patients showed clustering trend between the two groups in both ESI modes, Fig. 2a,b. The OPLS-DA statistics showed enhanced separation in both modes with excellent reproducibility (Fig. 2c, positive mode $Q^2 Y = 76\%$ and Fig. 2d, negative mode $Q^2 Y = 65\%$). The permutation test and CV-ANOVA p-value ($<0.001$) confirmed the model robustness, Fig. 2e,f.

**Impact of tumour location on the metabolic profile**
To address the impact of the anatomical location of CCA tumours on the endogenous metabolite profile, OPLS-DA models were computed using the spectral data from iCCA, pCCA and dCCA. All models projected poor classification, particularly between perihilar and distal tumours (ESI positive, $Q^2 Y = -85\%$, $p = 1.0$ and ESI negative: $Q^2 Y = -58\%$, $p = 1.0$) indicating a very similar urine metabolome. Perihilar and intrahepatic tumours also showed relatively indistinguishable molecular profiles, based on the model statistics in positive ($Q^2 Y = 4\%$ and $p = 0.770$) and negative ionisation mode ($Q^2 Y = 3\%$ and $p = 0.832$).

**Cholangiocarcinoma patients with jaundice vs. non-jaundiced patients**

To assess the impact of obstructive jaundice, a supervised model was calculated using the spectral data from jaundiced CCA patients, compared to non-jaundiced CCA patients (Fig. 3). The model was discriminant in both ESI modes (positive $Q^2 Y = 26\%$ and negative $Q^2 Y = 25\%$) (Fig. 3c,d). The CV-ANOVA p-value and the permutation test indicated reliable metabolic difference between the two groups, Fig. 3.

The discriminant metabolites generated from the OPLS-DA model of CCA patients with jaundice, compared to non-jaundiced CCA patients are listed in Table 3. A difference in urinary bile acid levels was most pronounced in patients with obstructive jaundice. For instance, the mean relative abundance (to total signal intensity) of urinary glycocholic acid was 68, 580 and 2356 in healthy, non-jaundiced and jaundiced CCA patients, respectively. Bile acids were significantly altered between jaundiced and non-jaundiced patients. An increase in hydroxylated acylcarnitine (C9-OH and C10:2-OH) and in vanilpyruvate, and a decrease in isocitrate, was associated with the urine metabotype of jaundiced-CCA participants.

**Non-jaundiced cholangiocarcinoma patients vs. healthy controls**

The difference between the metabotype of non-jaundiced CCA patients in comparison to healthy participants was examined, Fig. 4. The supervised model was highly discriminant with $Q^2 Y = 77\%$ and $Q^2 Y = 71\%$ in the positive and negative ESI modes, respectively. The differential features between for non-jaundiced CCA patients and healthy participants are summarized in Table 4. Disturbances in isocitrate, citrate and acylcarnitine metabolism were prominent. The leading 29 metabolites are shown in the table and whether the trends are upregulated or downregulated, compared to the control population.

**The diagnostic performance of candidate biomarkers**

Individual metabolites of interest are shown in Table 5. Of note, citrate and isocitrate were decreased in CCA patients with acylcarnitine (C9-OH) and acylcarnitine (C10-OH) were increased in the patient population, compared to the ethnically-matched Isaan healthy volunteers. AUROC curves for different metabolite panels were calculated. The AUC values of potential biomarker combinations yielded high values ranging from 0.942 using only 2 metabolites to 0.99 using 20 or 25 molecules. The best diagnostic accuracy of 93.4\% was achieved, with a total of 10 metabolic features.
Discussion

A total of 98 spot urine samples were used to characterise liver-fluke induced CCA signatures using a MS-based metabonomics approach. Clinical information including the anatomical location of the tumours and the presence of obstructive jaundice were used to stratify the patients and explore the effect on the detected metabolite profiles. The presence of obstructive jaundice had a clear impact on bile acid urinary elimination, whereas no obvious difference was observed, based on the tumour anatomical location. Acylcarnitine perturbation was pronounced in urine metabolic profile from CCA patients. Environmental and physiological factors such as drug intake, were found to be the most influential confounding factors.

Despite its promising potential, metabolic profiling holds some intrinsic flaws. The biological and experimental limitations associated with such a system biology approach can challenge its biological interpretation and clinical implementation. Various factors can be a source of variability in the MS data matrix produced using the training and validation samples, which are generally divided into 1) pre-analytical and 2) instrumental or analytical-related (5).

Pre-analytical causes of variability in the detected metabolic profiles can affect the sample at any point from the sample selection and collection until the sample preparation for chromatographic analysis. The stability of human urine specimens can be potentially jeopardised because of sample-related factors, such as prolonged storage and freeze thaw cycles (7). Additionally, physiological factors, including age, gender, ethnicity and disease statues, and environmental factors (such as diet, drugs, physical activity, diurnal cycles, and the host-gut microbiota metabolic interactions) can increase inter-individual biological diversity and mask the true biomarker associations relative to the disease by the addition of noise (5). For example, the presence of drug-related analytes in a sample can affect the presence of metabolic features, as a result of sample saturation and can potentially challenge the interpretation of the results.

It is therefore critical to study these non-disease-specific factors which can significantly induce a high level of noise in the urinary metabolic profiles. The pre-analytical variability can be partially controlled for by matching the participants on demographic characteristics (such as age and gender), collection of sufficient metadata, and standardizing the dietary recording and sample handling protocols.

Sources of technical variability can be accounted for by standardising the analytical pipeline, including the use of the same instrument, column type, solvent batch and instrument parameters (such as mass resolution and collision energy). Small variations between MS experiments in retention times and accurate mass measurements are expected. Furthermore, it is also possible to experience variation across the same analytical run such as retention time drift over the run. Therefore, intermittent injection of quality control samples is routinely used to assess the analytical variation in the MS system (5).

Nevertheless, it is difficult to assess precisely what influences the reproducibility and selectivity of untargeted MS signal detection. Ion suppression and/or ion-enhancement due to the endogenous compounds within the sample (such as salts, proteins and lipids) and the presence of xenobiotic substances (such as drugs) are named “matrix effects”. The suppression or enhancement response of an
analyte results when an analyte present in a sample, other than the analyte of interest compete to be ionised. The volatility of compounds in a sample is another factor that can impact the number of ions detected (8). Furthermore, the degree of droplet evaporation in the ESI source alters the amount of charged ions in the gas phase (8).

Evidence of perturbation in bile acid homoeostasis and increased urinary elimination of excess bile acids in patients with obstructive jaundice have been acknowledged in the late 1960s (9). In the current study, a dramatic increase in urinary bile acid excretion was the most pronounced feature in jaundiced-CCA patients (fold change increase from 75 to 4), compared to non-jaundiced CCA patients. Taurine- and glycine-conjugated cholic acid (taurocholic acid, glycocholic acid), conjugated glycocholic acid and glycochenodeoxycholic were significantly increased in jaundiced-CCA patients, compared to non-jaundiced patients.

To a lesser extent, the urinary excretion of acylcarnitine species was also dysregulated in jaundiced patients. Two medium-chain hydroxylated acylcarnitines, acylcarnitine (C9-OH) and acylcarnitine (C10:2-OH), were elevated in jaundiced patients compared to the non-jaundiced group urine metabolome. Greater urinary concentrations of acylcarnitine (C9-OH) have been detected in Chinese colorectal cancer patients (10) and a rise in the plasma level of acylcarnitine (C10:2-OH) was found implicated in recovery after exercise (11). Acylcarnitine implication in CCA with jaundice is uncertain, but it is possibly due to compromised liver capacity to oxidise fatty acid in clinically-icteric patients (12).

A reduction in the urinary relative abundance of citrate and its isomer, isocitrate, was generally associated with CCA patients, but, it was more pronounced in patients with biliary obstruction. This was contrary to the observation found in a recent Chinese LC-MS metabolic profiling study of urine samples from extrahepatic CCA patients, where urinary citrate levels were higher in cancer patients (13). Shao and colleagues identified general reduction in urinary citrate in presence of liver disease, both benign (cirrhosis) and malignant (HCC), but the metabolite exhibited no significant difference between the two groups (14). An observation which was confirmed by Chen and colleagues, where citrate was only significantly reduced in HCC patients with cirrhosis and hepatitis. (15). The decrease in citrate excretion in urine is possibly related to TCA cycle perturbation, resulting from mitochondria dysfunction in liver disease (16).

A compound putatively identified as vanilpyruvate, a catecholamine and phenylpyruvate derivative, was significantly associated with jaundiced patients (VIP = 3.5, p < 0.0001 and FC = 2.56). Vanilpyruvate is reported to be increased in the urinary metabolic profile of individuals with aromatic l-amino acid decarboxylase deficiency, a disorder that impairs the synthesis of serotonin, dopamine and catecholamines (17). Several other compounds were found to be differential between jaundiced and non-jaundiced CCA urine samples, but not to a level of significance. Greater abundance of creatinine, creatine, proline betaine, citrate, hippurate and indoxylsulfate was found to be associated with jaundiced patients, whereas, 4-phenylbutanic acid-O-sulfate and butyryl-L-carnitine were down-regulated, compared to the non-jaundiced group.
A multivariate statistical model was computed using the spectral data of CCA patients (excluding jaundiced participants) versus healthy controls to identify a panel of urinary metabolic markers related to CCA-genesis. Several metabolites belonging to various metabolic pathways, particularly related to acylcarnitine and steroid metabolism, were found to be perturbed. Assessment the validity of these metabolites and their relevance to hepatobiliary disease was carefully reviewed before considering these for future evaluation as potential markers in CCA.

As an example, leucyl- or isoleucyl- proline was ruled out from our diagnostic panel: the dipeptide was higher in men and it is therefore possibly influenced by protein intake. Previous MS studies on animals have shown influence of disease and diet on leucyl-proline levels. The urinary clearance of the peptide decreased in rats with atherosclerosis (18). In another study, the plasma levels of leucyl-proline decreased after 7 weeks of hypercholesterolemic diet (19). Leucyl-proline was also significantly (p < 0.001) reduced in human MS urinary profiles in bladder cancer patients (20), but was found to be higher in HCC (21). Therefore, because of the ambiguous kinetics leading to these observations, leucyl-proline was not included as one of the diagnostic markers.

Compounds classified as gut microbial co-metabolites and/or dietary phenols (hippurate, phenylacetylglutamine, indoxyl sulfate, pyrocatechol sulfate, 4-phenylbutanic acid-O-sulphate, and phenylalanine) are influenced by a number of factors other than cancer. For instance, hippurate excretion has been observed to be higher in females, as reviewed by (22).

Lees and colleagues also reported reduction in hippurate urinary clearance in obesity, diabetes, inflammatory bowel disease, parasitic infection and cancer, as observed in numerous publications investigated using a metabonomic approach (22). Analysis of the diagnostic power of these markers generally showed poor performance (AUC < 70) and/or wide confidence intervals. These markers did not contribute greatly to the multi-panel marker classification.

The AUC values and the corresponding CIs of potential biomarker combinations yielded high values ranging from 0.942 using only 2 metabolites to 0.99 using 20 or 25 molecules. The best diagnostic accuracy, 93.4%, was achieved with a total of 10 metabolic features. The performance compares favourably to CA-19, the currently used clinical marker, which has been reported to have a sensitivity and specificity of 40–70% and 50–80%, respectively in CCA patients (23).

Acylcarnitine and steroid species yielded the best classification performance and were selected in the multipanel model. Accumulation of hydroxylacylcarnitine species (C9-OH and C10-OH) was a distinctive pattern in the CCA urine metabolome. Levels of saturated (C9) and unsaturated (C8:1, C10:1, C10:2 and C10:3) long-chain acylcarnitine species were significantly reduced in CCA, particularly in patients without obstructive jaundice. An acylcarnitine, 2-trans,4-cis-decadienoylcarnitine (C10:2), was shown to be the most frequently selected metabolite, based on a repeated random sub-sampling cross validation (n = 50 runs).
Steroid-related analytes also showed excellent diagnostic utility. Four steroid metabolites, tetrahydroaldosterone-3-glucuronide, 3b,16a-dihydroxyandrostenone sulfate, pregnanediol-3-glucuronide and steroid glucuronide, were selected among the diagnostic panel. Low urinary elimination of steroids was observed in all CCA patients, regardless of jaundice. Urine steroid metabonomics studies revealed a promising diagnostic utility for distinguishing early HCC from cirrhosis (24) and adrenocortical carcinoma from adrenocortical adenomas (25). The strong male predominance of HCC in males prompts the question on the involvement of sex hormones (particularly androgens) in hepatocarcinogenesis (26).

The role of various agents in modulating biliary carcinogenesis has been thoroughly studied in experimental models (27, 28). Steroid hormones have been shown to play a role in supporting the growth of the biliary epithelium (28). The bile epithelial network possess a number of biological functions (such as secretion, absorption, proliferation, regenerative processes and signalling) that is regulated by several agents including neuropeptides, steroid hormones, cytokines and growth factors (27, 29). In the presence of cholangiopathies, the maintenance of biliary functions is greatly affected as the balance between proliferation/loss of cholangiocytes is lost (28).

Oestrogens, for example, are well known carcinogenic agents in oestrogen-responsive tissues and have been recently shown to play a role in promoting CCA cellular growth and apoptosis (28). Cholangiocytes of normal liver do not express oestrogen receptors, but in presence of biliary malignancy, oestrogen receptors were positive in >80% of cholangiocarcinoma cells (30). Oestrogens were found to favour the growth and proliferation of malignant mass by synergising the activities of growth factors, and by taking part at both receptor and post-receptor levels (31). Thus, they play a critical role in inducing neo-angiogenesis in oestrogen-sensitive cancers through the activities of vascular endothelial growth factor (VEGF) receptors, which in turn, mediate the proliferative effects of oestrogens (31). Mancinelli et al., suggested that CCA tumours could be consider oestrogen dependent and the use of anti-oestrogen drugs should control the growth of these tumours (32).

Interestingly, glycochenodeoxycholate-N-sulfate was also selected among the biomarker panel. The abundance of bile acid metabolites varied within the non-obstructive CCA group, implicating those factors other than jaundice status affected bile acid homeostasis in a subgroup of patients. Accumulation of bile acids in urine has been associated with an inflammatory signature in various forms of hepatobiliary diseases (33). Furthermore, dysregulation of bile acid metabolism was found to be implicated in biliary tract carcinogenesis by activating pathways that promote cholangiocellular proliferation and increases CCA invasiveness, such as the G protein-coupled bile acid receptor TGR5 (34).

In conclusion, the study highlights several metabolic alterations in biliary carcinoma and builds on the initial pilot study, where we studied 48 Thai subjects at high risk of infection, 41 with active _O. viverrini_ infection, 34 with periductal fibrosis and owing to the difficulty of sample collection, only 14 with CCA, all of whom were non-jaundiced (4). In the current study, pronounced bile acid urinary excretion was mostly associated with clinically icteric CCA patients. Similar to our initial study, patients without jaundice, acylcarnitine and steroid species were the most discriminant, compared to the disease-free cohort.
Targeted acylcarnitine and steroid assay could potentially be informative; it can provide a greater understanding with the benefits of higher specificity, lower potential for false positive identifications and better control for environmental confounders (such as diet and drugs). Gut microbial co-metabolites (such as hippurate and indoxyl sulfate) and bile acids have been shown to exhibit a different pattern in cancer patients. Emerging evidence has identified a distinct colonic microbial population “on” and “off” the tumour microenvironment, implicating an interplay between intestinal microbial ecology and carcinogenesis (35). Better understanding of the gut microbial-mammalian co-metabolism and the role of the oncogenic microbiome in disease initiation and progression may provide novel diagnostic and therapeutic tools.

The rare occurrence of CCA in Western countries limits biomarkers validation in large human cohorts. Nevertheless, CCA is one of the most common malignancies in Thailand and the framework for a large-scale surveillance programme is already available. Preclinical validation of a urinary prognostic and diagnostic metabolic panel could be implemented in parallel to the exciting ultrasound screening programme known as CASCAP (36).

**Abbreviations**

ANOVA = analysis of variance; BCAA = branched chain amino acids; CCA = cholangiocarcinoma; CID = collision-induced dissociation; CT = computed tomography; CV-ANOVA = ANOVA of cross-validated residuals; DDA = data-dependent acquisition; dCCA = distal cholangiocarcinoma; ESI = electrospray ionisation; ESI - = electrospray ionisation negative mode; ESI + = electro spray ionisation positive mode; iCCA = intrahepatic cholangiocarcinoma; LC-MS= liquid chromatography mass spectroscopy; MRI = magnetic resonance imaging; NMR = nuclear magnetic resonance; OPLS-DA = orthogonal projections to latent structures discriminant analysis; pCCA = perihilar cholangiocarcinoma; QC = quality control; ROC = receiver operating characteristic; RP = reverse phase; TOF = time of flight; UPLC = ultra performance liquid chromatography

**Declarations**

**Acknowledgements:** This study was funded by grants from the Wellcome Trust ISSF Fund at Imperial College London and AMMF – the Cholangiocarcinoma Charity (Stansted, Essex, UK). Munirah Alsaleh was funded by the StratiGrad PhD program at Imperial College London. Running costs were also provided by a generous donation from the friends and family of Mr and Mrs Barry Winter and of Mrs Suzy Dunn. All authors acknowledge the support of the United Kingdom National Institute for Health Research Biomedical Research Centre at Imperial College London for infrastructure support. We are most grateful to Mary M.E. Crossey, Nicola A. Cook and Yasmin Pasha for help with sample collection and storage.

**Authors’ contributions:** All authors made substantial contributions to the conception and design of the work. The acquisition of data was undertaken by T.O.'C., T.H., L.K., P.S., N.K., W.L., P.Y. and C.W. Analysis
was performed by M.A. with interpretation of data by I.J.C., P.S., N.K., E.H., R.H.A. and S.D.T.-R. M.A drafted the work with input from I.J.C., R.W., E.H. and S.D.T.-R. All authors revised it critically for important intellectual content. All authors approved the version to be published and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

**Declaration:** This paper is based on a chapter in the thesis of the first author, Munich Alsaleh - to which all authors had academic contribution or supervision. Professor Puangrat Yongvanit died during preparation of this manuscript. She contributed substantially to the work. Approval has been obtained from her legal representatives for her inclusion in this manuscript as an author. Further details are available from Professor Ross H. Andrews, Cholangiocarcinoma Research Centre, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand (rhandrews@gmail.com). Professor Roger Williams also died during manuscript preparation, but contributed substantially to the work. Approval has been obtained from his legal representatives for his inclusion in this manuscript as an author. Further details are available from Natalie Day at The Institute of Hepatology, 111 Coldharbour Lane, London SE5 9NT, United Kingdom (n.day@researchinliver.org.uk).

**Funding:** This study was funded by grants from the Wellcome Trust ISSF Fund at Imperial College London and AMMF – the Cholangiocarcinoma Charity (Stansted, Essex, UK). MA was funded by the StratiGrad PhD program at Imperial College London. Running costs were also provided by a generous donation from the friends and family of Mr and Mrs Barry Winter and of Mrs Suzy Dunn. Conflicts of interest/Competing interests (include appropriate disclosures).

**Ethics approval:** Ethical approval was obtained from Khon Kaen Human Research Ethics Committee, (references HE571283 and HE521209) and from Imperial College London REC, London, UK (REC Reference 09/H0712/82).

**Consent to participate:** Prior written, informed consent was obtained from each participant.

**Consent for publication:** Not applicable

**Availability of data and material:** Available from Larry Koomson, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, W2 INY, United Kingdom (l.koomson@imperial.ac.uk).

**Code availability** (software application or custom code): Not applicable

**References**


Tables

Tables 1 to 5 are available in the Supplementary Files section.

Figures
Figure 1

PCA scores plots of test samples and quality control (QC) samples derived from UPLC-MS urinary data (n= 98 participants and n= 11 QCs) for (a) positive and (b) negative ion mode.

(a) PCA scores plot - positive mode
$R^2_X = 31\%$ and $Q^2_X = 16\%$ from 2 PC model

(b) PCA scores plot - negative mode
$R^2_X = 21\%$ and $Q^2_X = 1\%$ from 2 PC model
Figure 2

PCA scores plots for (a) positive and (b) negative ion mode data of all CCA patients and healthy volunteers. OPLS-DA scores plots showing group separation for both (c) positive and (d) negative ion mode data and the corresponding permutation tests for (e) positive and (f) negative ion mode data.
Figure 3

PCA scores plots for (a) positive and (b) negative ion mode data of CCA patients with and without jaundice. OPLS-DA scores plots showing group separation for both (c) positive and (d) negative ion mode data and the corresponding permutation tests for (e) positive and (f) negative ion mode data.
Figure 4

PCA scores plots for (a) positive and (b) negative ion mode data of non-jaundiced cholangiocarcinoma patients and healthy controls. OPLS-DA scores plots showing group separation for both (c) positive and (d) negative ion mode data and the corresponding permutation tests for (e) positive and (f) negative ion mode data.
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 2022MunirahThaiurinevalidationWLIJCTABLESv27Apr22.pdf