**Serum *N*-glycans outperform CA19-9 in diagnosis of extrahepatic cholangiocarcinoma**

*Mengmeng Wang[1, 2, #], Meng Fang[1, #], Jianhui Zhu[2], Huijuan Feng[1], Elisa Warner[2, 3], Changhong Yi[1], Jun Ji[1], Xing Gu[1], Chunfang Gao[1,*]*

[1]*Department of Laboratory Medicine, Eastern Hepatobiliary Hospital, Second Military Medical University, Shanghai 200438, P. R. China*

[2]*Department of Surgery, University of Michigan Medical Center, Ann Arbor, MI 48109, USA*

[3]*Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI, 48109, USA*

**\*Corresponding author:** Professor Chunfang Gao, **E-mail:** gaocf1115@163.com; **Fax:** +86-21-65562400

[#]These authors contributed equally to this work.

Total number of words: 5,484

**Abbreviations**

AFP, α-fetoprotein; AFP-L3, core-fucosylated AFP; ALB, albumin; ALT, alanine aminotransferase; APTS, 8-aminonaphtalene-1, 3, 6-trisulphonic acid; AUC, area under the curve; BBD, benign biliary disease; CA125, carbohydrate antigen 125; CA19-9, carbohydrate antigen 19-9; CA72-4, carbohydrate antigen 72-4; CCA, cholangiocarcinoma; CE, capillary electrophoresis; CEA, carcinoembryonic antigen; DMSO, dimethyl sulfoxide; DSA-FACE, DNA sequencer-assisted fluorophore-assisted capillary electrophoresis; ECCA, extrahepatic cholangiocarcinoma; GLU, glucose; HC, healthy

control; HCC, hepatocellular carcinoma; MS, mass spectrometry; NaBH$_3$CN, sodium

cyanoborohydride; NPV, negative predictive value; PLT, platelet; PNGaseF, peptide *N*-glycosidase-F;

PPV, positive predictive value; PSA, prostate-specific antigen; ROC, receiver operating characteristic;

TBIL, total bilirubin; TP, total protein.

**Abstract**

Extensive efforts have been devoted to improve the diagnosis of extrahepatic cholangiocarcinoma

(ECCA) due to its silent clinical character and lack of effective diagnostic biomarkers. Specific

alterations in *N*-glycosylation of glycoproteins are considered a key component in cancer progression,

which can serve as a distinct molecular signature for cancer detection. This study aims to find potential

serum *N*-glycan markers for ECCA. In total, 255 serum samples from patients with ECCA (n=106),

benign bile tract disease (BBD, n=60) and healthy control (HC, n=89) were recruited. Only 2μl of

serum from individual patients was used in this assay where the *N*-glycome of serum glycoproteins was

profiled by DNA sequencer-assisted fluorophore-assisted capillary electrophoresis (DSA-FACE)

technology. Multi-parameter models were constructed by combining the *N*-glycans and carbohydrate

antigen 19-9 (CA19-9) which is currently used clinically. Quantitative analyses showed that among 13

*N*-glycan structures, the bifucosylated triantennary *N*-glycan (peak10, NA3F2) presented the best

diagnostic performance for distinguishing ECCA from BBD and HC. Two diagnostic models

(Glycotest1 and Glycotest2) performed better than single *N*-glycan or CA19-9. Additionally, two

*N*-glycan structures (peak9, NA3Fb; peak12, NA4Fb) were tightly related to lymph node metastasis in

ECCA patients. In conclusion, sera of ECCA showed relatively specific *N*-glycome profiling patterns.

Serum *N*-glycan markers and models are novel, valuable and noninvasive alternatives in ECCA

diagnosis and progression monitoring.

*Keywords:* biomarker/ diagnosis/ extrahepatic cholangiocarcinoma/ *N*-glycans

## 1. Introduction

Cholangiocarcinoma (CCA) accounts for about 3% gastrointestinal malignancies. It arises from the epithelium of either intrahepatic or extrahepatic bile ducts and can be divided into intra- (ICCA) or extra-hepatic CCA (ECCA) anatomically [1]. Approximate 80-90% of CCA were ECCA and could be further subdivided into perihilar CCA or distal CCA, separated by the insertion of the cystic duct [2]. A lot of risk factors including primary sclerosing cholangitis, bile duct abnormalities, infection, metabolic disorders and genetic susceptibility have been reported to be associated with ECCA, but no specific predisposing factors have been identified up to now [3, 4]. ECCA is a relatively uncommon but highly malignant cancer with the incidence of 0.82/100,000 in the United States [5]. At present, complete resection remains the most effective and only potentially curative treatment for ECCA, but most patients present with advanced unresectable disease mainly due to the lack of reliable serum biomarkers [6]. In spite of multiple studies on novel molecules for diagnosis in ECCA, the progress in clinical translation has so far been limited. The overall prognosis of ECCA patients is reportedly dismal with the median survival less than 24 months and once diagnosed, few cases could survive for more than 5 years [5, 7]. Thus, more precise markers with high sensitivity and specificity for ECCA are in urgent need which will benefit the patients and expedite the discovery of novel therapeutic strategies.

Glycosylation, one of the most common co- and post-translational modifications, has been reported to play vital roles in a variety of physiopathological conditions, including intercellular adhesion, cell migration, cell-cell signaling, host-microbial interaction and inflammation [8-10]. Accumulated evidences reveal that aberrant glycosylation is associated with the oncogenesis and progression of various cancers and this modification is the characteristic of serum glycoproteins in cancer patients [11-13]. As a result, altered glycans related to a specific cancer would make these glycans or their carriers potential tumor biomarkers [14-18]. In fact, most cancer biomarkers applied to clinical care are glycoproteins bearing *N*-glycosylation modifications such as prostate-specific antigen (PSA) and alpha-fetoprotein (AFP) [19, 20]. AFP is a commonly used tumor biomarker for hepatocellular carcinoma (HCC), and the result of aberrant *N*-glycosylation of AFP, core-fucosylated

AFP (AFP-L3) is more specific than AFP and is negative for most benign liver diseases [20, 21]. Approved by the FDA as a clinical biomarker for HCC in 2005, AFP-L3 might trigger a new era of glycan or glycoprotein biomarker discovery [22, 23].

Numerous techniques and methods have been established to identify and quantify variation of *N*-glycans in glycoproteins [24, 25]. However, most of these approaches involve chromatographic separation, mass spectrometry analysis or various microarrays, along with high complexity and low cost-efficiency [26-28]. Herein, we employ a modified DNA sequencer-assisted fluorophore-assisted capillary electrophoresis (DSA-FACE) technology, which is originally established by Callewaert and coworkers to decipher the global *N*-glycome of serum [29-31]. Discovery of differentially expressed glycans from whole serum may serve to present candidate biomarkers for the discrimination of ECCA.

## 2. Materials and methods

### 2.1 Case selection and blood samples

A total of 166 patients with ECCA (n=106) and benign bile tract disease (BBD, n=60) who underwent surgical resection at Eastern Hepatobiliary Hospital (EHBH), Shanghai from 2010 to 2012 were enrolled in our study. The histologic slides were retrieved and the diagnosis for each case was confirmed by two independent pathologists. Clinical staging is based on the TNM (tumor-node-metastasis) staging systems of the American Joint Committee on Cancer (AJCC, 7th edition) [32]. Demographic, clinical information and histologic characteristics for each patient were obtained from electronic medical record. For a healthy control (HC) group, 89 healthy volunteers who visited the same hospital for a regular physical examination were enrolled. We defined a healthy individual as someone who was deemed free of disease (including no history of cancer) at health check-up. Blood was collected using a standard protocol and serum samples were separated by centrifuging at 3000 rpm for 10 min, and then stored at −80°C. The study protocol was approved by the Chinese Ethics Committee of Human Resources, EHBH. Informed consents were obtained from the

patients and the healthy donors in advance. The main clinical and biochemical data of the study population were summarized in Table 1.

## 2.2 Laboratory tests and clinical information

Routine biochemical tests including total bilirubin (TBIL), total protein (TP), albumin (ALB), alanine aminotransferase (ALT) and glucose (GLU) were measured using standard methods and respective reagents recommended by the manufacturer (Hitachi 7600 Analyzer, Hitachi, Japan; Wako Diagnostics reagents, Wako Pure Chemical Industries Ltd., Japan). Tumor markers such as CA19-9, carcinoembryonic antigen (CEA), alpha-fetoprotein (AFP), carbohydrate antigen 72-4 (CA72-4) and carbohydrate antigen 125 (CA125) were determined on Roche E170 modular with matched reagents (Roche E170, Germany).

## 2.3 Serum protein *N*-glyome profiling

Serum protein *N*-glycan analysis was performed using DSA-FACE technology as described previously [33, 34]. Briefly, the *N*-glycans present on glycoproteins in 2μl of serum were released with peptide *N*-glycosidase-F (PNGaseF) (New England Biolabs, Boston, MA). Afterwards, the dried *N*-glycans were labeled with 8-aminonaphtalene-1, 3, 6-trisulphonic acid (APTS) (Invitrogen, Carlsbad, CA) using 2μl labeling solution (equal mixture of 20mM APTS in 1.2M of citric acid and 1M of $NaBH_3CN$ in DMSO) and incubated for 16 hours at 37°C. Sialic acid was then removed with 3μl neuraminidase/sialidase (Roche Bioscience, Palo Alto, CA) at a final concentration of 0.2mU in 5mM ammonium acetate and incubated at 37°C overnight. Finally, the processed *N*-glycans were separated using 3500 Series Genetic Analyzer (Applied Biosystems, Foster city, CA) and the peaks were analyzed via the GeneMapper v4.1 software (Applied Biosystems, Foster city, CA). The abundance of each *N*-glycan peak was quantified by normalizing its height to the sum of the heights of all 13 peaks based on a previous method which has been approved to relatively quantitate abundance [31, 35].

## 2.4 Statistical Analysis

All quantitative variables were expressed as median, (centile 25, centile 75) unless otherwise stated, categorical variables were expressed as number (percentage). Quantitative variables were compared with Student t test, ANOVA or nonparametric test. Pearson coefficients of correlation (spearman coefficients of correlation were calculated for ordinal categorical variables) and their associated probabilities ($p$) were used to evaluate the relationship between parameters. Forward stepwise logistic regression analysis was performed to construct the diagnostic multi-parameter models. The receiver operating characteristic (ROC) curve analysis was carried out and the area under curve (AUC) was presented to evaluate the diagnostic efficacy of each single marker or multi-parameter model. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy were calculated using cut-off values optimally determined upon the ROC curves. All reported $p$ values were 2-tailed, and a $p$ value less than 0.05 was considered statistically significant. Statistical analysis was performed with SPSS 21.0 for Windows software (IBM Corporation, Armonk, NY).

## 3. Results

### 3.1 Different profiling patterns and diagnostic powers of *N*-glycome in ECCA, BBD and HC group

By DSA-FACE, the *N*-glycan profiling of desialylated sera from subjects in the ECCA (n=106), BBD (n=60), and HC (n=89) groups were quantitatively examined and statistically compared. Thirteen *N*-glycan structures (peaks) were identified in the samples of each group (Figure 1). Structural analysis of these peaks was performed by correlating the migration time of *N*-glycans to those of dextran ladders as previously reported [31, 35]. The average relative abundance of these *N*-glycans is summarized in Table 2. Different *N*-glycan structures of serum total glycoprotein in three groups suggested that different *N*-glycan patterns appeared in specific pathophysiologic conditions and indicated the possibility of *N*-glycans as diagnostic markers. Compared to the HC group, peak1 (NGA2F), peak2 (NGA2FB), peak4 (NG1A2F), peak9 (NA3Fb), peak9' (NA3F), peak10 and peak12 (NA4Fb) were elevated ($p<0.05$), whereas the other peaks (except peak7, NA2FB) were decreased in the ECCA group ($p<0.05$). When comparing the *N*-glycome of the ECCA and BBD group, the abundance of peak3

(NG1A2F), peak7 and peak11 (NA4) were decreased in the ECCA group while the abundance of peak9, peak9', peak10, peak11 and peak12 were increased in ECCA patients.

Diagnostic powers of *N*-glycans for differentiating ECCA from HC and BBD were calculated by ROC curves and AUC analysis. In ROC analysis, peak9, peak9', peak10 and peak12 performed effectively in both diagnosis and differential diagnosis for ECCA (Supplementary Figure 1; AUC>0.7 was recognized as the threshold value). Peak10, a bifucosylated triantennary structure, showed the best diagnostic power among all 13 *N*-glycan peaks to distinguish ECCA from HC with the AUC of 0.948 (95% Confidence interval (95% CI): 0.918-0.978; Figure 2A). In addition, peak10 also revealed the best power to distinguish ECCA from BBD patients and the AUC was 0.909 (95% CI: 0.864-0.955; Figure 2B)

**3.2 Serum *N*-glycans were correlated with clinical and pathological characteristics.**

Apart from the diagnostic value of *N*-glycan based markers, the correlation between *N*-glycans and clinical laboratory testing data or pathological characteristics was also explored. The coefficients of correlation and corresponding *p* value were listed in Supplementary Table 1. Lymph nodes status is a strong determinant of clinical outcomes for patients with various gastrointestinal cancers [36, 37]. Intriguingly, we noticed that a branching fucosylated triantennary *N*-glycan (peak9, AUC=0.845) and a branching fucosylated tetra-antennary *N*-glycan (peak12, AUC=0.761) were positively correlated with lymph node involvement (*p*<0.05), whereas CA19-9, the commonly used marker, was not related to this pathological feature. Although both peak9 and peak12 were less helpful than CA19-9 in diagnosis (Supplementary Figure 1), they were more efficient than CA19-9 (AUC=0.660) in predicting lymph node involvement (Figure 3), indicating that *N*-glycans might play a critical role in progression of ECCA and can be used for prognosis assessment.

**3.3 Construction and assessment of diagnostic models based on *N*-glycan markers for ECCA**

To date, CA19-9 was the most frequent and effective tumor marker for identifying ECCA in routine clinical screening and tests, so we compared $N$-glycan markers with CA19-9. Based on ROC curves in Figure 2, the optimum cut-off value of peak10 was 0.38 (accuracy 88.2%) and 0.49 (accuracy 79.5%) when used to identify ECCA from HC and BBD, respectively. As shown in Table 3 and Table 4, peak10 performed better than CA19-9 in sensitivity (86.8% vs 83.0%) but not in specificity (90.0% vs 100.0%) to distinguish ECCA patients from normal controls. However, when used to separate ECCA from BBD patients, peak10 showed higher specificity (90.0% vs 83.3%) but lower sensitivity (73.6% vs 83.0%) than CA19-9.

To discover ideal diagnostic efficiencies, mathematical diagnostic models were constructed by combining $N$-glycan markers and CA19-9. Logistic regression coefficients were used to estimate odds ratios (OR) for each of the independent variables. Eventually, a mathematic formula named Glycotest1 was constructed to separate ECCA patients from normal controls (Glycotest1=0.089*CA19-9+11.444*peak10-0.331*peak5-0.540*peak6+17.058). To assess the contribution of Glycotest1 in differentiating ECCA, we determined the areas under the ROC curves. Compared with CA19-9 (AUC=0.918), Glycotest1 was more effective in identifying ECCA from normal controls at the cut-off value of 1.83 (AUC=0.992, Figure 2A). Table 3 listed the sensitivity, specificity, PPV, NPV, and accuracy of Glycotest1 for predicting ECCA. The diagnostic accuracy of Glycotest1 was 94.4%, whereas 90.8% of single CA19-9. Similarly, another mathematical formula named Glycotest2 was established to differentiate ECCA from BBD (Glycotest2 =0.009*CA19-9+9.112*peak10-4.370), and the diagnostic parameters of Glycotest2 were given in Table 4. When using the optimum diagnostic cut-off value of 0.56 based on ROC analysis (Figure 2B), Glycotest2 improved the diagnostic accuracy of CA19-9 (83.1%) and peak10 (79.5%) to 88.0%.

**Follow-up of changes of $N$-glycan markers and CA19-9 of ECCA patients before and after surgery**

After surgery, ECCA patients were followed up and a blood sample was obtained 7-14 days after treatment. The abundance of structures in peak1, 3, 4, 11 and 12 increased significantly after surgery versus before surgery by the DSA-FACE analysis (Supplementary Figure 2). Conversely, the postoperative abundance of peak2, 7 and 8 (NA3) was lower than those of preoperative samples. There was no significant difference in serum CA19-9 levels before and after surgery. These results presented the clues again that *N*-glycan markers might be used to monitor the progression and predict the prognosis of ECCA.

**Discussion**

ECCA is a rare but challenging cancer with poor prognosis. Early diagnosis plays a pivotal role in treating and managing ECCA cases [38, 39]. CA 19-9 is the most commonly used ECCA marker, but it does not provide sufficient sensitivity and reliability for early detection [1, 2]. As most other cancer biomarkers in use today, CA19-9 is a glycoprotein and measured immunochemically using monoclonal antibodies. However, the epitope for these antibodies were mostly designed against the protein moiety and not towards the *N*-glycan structures [17]. To date, *N*-glycan based biomarkers or models have been developed to help validate diagnosis and evaluate outcome for various cancers or other illnesses including liver cancer [33], rectal carcinoma [34, 40], ovarian cancer [41] and lung cancer [42]. Some reviews have highlighted the promising prospect of *N*-glycans as cancer biomarkers and the progress in glycomic research strategies recently [14, 18, 19]. Therefore, we believe that *N*-glycans could be more effective than individual glycosylated molecules for detecting ECCA occurrence and development.

In current research, a simple and practical *N*-glycan analysis technology, DSA-FACE, was adopted for measuring *N*-glycosylation changes in serum from ECCA patients, BBD patients, and healthy volunteers. We tried to identify the specific *N*-glycan profiling patterns occurring in the process of carcinogenesis and progression of ECCA. As shown in the results, differentially-expressed *N*-glycan structures of serum glycoprotein in three groups were found (Figure 1, Table 2). In order to determine

diagnostic value, we compared the *N*-glycosylation characteristics of ECCA with HC and BBD separately. A bifucosylated triantennary *N*-glycan (peak10) performed best among all 13 peaks in diagnosis and differential diagnosis for ECCA and acted better in some aspects than CA19-9 (Figure 2, Table 3&4). Previously, this *N*-glycan structure has also been reported to be a potential biomarker for liver cancer [23] and pancreatic cancer [43, 44]. To enhance diagnostic accuracy, *N*-glycans were combined with CA19-9 using mathematical models to identify ECCA from BBD patients and the healthy controls. Eventually, the models Glycotest1 and Glycotest2 were screened out and remarkably improved the accuracy of the single-marker index (CA19-9 or peak10) with AUC of 0.992 and 0.938, respectively (Figure 2, Table 3&4). Thus, both combined-marker methods demonstrated higher accuracy in discriminating ECCA from BBD than any individual marker alone.

As is well-known, assessment of nodal status might harbor important prognostic information and guided adjuvant treatment [36, 37]. Correlation analyses and ROC curves indicated that a branching fucosylated triantennary *N*-glycan (peak9) and branching fucosylated tetra-antennary *N*-glycan (peak12) were closely related with lymph metastasis. Conversely, CA19-9 could not reflect the status of regional lymph nodes (Figure 3). These branching/outer-arm fucosylation structures have been reported to be associated with ovarian cancer [41] and liver cancer [45, 46]. The findings implied the participation of *N*-glycans in oncogenesis and progression of ECCA.

In exploring the dynamic monitoring value of *N*-glycan markers, peaks of *N*-glycans and levels of CA19-9 in ECCA were compared before and after surgical treatment (Supplementary Figure 2). Changes of the abundance of some *N*-glycans pre- and post-operation proved the tumor specificity of certain structures and could be used in therapeutic observation, recurrence monitoring, and prognosis evaluation.

In this research, relatively specific *N*-glycan patterns of ECCA referring to BBD and normal individuals were identified by an efficient approach. This global *N*-glycomic strategy aimed to release and analyze the whole *N*-glycome from serum glycoproteins using enzymatic means without a pre-processing step for depletion of high-abundance proteins in serum, which avoids sample loss. We

observed the diagnostic value of *N*-glycans for ECCA and constructed two formulas to enhance the accuracy. The relationships between *N*-glycan markers and clinicopathological parameters, especially lymph node involvement suggested the potential of *N*-glycans as the monitoring indicators and underlying therapeutic targets. However, it should be noted that some limitations still existed in our study. Limited to the strategy and instrument, a relatively low number of *N*-glycans were determined and further differentiation of the specific isomers could not be achieved without combining capillary electrophoresis (CE) with mass spectrometry (MS) or MS/MS. However, the DSA-FACE method remains an effective glycomic approach for the discovery of cancer biomarkers. Differential biomarkers of clinical application potential can be identified with simple structural information using this method with higher repeatability and better quality control than the original method, as previously reported [30, 33]. Aberrant sialylation has been implicated in several cancers [47, 48]. However, desialylation reaction is adopted prior to the analysis, because sialic acids add negative charge on the *N*-glycans, which is a limitation in many CE-based methods. On the other hand, sialic acids add to the enormous heterogeneity of the N-glycans, and removal of sialic acids may thus add to the readability in complex mixtures, especially for N-glycans cleavaged from serum glycoproteins. Some groups have reported their strategies and methods for analyzing sialylated *N*-glycan linkage isomers, and these methods could be applied to discover the potential biomarkers with sialic linkage specific manner [49-51]. Further research for protein-specific or site-specific glycan markers, including sialic acid information may be carried out in the future.

**Conclusion**

*N*-glycan profiles of ECCA differ from the profiles of benign diseases and healthy controls. *N*-glycan based biomarkers and models are promising, noninvasive serum diagnostic tools and valuable

supplements to the serologic biomarkers already in use. Future in-depth study may investigate the prognostic and therapeutic value of *N*-glycans for ECCA and the mechanisms involved in pathogenesis.

**Conflict of Interest Statement**

The authors have declared no conflict of interest.

**References**

[1] Rizvi, S., Gores, G. J., *Gastroenterology* 2013, *145*, 1215-1229.

[2] Blechacz, B., Gores, G. J., *Hepatology* 2008, *48*, 308-321.

[3] Khan, S. A., Davidson, B. R., Goldin, R. D., Heaton, N., Karani, J., Pereira, S. P., Rosenberg, W. M., Tait, P., Taylor-Robinson, S. D., Thillainayagam, A. V., Thomas, H. C., Wasan, H., *Gut* 2012, *61*, 1657-1669.

[4] Razumilava, N., Gores, G. J., *Lancet* 2014, *383*, 2168-2179.

[5] Segura-Lopez, F. K., Aviles-Jimenez, F., Guitron-Cantu, A., Valdez-Salazar, H. A., Leon-Carballo, S., Guerrero-Perez, L., Fox, J. G., Torres, J., *Helicobacter* 2015, *20*, 223-230.

[6] Ben-Josef, E., Guthrie, K. A., El-Khoueiry, A. B., Corless, C. L., Zalupski, M. M., Lowy, A. M., Thomas, C. R., Jr., Alberts, S. R., Dawson, L. A., Micetich, K. C., Thomas, M. B., Siegel, A. B., Blanke, C. D., *Journal of clinical oncology* 2015, *33*, 2617-2622.

[7] Choi, S. B., Park, S. W., Kim, K. S., Choi, J. S., Lee, W. J., *Journal of surgical oncology* 2009, *99*, 335-342.

[8] Grewal, P. K., Uchiyama, S., Ditto, D., Varki, N., Le, D. T., Nizet, V., Marth, J. D., *Nat Med* 2008, *14*, 648-655.

[9] Ohtsubo, K., Marth, J. D., *Cell* 2006, *126*, 855-867.

[10] Helenius, A., Aebi, M., *Science* 2001, *291*, 2364-2369.

[11] Tan, Z., Yin, H., Nie, S., Lin, Z., Zhu, J., Ruffin, M. T., Anderson, M. A., Simeone, D. M., Lubman, D. M., *J Proteome Res* 2015, *14*, 1968-1978.

[12] Nie, S., Lo, A., Wu, J., Zhu, J., Tan, Z., Simeone, D. M., Anderson, M. A., Shedden, K. A., Ruffin, M. T., Lubman, D. M., *J Proteome Res* 2014, *13*, 1873-1884.

[13] Weiz, S., Wieczorek, M., Schwedler, C., Kaup, M., Braicu, E. I., Sehouli, J., Tauber, R., Blanchard, V., *Electrophoresis* 2016, *37*, 1461–1467.

[14] Mechref, Y., Muddiman, D. C., *Analytical and bioanalytical chemistry* 2017, *409*, 355-357.

[15] Hu, Y., Shihab, T., Zhou, S., Wooding, K., Mechref, Y., *Electrophoresis* 2016, *37*, 1498-1505.

[16] Sethi, M. K., Fanayan, S., *International journal of molecular sciences* 2015, *16*, 29278-29304.

[17] Taniguchi, N., Kizuka, Y., *Adv Cancer Res* 2015, *126*, 11-51.

[18] Kailemia, M. J., Park, D., Lebrilla, C. B., *Analytical and bioanalytical chemistry* 2017, *409*, 395-410.

[19] Banazadeh, A., Veillon, L., Wooding, K. M., Zabet-Moghaddam, M., Mechref, Y., *Electrophoresis* 2017, *38*, 162-189.

[20] Korekane, H., Hasegawa, T., Matsumoto, A., Kinoshita, N., Miyoshi, E., Taniguchi, N., *Biochimica et biophysica acta* 2012, *1820*, 1405-1411.

[21] Taketa, K., Sekiya, C., Namiki, M., Akamatsu, K., Ohta, Y., Endo, Y., Kosaka, K., *Gastroenterology* 1990, *99*, 508-518.

[22] Kumada, T., Toyoda, H., Tada, T., Kiriyama, S., Tanikawa, M., Hisanaga, Y., Kanamori, A., Tanaka, J., Kagebayashi, C., Satomura, S., *J Gastroenterol* 2014, *49*, 555-563.

[23] Zhu, J., Lin, Z., Wu, J., Yin, H., Dai, J., Feng, Z., Marrero, J., Lubman, D. M., *J Proteome Res* 2014, *13*, 2986-2997.

[24] Veillon, L., Huang, Y., Peng, W., Dong, X., Cho, B. G., Mechref, Y., *Electrophoresis* 2017, *0*, 1-15

[25] Mechref, Y., Hu, Y., Desantos-Garcia, J. L., Hussein, A., Tang, H., *Mol Cell Proteomics* 2013, *12*, 874-884.

[26] Kalay, H., Ambrosini, M., van Berkel, P. H., Parren, P. W., van Kooyk, Y., Garcia Vallejo, J. J., *Anal Biochem* 2012, *423*, 153-162.

[27] Wu, J., Xie, X., Liu, Y., He, J., Benitez, R., Buckanovich, R. J., Lubman, D. M., *J Proteome Res* 2012, *11*, 4541-4552.

[28] Chen, S., LaRoche, T., Hamelinck, D., Bergsma, D., Brenner, D., Simeone, D., Brand, R. E., Haab, B. B., *Nature methods* 2007, *4*, 437-444.

[29] Laroy, W., Contreras, R., Callewaert, N., *Nat Protoc* 2006, *1*, 397-405.

[30] Chen, J., Fang, M., Zhao, Y. P., Yi, C. H., Ji, J., Cheng, C., Wang, M. M., Gu, X., Sun, Q. S., Chen, X. L., Gao, C. F., *PloS one* 2015, *10*, e0127022.

[31] Callewaert, N., Van Vlierberghe, H., Van Hecke, A., Laroy, W., Delanghe, J., Contreras, R., *Nat Med* 2004, *10*, 429-434.

[32] Juntermanns, B., Sotiropoulos, G. C., Radunz, S., Reis, H., Heuer, M., Baba, H. A., Canbay, A., Schuler, M., Gerken, G., Paul, A., Kaiser, G. M., *Annals of surgical oncology* 2013, *20*, 277-284.

[33] Fang, M., Zhao, Y. P., Zhou, F. G., Lu, L. G., Qi, P., Wang, H., Zhou, K., Sun, S. H., Chen, C. Y., Gao, C. F., *International journal of cancer* 2010, *127*, 148-159.

[34] Guo, H., Nairn, A., dela Rosa, M., Nagy, T., Zhao, S., Moremen, K., Pierce, M., *The Journal of biological chemistry* 2012, *287*, 24941-24954.

[35] Liu, X. E., Desmyter, L., Gao, C. F., Laroy, W., Dewaele, S., Vanhooren, V., Wang, L., Zhuang, H., Callewaert, N., Libert, C., Contreras, R., Chen, C., *Hepatology* 2007, *46*, 1426-1435.

[36] Sakata, J., Wakai, T., Matsuda, Y., Ohashi, T., Hirose, Y., Ichikawa, H., Kobayashi, T., Minagawa, M., Kosugi, S. I., Koyama, Y., Akazawa, K., Ajioka, Y., *Annals of surgical oncology* 2016, *23*, 225-234.

[37] Ito, K., Ito, H., Allen, P. J., Gonen, M., Klimstra, D., D'Angelica, M. I., Fong, Y., DeMatteo, R. P., Brennan, M. F., Blumgart, L. H., Jarnagin, W. R., *Ann Surg* 2010, *251*, 675-681.

[38] van der Gaag, N. A., Kloek, J. J., de Bakker, J. K., Musters, B., Geskus, R. B., Busch, O. R., Bosma, A., Gouma, D. J., van Gulik, T. M., *Annals of oncology* 2012, *23*, 2642-2649.

[39] Esnaola, N. F., Meyer, J. E., Karachristos, A., Maranki, J. L., Camp, E. R., Denlinger, C. S., *Cancer* 2016, *122*, 1349-1369.

[40] Kaprio, T., Satomaa, T., Heiskanen, A., Hokke, C. H., Deelder, A. M., Mustonen, H., Hagstrom, J., Carpen, O., Saarinen, J., Haglund, C., *Mol Cell Proteomics* 2015, *14*, 277-288.

[41] Alley, W. R., Jr., Vasseur, J. A., Goetz, J. A., Svoboda, M., Mann, B. F., Matei, D. E., Menning, N., Hussein, A., Mechref, Y., Novotny, M. V., *J Proteome Res* 2012, *11*, 2282-2300.

[42] Arnold, J. N., Saldova, R., Galligan, M. C., Murphy, T. B., Mimura-Kimura, Y., Telford, J. E., Godwin, A. K., Rudd, P. M., *J Proteome Res* 2011, *10*, 1755-1764.

[43] Lin, Z., Simeone, D. M., Anderson, M. A., Brand, R. E., Xie, X., Shedden, K. A., Ruffin, M. T., Lubman, D. M., *J Proteome Res* 2011, *10*, 2602-2611.

[44] Mechref, Y., Hu, Y., Garcia, A., Zhou, S., Desantos-Garcia, J. L., Hussein, A., *Bioanalysis* 2012, *4*, 2457-2469.

[45] Tanabe, K., Deguchi, A., Higashi, M., Usuki, H., Suzuki, Y., Uchimura, Y., Kuriyama, S., Ikenaka, K., *Biochemical and biophysical research communications* 2008, *374*, 219-225.

[46] Mehta, A., Herrera, H., Block, T., *Adv Cancer Res* 2015, *126*, 257-279.

[47] Pearce, O. M., Laubli, H., *Glycobiology* 2016, *26*, 111-128.

[48] Bull, C., Stoel, M. A., den Brok, M. H., Adema, G. J., *Cancer Res* 2014, *74*, 3199-3204.

[49] Zhou, S., Dong, X., Veillon, L., Huang, Y., Mechref, Y., *Analytical and bioanalytical chemistry* 2017, *409*, 453-466.

[50] Hu, Y., Shihab, T., Zhou, S., Wooding, K., Mechref, Y., *Electrophoresis* 2016, *37*, 1498-1505.

[51] Li, H., Gao, W., Feng, X., Liu, B. F., Liu, X., *Analytica chimica acta* 2016, *924*, 77-85.

**Figure Captions**

Figure 1. A representative *N*-glycome profiling for total serum glycoproteins. Thirteen peaks were identified. The structure of each peak were shown below the panel.

NGA2F NGA2FB NG1A2F NA2 NA2F NA2FB NA3 NA3Fb NA3F NA3F2 NA4 NA4Fb

1    2    3/4    5    6    7    8    9    9'    10    11    12

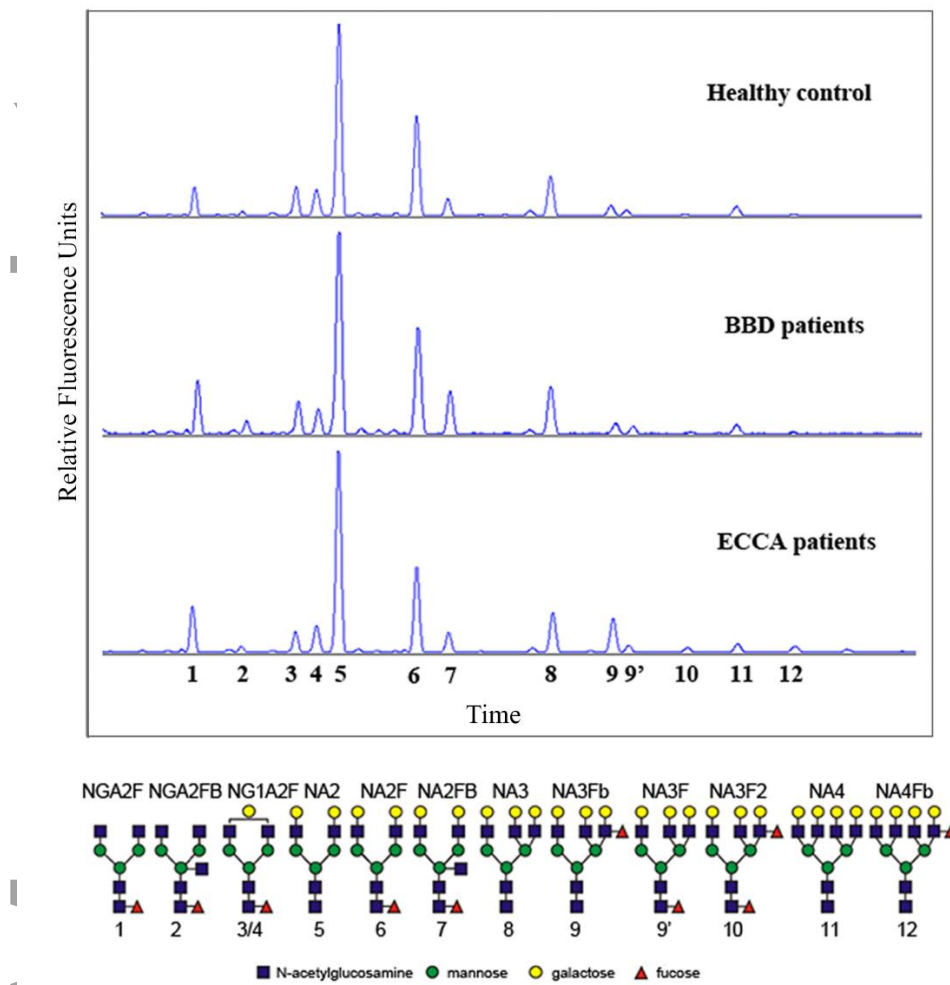■ N-acetylglucosamine  ● mannose  ○ galactose  ▲ fucose

Figure 2. ROC curves for the detection of ECCA. Figure 2A shows the ROCs of single parameter including CA19-9 (AUC=0.918), peak10 (AUC=0.948) and the diagnostic model Glycotest1 (AUC=0.992) for identifying ECCA from HC. Figure 2B shows the ROCs of CA19-9 (AUC=0.851), peak10 (AUC=0.909) and the model Glycotest2 (AUC=0.938) for distinguishing ECCA from BBD.

Glycotest1=0.089*CA19-9+11.444*peak10-0.331*peak5-0.540*peak6+17.058;
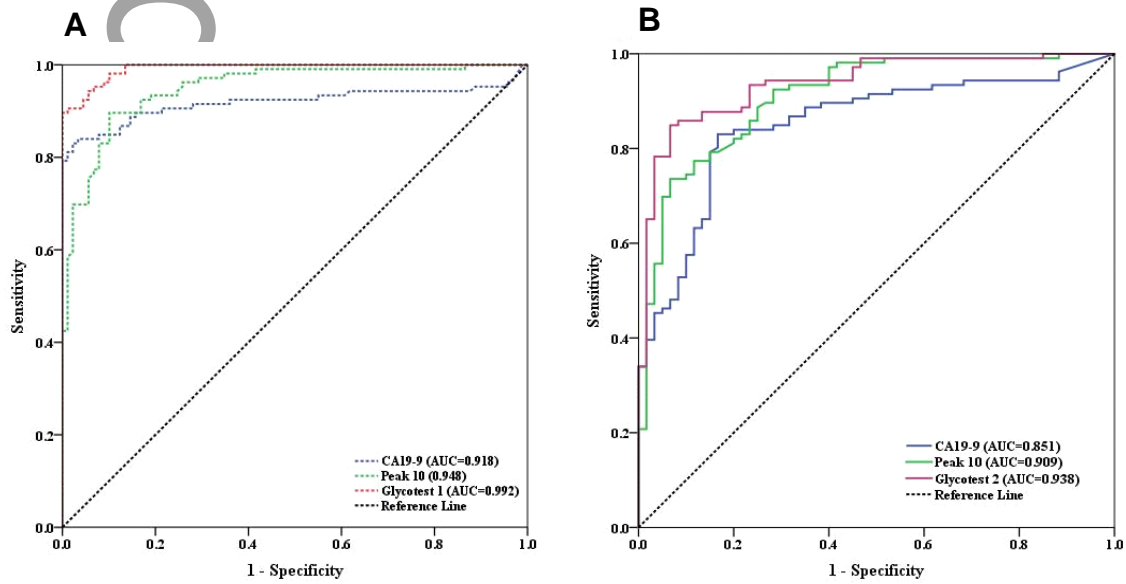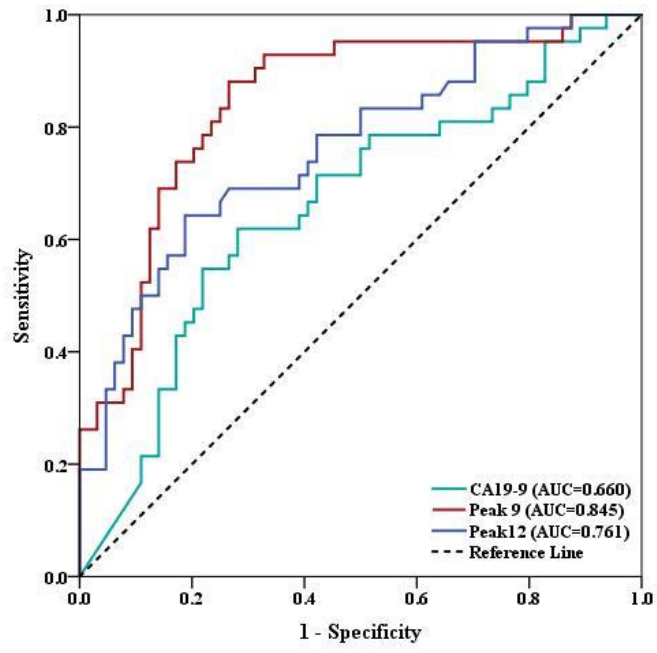
Glycotest2=0.009*CA19-9+9.112*peak10-4.370.



Figure 3. ROC curves of peak 9, peak 12 and CA19-9 for the detection of ECCA with lymph nodes involvement ($AUC_{peak9}$=0.845, $AUC_{peak12}$=0.761, $AUC_{CA19-9}$=0.660).

**Tables**

Table 1. Demographic and clinical characteristics of study population

| Variables | ECCA (n=106) | BBD (n=60) | HC (n=89) | *P* | |
|---|---|---|---|---|---|
| | | | | ECCA vs BBD | ECCA vs HC |
| Age (years) | 58.50, (51.00, 64.00) | 53.00, (35.25, 58.00) | 42.00, (31.50, 49.00) | <0.05 | <0.05 |
| TBIL (μmol/L) | 129.90, (38.55, 244.65) | 11.70, (9.30, 16.95) | 12.30, (10.00, 15.25) | <0.001 | <0.001 |
| TP (g/L) | 67.50, (62.35, 72.80) | 71.40, (67.85, 75.05) | 77.00, (74.00, 79.65) | 0.002 | <0.001 |
| ALB (g/L) | 39.40, (36.40, 42.15) | 41.60, (39.40, 45.05) | 47.00, (45.90, 48.65) | <0.001 | <0.001 |

| | | | | |
|---|---|---|---|---|
| ALT (IU/L) | 109.60, (64.40, 205.45) | 20.00, (14.55, 32.90) | 16.00, (11.15, 26.50) | <0.001 | <0.001 |
| GLU (mmol/L) | 5.45, (4.97, 6.28) | 4.93, (4.64, 5.29) | 4.94, (4.58, 5.20) | <0.001 | <0.001 |
| PLT(10^9/L) | 250.00, (239.78, 272.51) | 213.00, (195.01, 235.23) | 216.50, (205.44, 226.70) | 0.004 | 0.001 |
| CA19-9 (U/ml) | 166.10, (50.30, 470.10) | 10.70, (4.53, 26.53) | 7.20, (4.15, 15.43) | <0.001 | <0.001 |
| CEA (µg/L) | 2.70, (1.80, 5.30) | 1.50, (0.88, 2.13) | 1.65, (0.88, 1.88) | <0.001 | <0.001 |
| AFP (µg/L) | 2.90, (2.10, 4.10) | 2.25, (1.63, 3.00) | 2.20, (1.58, 3.55) | 0.370 | 0.290 |
| CA72-4 (U/ml) | 1.80, (1.20, 5.60) | 1.30, (0.90, 2.50) | 1.85, (0.90, 2.73) | 0.007 | 0.004 |
| CA125 (U/ml) | 14.30, (10.90, 21.9) | 13.75, (8.30, 18.35) | 12.50, (10.05, 19.03) | 0.113 | 0.038 |

Table 2. General *N*-glycome profiling results in 3 different groups

| Variables | ECCA (n=106) | BBD (n=60) | HC (n=89) | *P* | |
|---|---|---|---|---|---|
| | | | | ECCA vs BBD | ECCA vs HC |
| Peak1 (NGA2F) | 7.44, (5.98, 9.91) | 8.01, (5.74, 10.07) | 6.40, (5.09, 7.83) | 0.775 | <0.001 |
| Peak2 (NGA2FB) | 1.165, (0.96, 1.48) | 1.20, (0.87, 1.55) | 0.97, (0.74, 1.21) | 0.814 | <0.001 |
| Peak3 (NG1A2F) | 5.76 (4.81, 6.64) | 6.65, (5.44, 7.32) | 6.29, (5.58, 7.01) | 0.001 | 0.001 |
| Peak4 (NG1A2F) | 5.44, (4.92, 6.33) | 5.41, (4.97, 6.10) | 5.08, (4.53, 5.81) | 0.961 | 0.002 |
| Peak5 (NA2) | 38.22, (35.39, 41.56) | 37.28, (35.12, 40.04) | 39.35, (37.44, 41.635) | 0.35 | 0.038 |
| Peak6 (NA2F) | 19.22, (17.18, 21.28) | 20.24, (17.78, 22.56) | 21.80, (19.78, 24.23) | 0.052 | <0.001 |
| Peak7 (NA2FB) | 5.23, (4.62, 6.31) | 6.16, (5.03, 7.06) | 5.57, (4.905, 6.48) | 0.001 | 0.116 |
| Peak8 (NA3) | 6.52, (5.08, 8.66) | 7.35, (6.18, 8.37) | 8.07, (6.69, 9.38) | 0.118 | <0.001 |

| | | | | |
|---|---|---|---|---|
| Peak9 (NA2Fb) | 4.93, (3.89, 6.34) | 2.55, (1.92, 3.49) | 2.25, (1.61, 3.11) | <0.001 | <0.001 |
| Peak9' (NA3F) | 1.09, (0.83, 1.43) | 0.81, (0.66, 1.02) | 0.86, (0.73, 0.97) | <0.001 | <0.001 |
| Peak10 (NA3F2) | 0.63, (0.47, 0.822) | 0.28, (0.19, 0.39) | 0.26, (0.21, 0.325) | <0.001 | <0.001 |
| Peak11 (NA4) | 1.38, (0.93, 2.03) | 1.78, (1.38, 2.25) | 1.8, (1.52, 2.18) | 0.004 | <0.001 |
| Peak12 (NA4Fb) | 0.80, (0.65, 1.00) | 0.53, (0.41, 0.84) | 0.49, (0.33, 0.655) | <0.001 | <0.001 |

Table 3. Diagnostic power of each index in differentiating ECCA from healthy individuals

| Cut-off value | Test Result | Actual Status | | Sen (%) | Spe (%) | PPV (%) | NPV (%) | Accuracy (%) |
|---|---|---|---|---|---|---|---|---|
| | | ECCA+ | ECCA- | | | | | |
| CA19-9 (32.00U/ml) | ECCA+ | 88 | 0 | 83.0 | 100.0 | 100.0 | 83.2 | 90.8 |
| | ECCA- | 18 | 89 | | | | | |
| Peak10 (0.38) | ECCA+ | 92 | 9 | 86.8 | 90.0 | 91.1 | 85.1 | 88.2 |
| | ECCA- | 14 | 80 | | | | | |
| Glycotest1 (1.83) | ECCA+ | 95 | 0 | 89.6 | 100.0 | 100.0 | 89.0 | 94.4 |
| | ECCA- | 11 | 89 | | | | | |

Glycotest1=0.089*CA19-9+11.444*peak10-0.331*peak5-0.540*peak6+17.058

Sen: Sensitivity; Spe: Specificity; PPV: Positive predictive value; NPV: Negative predictive value

Table 4. Diagnostic power of each index in differentiating ECCA from BBD

| Cut-off value | Test Result | Actual Status | | Sen (%) | Spe (%) | PPV (%) | NPV (%) | Accuracy (%) |
|---|---|---|---|---|---|---|---|---|
| | | ECCA+ | ECCA- | | | | | |
| CA19-9 (32.00U/ml) | ECCA+ | 88 | 10 | 83.0 | 83.3 | 89.8 | 73.5 | 83.1 |
| | ECCA- | 18 | 50 | | | | | |
| Peak10 (0.49) | ECCA+ | 78 | 6 | 73.6 | 90.0 | 92.9 | 65.9 | 79.5 |
| | ECCA- | 28 | 54 | | | | | |
| Glycotest2 (0.56) | ECCA+ | 90 | 4 | 84.9 | 93.3 | 95.7 | 77.8 | 88.0 |

Glycotest2=0.009*CA19-9+9.112*peak10-4.370

Sen: Sensitivity; Spe: Specificity; PPV: Positive predictive value; NPV: Negativpredictive value