Analysis of N-linked glycan alterations in tissue and serum reveals promising biomarkers for intrahepatic Cholangiocarcinoma

Shaaron Ochoa-Rios[1], Calvin R. K. Blaschke[1], Mengjun Wang[1], Kendell D Peterson[2], Andrew DelaCourt[1], Stéphane Elie Grauzam [1], David Lewin[3], Peggi Angel[1], Lewis R. Roberts[4], Richard Drake[1], and Anand S. Mehta[1]

1)Department of Cell and Molecular Pharmacology, Medical University of South Carolina, Charleston, SC, 29425
2)Department of Pediatrics, Medical University of South Carolina, Charleston, SC, 29425
3)Department of Pathology and Laboratory Medicine, Medical University of South Carolina, Charleston, SC, 29425
4)Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, 55905

Running Title: N-glycan modifications as biomarkers of intrahepatic CCA

Corresponding Author: Dr. Anand Mehta

SmartState Endowed Chair of Proteomic Biomarkers and Professor, Department of Cell and Molecular Pharmacology
Medical University of South Carolina, Basic Science Building Room 310
173 Ashley Avenue. Charleston, SC 29425
mehtaa@musc.edu; (+1) 843-792-9946 (office)

Conflict of interest

The authors declare no potential conflicts of interest

Statement of significance

This work elucidates the N-glycan alterations that occur directly in intrahepatic cholangiocarcinoma tissue and utilizes this information to discover serum biomarkers that can be used for the noninvasive detection of intrahepatic cholangiocarcinoma.

Abstract

There is an urgent need for the identification of reliable prognostic biomarkers for patients with intrahepatic cholangiocarcinoma (iCCA) and alterations in N-glycosylation have demonstrated an immense potential to be used as diagnostic strategies for many cancers, including hepatocellular carcinoma (HCC). N-glycosylation is one of the most common post-translational modifications known to be altered based on the status of the cell. N-glycan structures on glycoproteins can be modified based on the addition or removal of specific N-glycan residues, some of which have been linked to liver diseases. However, little is known concerning the N-glycan alterations that are associated with iCCA. We characterized the N-glycan modifications quantitatively and qualitatively in three cohorts, consisting of two tissue cohorts: a discovery cohort (n=104 cases) and a validation cohort (n=75), and one independent serum cohort consisting of patients with iCCA, HCC, or benign chronic liver disease (n=67). N-glycan analysis in-situ was correlated to tumor regions annotated on histopathology and revealed that bisected fucosylated N-glycan structures were specific to iCCA tumor regions. These same N-glycan modifications were significantly upregulated in iCCA tissue and serum relative to HCC and bile duct disease, including primary sclerosing cholangitis (PSC) (p<0.0001). N-glycan modifications identified in iCCA tissue and serum were used to generate an algorithm that could be used as a biomarker of iCCA. We demonstrate that

this biomarker algorithm quadrupled the sensitivity (at 90% specificity) of iCCA detection as compared to CA19-9, the current "gold standard" biomarker of CCA.

Introduction

Cholangiocarcinoma (CCA) is an epithelial cancer arising in the biliary mucosa lining the ducts that carry bile from the liver to the small intestine (1). CCA is the second most common type of liver cancer after hepatocellular carcinoma (HCC) and is a highly lethal cancer (2). CCA is classified based on the anatomic location into intrahepatic CCA (iCCA), perihilar CCA (pCCA), or distal (dCCA) subtypes (3). Cholangiocarcinomas frequently consist of small nests of epithelial cancer cells surrounded by dense stromal regions of cancer-associated fibroblasts, immune cell populations, and extracellular matrix. In addition, glandular formations are also a histologic characteristic of cholangiocarcinoma (4). CCA risk factors can be subtype-specific: Diabetes, obesity, smoking, chronic viral hepatitis, cirrhosis, chronic pancreatitis, and chemical exposure have been linked to iCCA compared to other subtypes (2,3,5). Other risk factors have been associated with geographic regions, where primary sclerosing cholangitis (PSC) is the most common risk factor in Western countries and liver fluke infection in Southeast Asia (5). Despite advances made so far in understanding the risk factors and biological mechanisms of the disease, a definitive diagnosis of CCA at an early stage continues to be challenging. For CCA diagnosis, a combination of diagnostic methods is necessary and a biopsy should be performed when feasible and taken into consideration for the final diagnosis (3). Conventional diagnostic techniques do not account for the heterogeneity of the tumor location, size, and pathological and cellular characteristics

3

(1,2). Consequently, there is an urgent need for improved biomarkers and treatments for CCA.

Glycosylation, one of the most common post-translational modifications, regulates biological functions including cell-cell communication, protein folding, and receptor signaling (1,6). Dysregulation in glycosylation has been reported in many cancers, including CCA (1,7–9). N-linked glycosylation is the addition of glycan structures to the glycoprotein at the asparagine residue. This process occurs through a well-established biosynthetic processing pathway in the endoplasmic reticulum (ER) and Golgi apparatus. The composition of these N-glycan structures is influenced by the abundance of glycosyltransferases and glycosidases, which add and remove monosaccharides, respectively, as well as the availability of nucleotide-monosaccharide donors (6,10). N-glycan alterations like fucosylation (addition of a fucose residue), sialyation (addition of a sialic acid residue) and complex branching (addition of a GlcNAc residue) have well-established associations with many different cancers, including HCC (11–14). Alterations in fucosylation can be classified as (1) core fucosylation or (2) outer-arm fucosylation. Fucosyltransferases (FUTs 1-11) catalyzes the addition of the respective fucose residue at a specific linkage of the N-glycan structure. FUT8 is the only enzyme capable of adding a fucose residue at the α1-6 linkage of the N-glycan structure (known as core fucosylation). While the remaining FUTs can add fucose residues at different linkages on the antenna of the N-glycan structure (known as outer-arm fucosylation). Due to the many biological roles, N-glycosylation is responsible for, N-glycan structures can also be altered in healthy tissue. We have previously reported N-glycan alterations like high mannose, and bi-

4

antennary N-glycans with limited fucosylation and branching in healthy liver tissue (15). N-glycosylated antigens have been widely used as biomarkers for different types of cancers: alpha-fetoprotein (AFP) for HCC, prostate-specific antigen (PSA) for prostate cancer, and Carbohydrate Antigen 19-9 (CA19-9) for CCA. However, the low specificity of CA19-9 reduces its clinical utility, and a higher-performing biomarker for CCA is urgently needed. Previous research has identified alterations in the abundance and/or N-glycosylation of certain serum or plasma glycoproteins that are correlated to liver damage caused by CCA and CCA tumor progression (1,8,9). We, and others, have previously correlated changes in glycosylation in the tissue and serum of individuals with the development of HCC (15–22). However, it is unclear if a similar change would be observed in the other major type of liver cancer, CCA.

To address this gap, we utilized matrix-assisted laser desorption ionization (MALDI) imaging mass spectrometry (IMS) for the characterization of the N-glycan-related molecular changes occurring in tissue and serum. A total of three cohorts were analyzed, consisting of two tissue cohorts: a discovery cohort (n=104 cases) and a validation cohort (n=75), and one independent serum cohort consisting of patients with iCCA, HCC, or benign chronic liver disease (n=67). Finally, we use the identified N-glycomic profiles for the development of a potential biomarker that could distinguish iCCA from other types of liver damage (23,24).

Materials and Methods

Tissues and Tissue Microarrays

Initial analysis was performed using normal liver tissue (HuFPT074), hepatocellular carcinoma (HCC) tissue (HuCAT081), and intrahepatic cholangiocarcinoma (iCCA) tissue (HuCAT086) (Biomax, Inc., Rockville, MD). H&E stains of each tissue were annotated by a pathologist with the tumor, adjacent to the tumor, and fibrotic regions. Subsequently, analysis was performed using a tissue microarray (TMA) (#LV2081, Biomax, Inc) that contained 208 cores with 103 cases (duplicated cores per case): consisting of fifty HCC, twenty iCCA, one clear cell carcinoma cyst, five metastatic HCC (spleen, chest wall, cerebrum, costal bone, and lymph node), two hepatic cyst, eight tissues with cirrhosis and dysplastic nodules, ten hepatitis infected tissues, two adjacent normal tissues, and six independent normal tissues. The cores were 1.0 mm in diameter and validated by pathology. The validation tissue cohort TMA (DLV753, US Biolab, Rockville, MD) contained 75 cases with 75 cores: forty-five cases of HCC, twenty-three cases of iCCA, two cases of mixed carcinoma, and five cases of normal liver tissue. Cores were 1.5 mm in diameter and validated by pathology. All tissue samples were formalin-fixed paraffin-embedded (FFPE) cut into a 5 µm thick section and unstained before analysis. H&E stainings of all tissues and TMAs annotated by a pathologist for small duct and large duct classifications for each core are provided in Supplementary Figure 1A-C. Clinical information for TMAs is listed in Supplementary Figure 2A-B.

Serum samples

Samples were from 30 patients with iCCA, 17 samples with primary sclerosing cholangitis (PSC), and 20 patients with other liver diseases but not iCCA or PSC seen at Mayo Clinic, Rochester, MN between January 2000, and May 2010. Peripheral blood was collected from each participant at the time of the office visit before treatment. Sera were stored at -80°C. The following data elements were abstracted from the medical record: demographics (age, gender, ethnicity, race, weight, height), medical history, etiology of liver disease, laboratory data including CA19-9 and AFP, and imaging results (US, CT, or MRI). Histopathology results and radiologic findings from the medical records of all patients were reviewed to ascertain the diagnosis of iCCA and identify tumor location. The diagnosis of iCCA in all patients was confirmed by histopathology. The anatomic location of cholangiocarcinomas was categorized as "intrahepatic" if the mass lesion arose within the hepatic parenchyma and did not extend to or involve the secondary branches of the biliary trees as demonstrated either by computerized tomography imaging, magnetic resonance imaging or endoscopic retrograde cholangiopancreatography findings. The etiology of liver disease was based on the laboratory, imaging, and histopathology results and the judgment of the treating physician. For patients with viral hepatitis, anti-HCV antibody, serum HCV RNA, HBV surface antigen, HBV e-antigen, and HBV DNA levels were recorded. Clinical information of the serum cohort is listed in Supplementary Figure 2C.

Enzymes and Reagents

Trifluoroacetic acid, Harris-modified hematoxylin, and α-cyano- 4-hydroxycinnamic acid (CHCA) were obtained from Sigma- Aldrich. HPLC grade methanol, ethanol, aceto-nitrile, xylene, hydrogen peroxide, and water were obtained from Fisher Scientific. Recombinant peptide N-glycosidase F (PNGase F) PRIME[TM] and endoglycosidase F3 Prime[TM] (Endo F3) were obtained from N-Zyme Scientifics (Doylestown, PA).

Tissue preparation for MALDI-IMS

Unstained FFPE tissue slides were processed using standardized N-glycan imaging workflows of matrix-assisted laser desorption ionization mass spectrometry imaging (MALDI-IMS) as described (15,25) Briefly, tissues were heated at 60°C for 1 hour and cooled to room temperature before deparaffinization. The slides were washed with xylene to remove the paraffin and then rehydrated using a series of water and ethanol washes. Antigen retrieval was performed using citraconic anhydride (Thermo Scientific) as the buffer and placed in a decloaker for 30 minutes. The buffer was then cooled to room temperature and buffer exchange was performed to replace the slides in 100% water. To release the N-glycans, an M5 TM-Sprayer[TM] Tissue MALDI Sample Preparation System (HTX Technologies, LLC) was used to spray 0.5 mL of 0.1 μg/μl aqueous solution PNGase F PRIME[TM] (N-zyme Scientifics) as previously described (15,26). To elucidate the core fucosylated N-glycan profile, we treated the validation TMA with the enzyme Endo F3 Prime[TM], which preferentially cleaves core-fucosylated N-glycans between the two core N-acetylglucosamine residues and results in a mass

shift of 349.137 m/z for core fucosylated N-glycans compared to N-glycans released only by PNGaseF (27). Following the spray of the respective enzyme, the slides were placed in a humidified chamber and incubated at 37°C for 2 hours. Slides were then desiccated and dried before matrix application. The matrix was assembled using α-cyano-4-hydroxycinnamic acid (0.042 g CHCA in 6 mL 50% acetonitrile/49.9% water/0.1% TFA) and sprayed using the same M5 TM-Sprayer™.

Serum preparation for MALDI-IMS

Serum samples were processed for N-glycan analysis as previously described (28). Briefly, 1 ul of serum was diluted in 2 ul of sodium bicarbonate (100Mm, pH 8.0). After mixing, 1 ul was spotted on a hydrogel-coated slide. Serum samples were spotted in triplicate and a blank well was always included in the same slide as the serum samples. Spots were left to immobilize onto the slide at room temperature for 1 hour and washed a total of three times with Carnoy's solution and one time with double distilled water. PNGase F PRIME™ and matrix were sprayed for tissue samples described above.

N-Glycan Imaging using MALDI-IMS

All tissues and serum samples were imaged using a timsTOF Flex MALDI-QTOF mass spectrometer (Bruker Daltonics) (m/z 500–4000) operating in a positive mode. Focus Pre TOF parameters were set as followed: transfer time 120.0 μs and pre-pulse storage 25.0 μs. For whole tissues (normal, HCC, and iCCA) and TMAs images were collected at 100 μm raster. For serum samples, images were collected at 150 μm raster. All images were collected at 200 laser shots per pixel.

9

Data Processing

Data analysis was done in SCiLs lab 2021 imaging software (Bruker) for analysis of the mass range m/z 500 to 4000. SCiLs-generated N-glycan spectras were normalized to the total ion count (TIC). N-glycan structure annotations for each m/z (mass to charge ratio) value were made based on an in-house database of known N-glycan generated using GlycoWorkBench and GlycoMod for annotation or to MS/MS data previously done by our group (29,30). For tissue, the maximum mean value of each m/z was extracted and used to calculate the relative intensities for each of the N-glycan peaks identified. For serum analysis, the following steps were applied to the extracted maximum mean value data: a blank well was included in the experimental plan and was subtracted from each N-glycan. Only N-glycans present in at least 80% of the samples were used for analysis. The standard sample was used to create normalization factors for each N-glycan on each slide, where the intensity of the individual slide's standard was divided by the average intensity of the standards across all slides for each N-glycan. Each slide's N-glycans were then multiplied by the corresponding normalization factor and the glycan intensities were converted to the relative intensity. A list of all N-glycans identified in tissue and serum can be found in Supplementary Table 1.

Statistical analysis

N-glycan data was explored with a boxplot and scatter plot (outliers were included). Descriptive statistics were presented as mean values +/- standard deviation unless otherwise stated. Statistical inference between two groups was applied with a t-test or

10

Mann-Whitney test based on the distribution of the data, normality of data was checked by the Shapiro-Wilk test. To explore the statistical inference of associations between outcomes with glycans, we applied logistic regression for a binary outcome, ordinal logistic regression, tree analyses, and random forest algorithm for multiple ordinal outcomes. Logistic regression was also used to derive the rate of change and corresponding p-value of each glycan change from those without iCCA to those with iCCA in both TMA and serum. Receiver operating characteristic curves (ROC curves) were built to assess the discriminating ability of individual glycans or panels of glycans. The Area under the ROC curve (AUC) was also chosen as a criterion in each step of feature selection to remove noninformative glycans or redundant glycans during optimization.

To alleviate the bias of feature selection, we applied various methods such as step-wise logistic regression, Lasso algorithm, correlation filter, random forest algorithm for feature selection to optimize the combination of glycans, and cross-validations (Leave-one-out cross-validation (LOOCV) and 3-fold CV) were further applied during feature selection to avoid bias (31). Based on the robustness, interpretability, and suitability for the glycan panel structure of interest, we chose a logistic regression model to build a predictive algorithm with selected features. The relative importance of the chosen glycans in the algorithm was derived. The performance of the predictive algorithm was explored by apparent validation (classification) and LOOCV. AUC was calculated, its standard error was derived using a bootstrap method with 2000 iterations, and the 95%CI of AUC and corresponding p-value were derived from the standard error. Statistical comparison between two AUCs was performed using Delong's test.

11

Cluster analysis was also applied to explore the similarity of the data structure of the glycans of interest based on statistical distances. Tanglegram (Cophylo plot) of serum and TMA were plotted to explore the congruence of the two dendrograms (32,33). Principal component Analysis (PCA) was used for further exploring glycan data structure based on its variation-covariance (information of each glycan), biplots of selected glycans of serum and TMA are provided for visual inspection in relationship of principal components.

Statistical analysis was performed by Graph Pad Prism 9.0 software package (Graph Pad, Inc, San Diego CA.) and R (version 4.1, https://www.r_project.org)

Study approval

Human serum samples were obtained from the Mayo Clinic Hepatobiliary Neoplasia Registry and Biorepository under an IRB-approved protocol. This study was approved by Mayo Clinic Institutional Review Board and met all guidelines set forth by the 1975 Declaration of Helsinki for good clinical practice. All participants provided written informed consent for this study.

Data Availability Statement

Data sets analyzed in this manuscript are available upon request from the corresponding author.

Results

N-glycan alterations correlate to histopathological changes

12

To elucidate the in-situ N-glycan changes that occur in iCCA tissue, we utilized MALDI-IMS methodology. Relative intensities across each tissue are presented via a heat map of individual m/z (mass to charge ratio) values, where blue is low abundance and red is high abundance. Here, m/z values are representative of specific N-glycan structures.

In Figure 1, specific N-glycans that are altered with the development of iCCA, and HCC are shown based on the histopathological changes annotated by a pathologist. N-glycans at m/z 1809.646 and 2012.717 (bi-antennary and bisected fucosylated N-glycans, respectively) were predominantly expressed in iCCA tissue, specifically in the tumor region while in the HCC tissue, these were present in only the fibrotic areas within the tumor region (Figure 1A-B, and 1E). However, an N-glycan at m/z 2393.840 (complex highly branched N-glycan) was highly expressed in HCC tissue compared to iCCA and was also expressed in the tumor region of iCCA tissue but at a lower intensity (Figure 1C and 1E). Finally, an N-glycan at m/z 1905.630 (high mannose N-glycan) did not show a definitive localization to the histology of iCCA or HCC tissues (Figures 1D and 1E). Overall, we demonstrate by MALDI-IMS that the origin of specific N-glycan modifications can change based on the type of liver cancer or the histopathological changes in each tissue.

Bisected fucosylated N-glycan alterations are specific to iCCA patients in tissue

To determine whether the N-glycan alterations previously identified in iCCA tissue could be observed in a larger set of tissue samples, we examined a discovery tissue microarray (TMA) consisting of a total of one hundred four tissue cores: fifty HCC, twenty iCCA, one clear cell carcinoma cyst, five metastatic HCC (spleen, chest wall,

cerebrum, costal bone, and lymph node), two hepatic cysts, eight tissues with cirrhosis and dysplastic nodules, ten hepatitis infected, two adjacent normal tissues, and six independent normal tissues. Figure 2A-E shows representative images from this discovery TMA (TMA 1) demonstrating the N-glycan changes from MALDI-IMS N-glycan imaging data. Specific N-glycans at m/z values of 1850.667 (presumed bisected N-glycan with one fucose residue), 1996.720 (presumed bisected N-glycan with two fucose residues), 2158.791 (presumed bisected N-glycan with two fucose residues and with an additional galactose residue), 2012.717 (presumed bisected N-glycan with one fucose residue), and 1809.646 (bi-antennary N-glycan with one fucose residue) were found primarily in iCCA tissue (Figure 2A-C and Supplementary Figure 2D-E). In contrast, N-glycan at m/z 2393.824 was specific to HCC tissue when compared to iCCA or normal samples (Figure 2D). Additionally, N-glycan at 1905.644 followed a similar trend as in Figure 1, where it was expressed in all groups with a decrease in the iCCA samples compared to other groups (Figure 2E). Normal tissue was associated with high mannose N-glycans, and bi-antennary type N-glycan with and without the terminal sialic acid, with limited levels of fucosylation and branching, as we have observed before (15). Next, we used a second TMA (TMA 2) for validation which consisted of a total of seventy-five tissue cores: forty-five HCC, twenty-three iCCA, two mixed HCC-iCCA, and five normal liver tissues (Figures 2F-J). We identified the same N-glycans in our validation TMA as in our discovery of TMA in iCCA tissues. As before, bisected fucosylated N-glycans at m/z values of 1850.667, 1996.720, 2158.791, and 2012.717 are significantly increased in iCCA as compared to the other groups (Figure 2F-H, Supplementary Figure 2D), while N-glycans at 2393.824 and 1905.644 are either not

14

elevated or under-expressed in iCCA (Figure 2I-J). Each N-glycan alteration from the two TMAs was further analyzed by comparing their relative intensity in iCCA tissue to non iCCA tissue which included normal, and other liver diseases (described in TMA 1). As figure 2K-M show, bisected fucosylated N-glycans were significantly altered in the iCCA tissue (Figure 2K-M) while complex highly branched and high mannose N-glycans were significantly decreased in iCCA tissue (Figure 2N-O). While only N-glycan changes between iCCA, HCC, and normal tissues are detailed here, Supplementary Figure 2E details N-glycan changes in other types of liver diseases present in TMA 1.

Bisected core fucosylated N-glycan alterations are specific to iCCA patients in tissue

TMA 1 and TMA 2 representative images from two different bisected N-glycans each with two fucose residues (Figure 2B, 2C, and 2G, 2H) demonstrated a higher specificity to iCCA tissues relative to Figure 2A and 2F which is also a bisected N-glycan with only one fucose residue. Based on this observation, we were interested to elucidate the origin of this fucosylation, that is core fucosylation or outer-arm fucosylation (Figure 3A). We hypothesized that the specificity of bisected double fucosylated N-glycans in iCCA tissues was most likely due to core fucosylation since FUT8 and core fucosylation have been the most common modifications reported in cancer (14). Figure 3B-D, demonstrates a higher intensity and specificity of core fucosylated N-glycans to iCCA tissues in bisected N-glycans, suggesting that the fucosylation previously observed is core fucosylation. While a fucosylated tetra-antennary branched N-glycan (at 2190 m/z) had a high intensity of core fucosylated N-glycans, this was not specific to any of the

15

groups (Figure 3E), confirming that the specificity of these N-glycan structures to iCCA samples is due to the combination of bisected and core fucosylated N-glycans.

Characterization of N-glycosylation alterations in serum of iCCA patients

To determine the translational potential of the N-glycan modifications seen in tissue to serum assays, we utilized a MALDI-IMS serum N-glycan profiling (28)**.** Figure 4A outlines the workflow used to process serum samples. Serum imaging data was analyzed based on the relative intensity of each N-glycan and used for sample comparison between non iCCA (samples from patients at risk of developing CCA or other types of liver disease including HCC, cirrhosis, hepatitis, PSC, and fatty liver diseases) and iCCA patients. Consistent with the trend observed in the discovery and validation TMAs, we found that the N-glycan bisected double fucosylated was significantly altered in this independent serum cohort (Figure 4B). Similarly, other fucosylated (in this case tri- and tetra-antennary) N-glycans (2174 m/z, 2320 m/z, and 2361 m/z) were also significantly altered in the iCCA cohort compared to non iCCA (Figure 4C-3E). In addition, we identified N-glycan at 1339.463 m/z, a non-fucosylated bisected N-glycan (known as the core structure of an N-glycan structure) to be significantly decreased in the iCCA tissues (Figure 4F). Overall, serum analysis revealed a very similar trend to what was observed in tissue, where bisected fucosylated and other fucosylated N-glycans continue to be highly altered in iCCA compared to any other group.

Association of N-glycan modifications identified in tissue to the serum of iCCA patients

16

Next, we performed a comprehensive analysis of all the N-glycan modifications observed in tissue and serum to determine the association between N-glycans found in tissue and serum. As before, we used the relative intensity of each N-glycan for the respective dataset (discovery TMA and validation TMA were analyzed and are referred to here as TMAs). Analysis of TMAs and serum revealed a total of 12 N-glycans that followed the same trend (upregulation or downregulation) with a significant *p*-value in serum, TMA, or both for some N-glycans (Table 1). Next, these N-glycans were analyzed by their data structure with clustering analysis (Figure 5A) and found that 6 specific N-glycans had a very similar data structure (Figure 5B), where most of these 6 N-glycans were fucosylated with different N-glycan types: bi-antennary, bisecting, tri-antennary and tetra-antennary N-glycans (Figure 5B). We continued our analysis using the initial 12 N-glycans identified to determine their data structure based on variability (Entropy) by principal component analysis (PCA). Figure 5C for serum and Figure 5D for TMAs, show that N-glycans profile cluster into three specific types based on the level of fucosylation (First, bi-antennary, not fucosylated; second, single fucosylated; and third, double fucosylated). Biplot of first principal component (dimension 1) (accounts for 39% of the total variation in TMAs and serum datasets) and second principal component (dimension 2) (accounts for 17% of the total variation in TMAs and serum datasets). Overall, double fucosylated N-glycans demonstrated higher informative contributions in the data set. Supplementary Figure 3A specifies the relative contribution per glycan to the first and second principal components indicated in plots with their relative contribution per N-glycan.

To identify specific N-glycans that could distinguish between iCCA and non iCCA patients,

we optimized the 12 N-glycans, this revealed three main N-glycans per data set (Supplementary Figure 3B). In the TMAs analysis, the N-glycans were at 1339 m/z, 1257 m/z (high mannose N-glycan), and 2158 m/z (Supplementary Figure 3B). For the serum analysis, the N-glycans were a 1339 m/z, 2158 m/z, and 2361 m/z (highly branched tetra-antennary, double fucosylated) (Supplementary Figure 3B). In addition, we investigated if the different types of iCCA (small duct and large duct) could express a different N-glycan profile, and found that there were no significant differences in any of these glycans, suggesting that these N-glycans are independent of the type of iCCA (Supplementary Figure 3C). From these three main N-glycans, only the two common N-glycans (1339 m/z and 2158 m/z) between tissue and serum were further analyzed (Supplementary Figure 3D). The receiver operating characteristic (ROC) curve from common N-glycans combined in serum had an AUC of 0.7656 (Figure 6A); while these same common N-glycans in the TMAs had an AUC of 0.9317 (Figure 6B). Interestingly, N-glycan 1339 m/z and 2158 m/z had an opposing trend in serum (Figure 6C) and tissue (Figure 6D), where 1339 m/z was significantly decreased in iCCA while 2158 m/z was significantly increased. This analysis suggests that the use of bisecting fucosylated N-glycans and the core N-glycan structure have a powerful discrimination ability that could be a promising differentiator strategy between iCCA and non iCCA patients.

Next, we explored further the importance of 1339 m/z as one of the differentiators between iCCA and non iCCA in serum since this N-glycan modification has not been associated with liver cancer before to our knowledge. We excluded this N-glycan from

the 12 N-glycans from our analysis and performed a feature selection from a random forest algorithm. We identified, 5 N-glycans: 2465 m/z and 2487 m/z (tri-antennary fucosylated with a sialic acid residue and same structure sulfated, respectively), 2158 m/z, 2320 m/z, and 2174 m/z. Supplementary Figure 3E shows the importance of each N-glycan selected, with all N-glycans being bisected/tri-antennary fucosylated. Combining these N-glycans to determine their ability to distinguish between iCCA and non iCCA resulted in an AUC of 1 for classification and an AUC of 0.7151 for Leave-One-Out Cross-Validation (LOOCV) (Supplementary Figure 3F). This analysis demonstrates the importance of N-glycan 1339 m/z in our model since its exclusion resulted in the need for 5 N-glycans to compensate for the use as a powerful discriminator (Supplementary Figure 3F). However, this still confirms our previous analysis in which fucosylated bisected and tri-antennary structures have important roles in iCCA.

Finally, to implement a clinical translational aspect for biomarker discovery using the N-glycan molecular changes presented here, we added CA19-9 information to our analysis. Serum CA19-9 is a biomarker used to identify those who might require diagnostic imaging for CCA. For this analysis, we focused on patients for whom we had CA19-9 information available (30 iCCA patients and 17 PSC patients). Strategies for the diagnosis of iCCA in PSC are urgently needed since early detection of iCCA can improve patients' survival and the conventional strategies share many features between conditions making diagnosis unsuccessful in most patients (3). Supplementary Table 2 lists the small capability of individual glycans (AUC: 0.500-0.727), CA19-9 (AUC: 0.512), age (AUC: 0.654), or liver enzymes (AUC:0.558-0.697) to differentiate between patients

with iCCA or PSC. However, when we combined some of the most significant N-glycans (1339, 1257, and 2158), this resulted in a significant ability of this combination to differentiate between patients with PSC to those with iCCA (AUC: 0.8431, p=<0.00001) (Figure 7A and Supplementary Figure 3G). We demonstrate this is independent of CA19-9 since including CA19-9 information in the N-glycan combination, did not significantly improve biomarker performance (AUC: 0.8472) (Figure 7B and Supplementary Figure 3G). Finally, an additional multivariate model on N-glycans 1339, 1257, and 2158 and the clinical information available for PSC and iCCA patients revealed that the ability of these N-glycan to differentiate between diagnosis is independent of the clinical information applied (Supplementary Figure 4A-C). A complete N-glycan profile quantification between the serum and tissue is presented in Supplementary Figure 5A. Overall, we demonstrate that the combination of specific N-glycan modifications can be a promising biomarker for identifying iCCA patients from those with PSC.

Discussion

Alterations in N-glycosylation have been long observed with HCC (18,21,34–36), and we and others have shown that increased levels of some fucosylated glycoproteins could be observed in the serum of patients with CCA (7–9,37). However, a study that elucidates the origin of these N-glycan modifications while also exploring the same modifications in iCCA serum as a possible biomarker has not been done before.

In this study, we identified bisecting, branching (tri-antennary), and fucosylation as specific N-glycan structure modifications in iCCA tissue and serum. These N-glycan

modifications are known to be catalyzed by the following glycosyltransferases: First, β-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase (MGAT3) for bisecting, which has been previously reported to differentiate between iCCA and HCC in serum (38). In addition, MGAT3 has been considered a malignancy suppressor where its overexpression can inhibit metastatic profiles of cancer cells (39). Second, α-1,6-mannosylglycoprotein 6-beta-N-acetylglucosaminyltransferase A (MGAT5) for branching, which has been linked to malignancy and correlates with disease progression (40). In addition, its activation in many cancers has been reported to be through the upregulation of the RAS-RAF-MAPK signaling pathway (39). Finally, another major alteration observed in the CCA patients was increased fucosylation. We and others have shown that core fucosylation, catalyzed by alpha-1,6-fucosyltransferase (FUT8), is one of the main N-glycan modifications in early liver disease like non-alcoholic fatty liver disease (NASH) and liver cancers (7,41–43). FUT8 expression has also been shown to increase as cells undergo an epithelial-mesenchymal transition (EMT) by remodeling core fucosylation on the TGF-B receptor (44). Previous studies in CCA serum have reported increased double fucosylation with tri-antennary N-glycans (7,8,37). Similarly, we demonstrate here that alterations in fucosylation are due to outer-arm and core fucosylation, but core fucosylation had a higher specificity to iCCA in tissue (Figure 3B-D).

This study elucidates N-glycan modifications that can be specific to each type of liver cancer in tissue. iCCA patients had a very specific expression of 2158 m/z (presumed bisected double fucosylated N-glycan), which was not present in HCC or normal samples (Figure 2C, 2H, and 2M). In contrast, only HCC patients expressed 2393 m/z

(tetra-antennary branched N-glycan), which was not present in iCCA and/or normal samples (Figure 2D, 2I, and 2N). This suggests N-glycosyltransferases MGAT 3 and MGAT 5 have different roles in cancer progression according to the type of liver cancer. The changes in specific N-glycan structures according to the type of cancer observed here can be explained by the opposing roles of MGAT3 and MGAT5. MGAT3-generated bisecting structures are less readily bound and modified by MGAT5, resulting in a decrease in tri- and tetra antennary structures. Accordingly, increased intensity of bisecting structures correlated with decreased intensity of tetra antennary structures in iCCA patients, but HCC patients had no changes in bisecting structures and increased highly branched structures. Future studies should investigate the exact roles of MGAT3 and MGAT5 in the context of each type of liver cancer.

An interesting N-glycan modification that has not previously been linked to iCCA was 1339 m/z (bi-antennary non-fucosylated N-glycan) which in this study had a decreased intensity in iCCA tissue and serum. Statistical analysis revealed the importance of this modification when coupled with bisected fucosylated structures as a biomarker candidate for differentiating between iCCA v non iCCA and iCCA v PSC. We hypothesize that the reduction in the expression of this N-glycan is due to this structure being modified to higher/complex N-glycan structures since this is the basic core of an N-glycan structure.

A translational point of this study was correlating tissue and serum analysis for the characterization of N-glycan-related molecular changes since most of the serum N-glycoproteins are synthesized by the hepatobiliary system and can reflect liver health. Utilizing N-glycan alterations that were correlated in tissue and serum analysis improved

biomarker detection and should be pursued for the investigation of other diseases, especially those impacting the liver. It is important to note that not all N-glycans followed the same trend from tissue to serum. We hypothesize these inconsistencies may be due to the large contribution of immunoglobulin G to the serum N-glycan profile.

Here, we utilized MALDI-IMS N-glycan imaging to identify the modifications that occur directly in iCCA patient tissue and serum samples. In our whole tissue analysis, we elucidate the histopathological origin of N-glycan structures in the iCCA and HCC tissue. Next, we analyzed two tissue TMA cohorts (a discovery and validation cohort) and an independent serum cohort. Finally, we used tissue (from discovery and validation cohorts) and serum N-glycan alterations and identified 2158 m/z (presumed bisecting double fucosylated) and 1339 m/z (bi-antennary non-fucosylated) as the most effective combination to distinguish between iCCA and non iCCA in tissue and serum. While 1339 m/z, 2158 m/z, and 1257 m/z (high mannose N-glycan) as biomarker candidates to distinguish between iCCA patients and PSC patients in serum.

Together, the data presented here suggest that the changes characterized here in tissue and serum could originate from cancer itself and prior analytical glycan tools were not able to detect these changes within the tissue. A matched serum and tissue cohort would be needed to confirm the origin of these N-glycan alterations. To conclude, we propose the use of N-glycan alterations as promising biomarkers for their ability to differentiate between intrahepatic iCCA, HCC, PSC, and benign chronic liver disease. Further studies should be done using a larger set of serum samples to validate the value of the biomarker candidates proposed here.

References

1. Sandhu DS, Roberts LR. Diagnosis and management of cholangiocarcinoma. Curr. Gastroenterol. Rep. 2008.

2. Sarcognato S, Sacchi D, Fassan M, Fabris L, Cadamuro M, Zanus G, et al. Cholangiocarcinoma. Pathologica. 2021;113:158–69.

3. Brindley PJ, Bachini M, Ilyas SI, Khan SA, Loukas A, Sirica AE, et al. Cholangiocarcinoma. Nat Rev Dis Prim. 2021;7.

4. Basturk O, Farris AB, Adsay NV. Chapter 15 - Immunohistology of the Pancreas, Biliary Tract, and Liver. In: Dabbs DJ, editor. Diagnostic Immunohistochem (Third Ed. 2011. page 541–92.

5. Alsaleh M, Leftley Z, Barbera TA, Sithithaworn P, Khuntikeo N, Loilome W, et al. Cholangiocarcinoma: A guide for the nonspecialist. Int J Gen Med. 2019;12:13–23.

6. Stanley P, Taniguchi N, Aebi M. N-Glycans. Essentials Glycobiol [Internet]. 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2017. Available from: https://www.ncbi.nlm.nih.gov/books/NBK453020/

7. Wang M, Fang M, Zhu J, Feng H, Warner E, Yi C, et al. Serum N-glycans outperform CA19-9 in diagnosis of extrahepatic cholangiocarcinoma. Electrophoresis. 2017;38:2749–56.

8. Talabnin K, Talabnin C, Ishihara M, Azadi P. Increased expression of the high-mannose M6N2 and NeuAc3H3N3M3N2F tri-antennary N-glycansin cholangiocarcinoma. Oncol Lett. 2018;15:1030–6.

9. Chang TT, Cheng JH, Tsai HW, Young KC, Hsieh SY, Ho CH. Plasma proteome plus site-specific N-glycoprofiling for hepatobiliary carcinomas. J Pathol Clin Res. 2019;5:199–212.

10. M. RJ, Esko JD. Glycosyltransferases and Glycan-Processing Enzymes. Essentials Glycobiol [Internet]. 3rd editio. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory

Press; 2017. Available from: https://www.ncbi.nlm.nih.gov/books/NBK453021/ doi: 10.1101/glycobiology.3e.006

11. Christiansen MN, Chik J, Lee L, Anugraham M, Abrahams JL, Packer NH. Cell surface protein glycosylation in cancer. Proteomics. 2014;14:525–46.

12. Arnold JN, Saldova R, Abd Hamid UM, Rudd PM. Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation. Proteomics. 2008;8:3284–93.

13. Hakomori S. Glycosylation defining cancer malignancy: New wine in an old bottle. Proc Natl Acad Sci U S A. 2002;99:10231–3.

14. Keeley TS, Yang S, Lau E. The diverse contributions of fucose linkages in cancer. Cancers (Basel). 2019;11:1–25.

15. West CA, Wang M, Herrera H, Liang H, Black A, Angel PM, et al. N-Linked Glycan Branching and Fucosylation Are Increased Directly in Hcc Tissue As Determined through in Situ Glycan Imaging. J Proteome Res. 2018;17:3454–62.

16. Blomme B, Van Steenkiste C, Callewaert N, Van Vlierberghe H. Alteration of protein glycosylation in liver diseases. J Hepatol [Internet]. European Association for the Study of the Liver; 2009;50:592–603. Available from: http://dx.doi.org/10.1016/j.jhep.2008.12.010

17. Debruyne EN, Vanderschaeghe D, Van Vlierberghe H, Vanhecke A, Callewaert N, Delanghe JR. Diagnostic value of the hemopexin n-glycan profile in hepatocellular carcinoma patients. Clin Chem. 2010;56:823–31.

18. Liu XE, Desmyter L, Gao CF, Laroy W, Dewaele S, Vanhooren V, et al. N-glycomic changes in hepatocellular carcinoma patients with liver cirrhosis induced by hepatitis B virus. Hepatology. 2007;46:1426–35.

19. Zhang Y, Zhu J, Yin H, Marrero J, Zhang XX, Lubman DM. ESI-LC-MS Method for Haptoglobin Fucosylation Analysis in Hepatocellular Carcinoma and Liver Cirrhosis. J Proteome Res. 2015;14:5388–95.
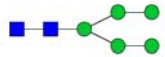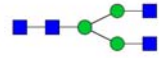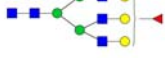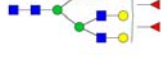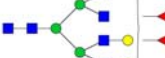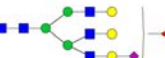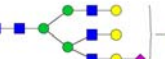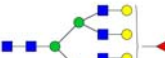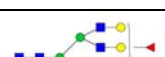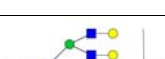
20.   Zhu J, Wu J, Yin H, Marrero J, Lubman DM. Mass spectrometric N-glycan analysis of haptoglobin from patient serum samples using a 96-well plate format. J Proteome Res. 2015;14:4932–9.

21.   Zhu J, Lin Z, Wu J, Yin H, Dai J, Feng Z, et al. Analysis of serum haptoglobin fucosylation in hepatocellular carcinoma and liver cirrhosis of different etiologies. J Proteome Res. 2014;13:2986–97.

22.   Yin H, Lin Z, Nie S, Wu J, Tan Z, Zhu J, et al. Mass-selected site-specific core-fucosylation of ceruloplasmin in alcohol-related hepatocellular carcinoma. J Proteome Res. 2014;13:2887–96.

23.   Blaschke CRK, McDowell CT, Black AP, Mehta AS, Angel PM, Drake RR. Glycan Imaging Mass Spectrometry: Progress in Developing Clinical Diagnostic Assays for Tissues, Biofluids, and Cells. Clin Lab Med. 2021;41:247–66.

24.   McDowell CT, Lu X, Mehta AS, Angel PM, Drake RR. Applications and continued evolution of glycan imaging mass spectrometry. Mass Spectrom Rev. 2021;

25.   Drake RR, Powers TW, Norris-Caneda K, Mehta AS, Angel PM. In Situ Imaging of N-Glycans by MALDI Imaging Mass Spectrometry of Fresh or Formalin-Fixed Paraffin-Embedded Tissue. Curr Protoc Protein Sci. 2018;94:1–21.

26.   Drake RR, Powers TW, Jones EE, Bruner E, Mehta AS, Angel PM. MALDI Mass Spectrometry Imaging of N-Linked Glycans in Cancer Tissues. Adv Cancer Res. 2017;134:85–116.

27.   West CA, Liang H, Drake RR, Mehta AS. New Enzymatic Approach to Distinguish Fucosylation Isomers of N-Linked Glycans in Tissues Using MALDI Imaging Mass Spectrometry. J Proteome Res. 2020;19:2989–96.

28.   Blaschke CRK, Black AP, Mehta AS, Angel PM, Drake RR. Rapid N-Glycan Profiling of Serum and Plasma by a Novel Slide-Based Imaging Mass Spectrometry Workflow. J Am Soc Mass Spectrom. 2020;31:2511–20.

29. Lütteke T, Editors MF. Glyco- informatics Methods in Molecular Biology 1273.

30. McDowell CT, Klamer Z, Hall J, West CA, Wisniewski L, Powers TW, et al. Imaging mass spectrometry and lectin analysis of n-linked glycans in carbohydrate antigen-defined pancreatic cancer tissues. Mol Cell Proteomics [Internet]. The Authors; 2021;20:100012. Available from: https://doi.org/10.1074/mcp.RA120.002256

31. Kuhn M, Johnson K. Applied Predictive Modeling. 2013.

32. de Vienne DM. Tanglegrams are misleading for visual evaluation of tree congruencele. HAL open Sci [Internet]. 2018; Available from: https://hal.archives-ouvertes.fr/hal-01807411

33. Borcard D, Gillet F, Legendre P. Numerical Ecology with R. 2018.

34. Block TM, Comunale MA, Lowman M, Steel LF, Romano PR, Fimmel C, et al. Use of targeted glycoproteomics to identify serum glycoproteins that correlate with liver cancer in woodchucks and humans. Proc Natl Acad Sci U S A. 2005;102:779–84.

35. Pj J. Structures of disease-specific serum alpha-fetoprotein isoforms. 2000;83:1330–7.

36. Johnson PJ, Poon TCW, Hjelm NM, Ho CS, Ho SKW, Welby C, et al. Glycan composition of serum alpha-fetoprotein in patients with hepatocellular carcinoma and non-seminomatous germ cell tumour. Br J Cancer. 1999;81:1188–95.

37. Betesh L, Comunale MA, Wang M, Liang H, Hafner J, Karabudak A, et al. Identification of fucosylated Fetuin-A as a potential biomarker for cholangiocarcinoma. Proteomics - Clin Appl. 2017;11:1–16.

38. Huang C, Xu X, Wang M, Xiao X, Cheng C, Ji J, et al. Serum N-glycan fingerprint helps to discriminate intrahepatic cholangiocarcinoma from hepatocellular carcinoma. Electrophoresis. 2021;1–9.

39. Pinho SS, Reis CA. Glycosylation in cancer: Mechanisms and clinical implications. Nat Rev Cancer. Nature Publishing Group; 2015;15:540–55.

40. Granovsky M, Fata J, Pawling J, Muller WJ, Khokha R, Dennis JW. Suppression of tumor

growth and metastasis in Mgat5-deficient mice. Nat Med. 2000;6:306–12.

41.   Noda K, Miyoshi E, Uozumi N, Yanagidani S, Ikeda Y, Gao CX, et al. Gene expression of α1-6 fucosyltransferase in human hepatoma tissues: A possible implication for increased fucosylation of α-fetoprotein. Hepatology. 1998;28:944–52.

42.   Cheng L, Gao S, Song X, Dong W, Zhou H, Zhao L, et al. Comprehensive N-glycan profiles of hepatocellular carcinoma reveal association of fucosylation with tumor progression and regulation of FUT8 by microRNAs. Oncotarget. 2016;7:61199–214.

43.   Ochoa-Rios S, O'Connor IP, Kent LN, Clouse JM, Hadjiyannis Y, Koivisto C, et al. Imaging Mass Spectrometry Reveals Alterations in N-linked glycosylation that are Associated with Histopathological Changes in Non-alcoholic Steatohepatitis in Mouse and Human. Mol Cell Proteomics [Internet]. The Authors; 2022;21:100225. Available from: https://doi.org/10.1016/j.mcpro.2022.100225

44.   Tu CF, Wu MY, Lin YC, Kannagi R, Yang RB. FUT8 promotes breast cancer cell invasiveness by remodeling TGF-β receptor core fucosylation. Breast Cancer Res. Breast Cancer Research; 2017;19:1–15.

Table 1. Identified N-glycans with the same trend between TMA and serum.

m/z[1]	Proposed structure[2]	Formula[3]	Rate of trend[4]	p value[5]
1257.417 m/z		Hex5HexNAc2 + 1Na	TMA: -1.2654 Serum: -1.6486	TMA: 0.000021 Serum: 0.310389
1339.464 m/z		Hex3HexNAc4 + 1Na	TMA: -10.179 Serum: -11.1094	TMA: 0.000007 Serum: 0.001252
1501.527 m/z		Hex4HexNAc4 + 1Na	TMA: -4.3148 Serum: -3.0030	TMA: 0.000715 Serum: 0.033319
2158.792 m/z		Hex5dHex2HexNAc5 + 1Na	TMA: 8.4103 Serum: 8.3182	TMA: 0.000011 Serum: 0.012718
2174.779 m/z		Hex6dHex1HexNAc5 + 1Na	TMA: 0.5213 Serum: 0.4505	TMA: 0.000137 Serum: 0.295533
2320.808 m/z		Hex6dHex2HexNAc5 + 1Na	TMA: 0.8466 Serum: 2.8109	TMA: 0.148806 Serum: 0.022350
2361.849 m/z		Hex5dHex2HexNAc6 + 1Na	TMA: 42.0811 Serum: 9.7265	TMA: 0.000008 Serum: 0.009994
2465.871 m/z		Hex6dHex1HexNAc5NeuAc1 + 1Na	TMA: 3.5604 Serum: 3.1887	TMA: 0.001474 Serum: 0.022080
2487.861 m/z		Hex6dHex1HexNAc5NeuAc1 + 2Na	TMA: 0.5366 Serum: 1.5437	TMA: 0.271611 Serum: 0.010348
2539.896 m/z		Hex7dHex1HexNAc6 + 1Na	TMA: 0.5737 Serum: 1.4326	TMA: 0.01425 Serum: 0.507898
2685.953 m/z		Hex7dHex2HexNAc6 + 1Na	TMA: 2.1893 Serum: 1.6233	TMA: 0.004647 Serum: 0.473449
2852.954 m/z		Hex7dHex1HexNAc6NeuAc1 + 2Na	TMA: 0.9399 Serum: 3.8909	TMA: 0.446394 Serum: 0.160911

[1]Observed mass-to-charge ratio (m/z) value. [2]The proposed glycan structure based upon the m/z value [3]Composition of the identified m/z value. [4]Rate of trend in TMA and serum datasets. [5]The p value in TMA and serum datasets. TMA: Tissue Microarray. For N-glycans, red triangle, fucose; blue square, N-acetylglucosamine; green circles, mannose; yellow circles, galactose.

Figure 1. Bisected and biantennary fucosylated structures are highly expressed in iCCA tumor. Representative images of the relative intensity of **A.** Bi-antennary fucosylated N-glycan (1809.646 m/z), **B.** Bisected fucosylated N-glycan (2012.717 m/z), **C.** Tetra-antennary branched N-glycan (2393.840 m/z), **D.** High mannose N-glycan (1905.630 m/z). Proposed N-glycan structures at the bottom correspond to the respective m/z value (observed mass). **E.** Hematoxylin & Eosin (H&E) staining. Tumor regions are outlined in red, normal areas are outlined in black and fibrotic regions are outlined in blue. Intrahepatic Cholangiocarcinoma (iCCA), and Hepatocellular Carcinoma (HCC). For N-glycans, red triangle, fucose; blue square, N-acetylglucosamine; green circles, mannose; yellow circles, galactose.

Figure 2. Bisected fucosylated N-glycan alterations are specific to iCCA patients in tissue– analysis in a discovery and validation tissue cohort. Representative images of the relative intensity of **A.** Bisected/tri-antennary single fucosylated N-glycan (1850.667 m/z), **B.** Double fucosylated N-glycan (1996.720 m/z), **C.** Double fucosylated with a galactose N-glycan (2158.791 m/z), **D.** Tetra-antennary branched N-glycan (2393.824 m/z) N-glycan and **E.** High mannose N-glycan (1905.644 m/z) in tissue microarray (TMA) 1. This TMA includes 2 cores per patient: 50 HCC, 20 iCCA, and 6 normal hepatic tissue. OLD: other liver diseases. Panels **F-J:** same N-glycans as in A-E but analyzed in a second independent validation tissue cohort (TMA 2). This TMA includes 1 core per patient: 45 HCC, 23 iCCA, and 5 normal hepatic tissue. Panels **K-O:** relative intensity quantification of both TMAs comparing iCCA (n=43) v non iCCA (n=108). Each point in box plots represents a patient. The mass defect used for each m/z value is based on TMA 1 run. The asterisk indicates statistical difference (Mann-Whitney, P<0.001) and error bars represent the standard deviation. Darker red colors represent a higher intensity for the specific glycan while more blue tones represent less intensity. For N-glycans, red triangle, fucose; blue square, N-acetylglucosamine; green circles, mannose; yellow circles, galactose.

31

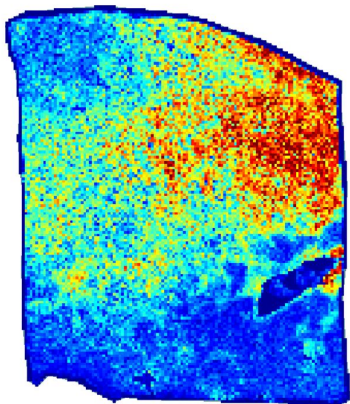**Figure 3. Bisected core fucosylated N-glycan alterations are specific to iCCA patients in tissue. A.** Cartoon description of Endo F3 (including mass shift) and PNGase F cleavage sites, and the respective fucosyltransferases (FUTs) that catalyze the addition of the fucose residue. Representative images of core fucosylated N-glycans after Endo F3 treatment on TMA2, **B.** Bisected core fucosylated N-glycan, **C.** Bisected core, and outer-arm fucosylation N-glycan**, D.** Bisected core, and outer-arm fucosylation with two galactose residues N-glycan, and **E.** Tetra-antennary core fucosylated N-glycan**.** For N-glycans, red triangle, fucose; blue square, N-acetylglucosamine; green circles, mannose; yellow circles, galactose.

Figure 4. Fucosylated N-glycans are significantly altered in iCCA serum. A. Serum N-glycan imaging workflow. Relative intensity quantification of **B.** Bisected double fucosylated N-glycan (2158.791 m/z), **C.** Tri-antennary fucosylated N-glycan (2174.779 m/z), **D.** Tri-antennary double fucosylated N-glycan (2320.808 m/z), and **E.** Tetra-antennary double fucosylated N-glycan (2361.849 m/z), **F.** Bi-antennary N-glycan (1339.463 m/z). Each point in box plots represents a patient. The asterisk indicates statistical difference (Mann-Whitney, $P<0.001$) and error bars represent the standard deviation. Non iCCA n=62 and iCCA n=30. The mass defect used for each m/z value is based on TMA 1 run. For N-glycans, red triangle, fucose; blue square, N-acetylglucosamine; green circles, mannose; yellow circles, galactose.

Figure 5. Profile data clustering reveals N-glycan grouping based on the number of fucose residues in iCCA serum and tissue. Dendrograms representing clustering for **A.**12 N-glycans and **B.** 6 N-glycans in serum and TMAs. Glycan (G), G1: 1217, G2: 1339, G3:1501, G4:2158, G5:2174, G6:2320, G7:2631, G8:2465, G9:2487, G10:2539, G11:2685, G12:2852. Principal Component Analysis (PCA) analysis and their dimension (Dim) of **C.** Serum and **D.** TMAs showing N-glycan clustering based on the number of fucose residues. 2Na (doubly sodiated N-glycan). For N-glycans, red triangle, fucose; blue square, N-acetylglucosamine; green circles, mannose; yellow circles, galactose.

32

Figure 6. N-glycans from serum and TMA as promising biomarkers to differentiate iCCA from other liver diseases. A. ROC (Receiving Operator Characteristic) curve for serum. Non iCCA n=62 and iCCA n=30 **B.** ROC curve for TMA, non iCCA n=108 and iCCA n=43. For ROC curve labeling: N-glycan 1339m/z (green-dashed line), N-glycan 2158m/z (blue-dashed line), and a combination of N-glycans 1339m/z and 2158m/z (red-solid line). **C.** Relative intensity quantification boxplots of N-glycan 1339m/z and 2158m/z in serum, non iCCA n=62, and iCCA n=30 **D.** Representative images of TMA 1 showing the N-glycan intensity for 1339m/z and 2158m/z. The asterisk indicates statistical difference (Mann-Whitney, $P < 0.001$) and error bars represent the standard deviation. For N-glycans, red triangle, fucose; blue square, N-acetylglucosamine; green circles, mannose; yellow circles, galactose.

Figure 7. N-glycan combinations as promising biomarkers to differentiate iCCA from PSC. A. ROC (Receiving Operator Characteristic) curve classification of a combination of 1339m/z, 2158m/z, and 1257m/z. **B.** ROC curve of combination of 1339m/z, 2158m/z, 1257m/z, and CA19-9. For ROC curves: iCCA (n=30) and PSC (n=17). For N-glycans, red triangle, fucose; blue square, N-acetylglucosamine; green circles, mannose; yellow circles, galactose.

Supplementary information

Additional supporting information is available in the online version of the article.

1

iCCA

HCC
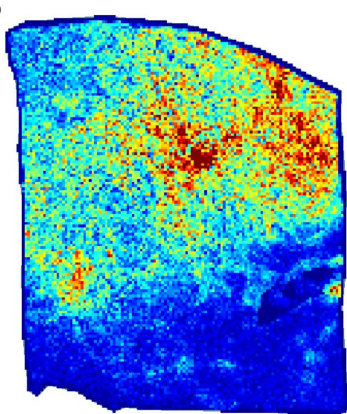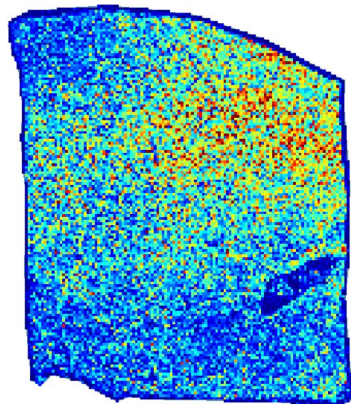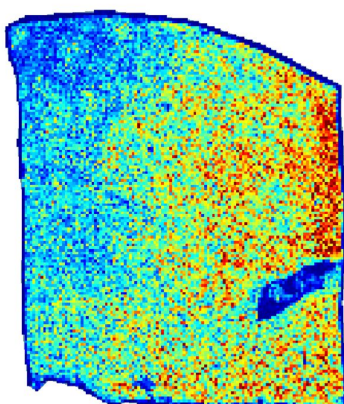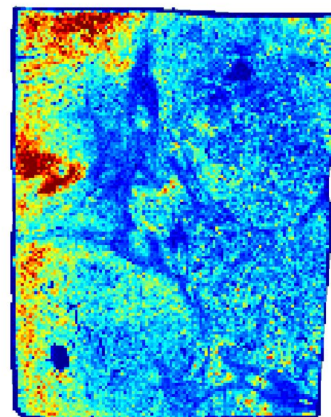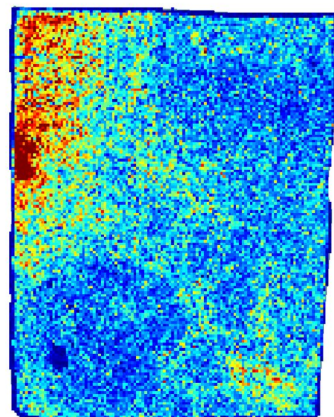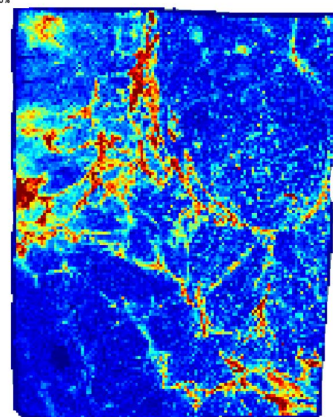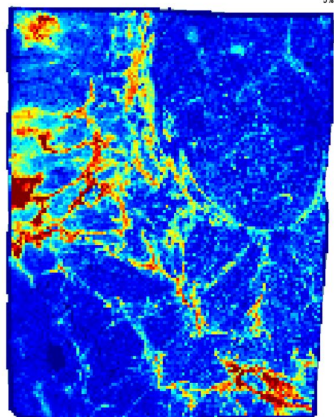
Proposed structure:
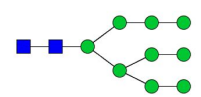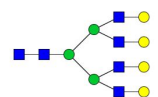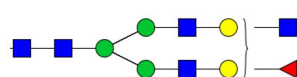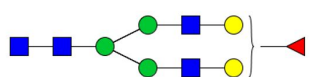
Observed mass:

A	B	C	D
1809.646 m/z	2012.717 m/z	2393.840 m/z	1905.630 m/z

E

H&E

2

A — TMA 1 — 1850.667 m/z
B — 1996.720 m/z
C — 2158.791 m/z
D — 2393.824 m/z
E — 1905.644 m/z

F — TMA 2 — 1850.667 m/z
G — 1996.720 m/z
H — 2158.791 m/z
I — 2393.824 m/z
J — 1905.644 m/z

K, L, M, N, O — Relative Intensity: Non iCCA vs iCCA

HCC, iCCA, OLD, Normal

3

A

FUTs 3, 4, 5, 6, 7, 9
Outer-arm Fucosylation

FUT 8
Core Fucosylation

Endo F3 Cleavage Site
Δ349.115 m/z

PNGase F Cleavage Site

Asn

Glycoprotein

B
Proposed structure:
Observed mass: 1501.527 m/z
iCCA
0%   100%   1501.527

C
Proposed structure:
Observed mass: 1647.588 m/z
iCCA
0%   100%   1647.588

D
Proposed structure:
Observed mass: 1809.653 m/z
iCCA
0%   100%   1809.653

E
Proposed structure:
Observed mass: 2190.784 m/z
iCCA
0%   100%   2190.784

4

A

Dilute serum with NaHCO₃

Spot/Immobilize on to an Amine-Reactive Slide

Wash with Carnoy's Solution and Water

Spray PNGase F Prime—Incubate and spray MALDI Matrix

Profile by MALDI-IMS

B
Proposed structure:
Observed mass: 2158.791 m/z

C
Proposed structure:
Observed mass: 2174.779 m/z

D
Proposed structure:
Observed mass: 2320.808 m/z

E
Proposed structure:
Observed mass: 2361.849 m/z

F
Proposed structure:
Observed mass: 1339.463 m/z

5

6

A

Serum, Non iCCA v iCCA

Glycan1339,AUC=0.7126
Glycan2158,AUC=0.6839
Glycan1339+Glycan2158,AUC=0.7656

B

TMAs, Non iCCA v iCCA

Glycan1339,AUC=0.8024
Glycan2158,AUC=0.9059
Glycan1339+Glycan2158,AUC=0.9317

C

Serum

1339m/z

2158m/z

D

TMA

1339m/z

2158m/z

A

B