Talanta 274 (2024) 126056

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Single cell glycan-linkages profiling for hepatocellular carcinoma early diagnosis using lanthanide encoded bacteriophage MS2 based ICP-MS

Yong Liang [a,c,*], Zhen Liu [a], Dongliang Zuo [b,d], Shi Chen [a], Jianbin Chen [a], Xiaowen Yan [a], Pingguo Liu [b,**], Qiuquan Wang [a,***]

[a] Department of Chemistry & the MOE Key Laboratory of Spectrochemical Analysis and Instrumentation, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen, 361005, China
[b] Department of Hepatobiliary Surgery, Zhongshan Hospital, Xiamen University, Fujian Provincial Key Laboratory for Chronic Liver Disease and Hepatocellular Carcinoma, Xiamen, 361004, China
[c] Science and Technology Innovation Center, Guangzhou University of Chinese Medicine, Guangzhou, 510405, China
[d] The Central Hospital of Enshi Tujia and Miao Autonomous Prefecture, Enshi, 445000, China

ARTICLE INFO

Handling Editor: Prof. J. Wang

Keywords:
HCC
Early diagnosis
Glycan-linkages
Single cell
ICP-MS

ABSTRACT

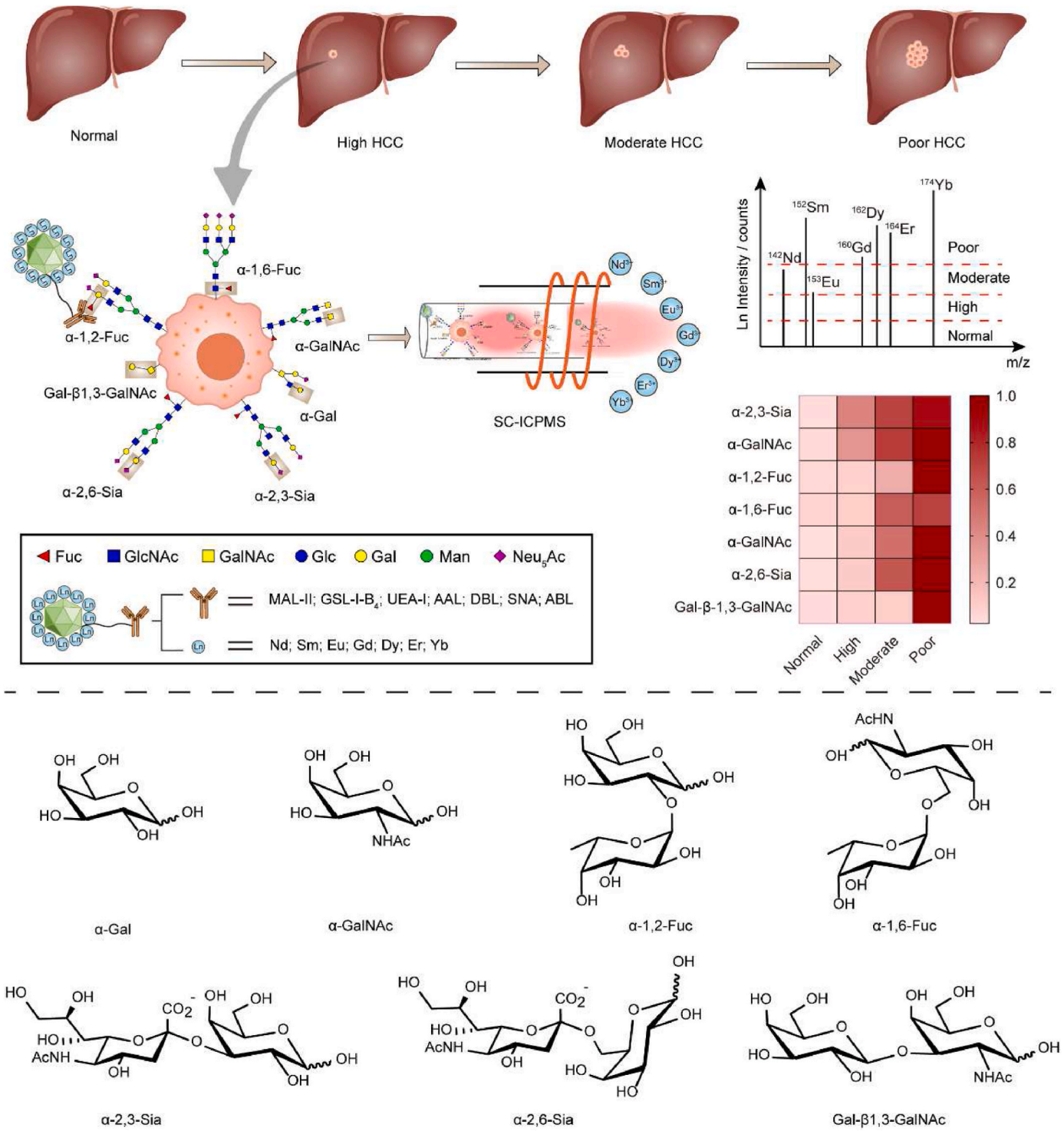
Early diagnosis is paramount for enhancing survival rates and prognosis in the context of malignant diseases. Hepatocellular carcinoma (HCC), the second leading cause of cancer-related deaths worldwide, poses significant challenges for its early detection. In this study, we present an innovative approach which contributed to the early diagnosis of HCC. By lanthanide encoding signal amplification to map glycan-linkages at the single-cell level, the minute quantities of "soft" glycan-linkages on single cell surface were converted into "hard" elemental tags through the use of an MS2 signal amplifier. Harnessing the power of lanthanides encoded within MS2, we achieve nearly three orders of magnitude signal amplification. These encoded tags are subsequently quantified using single-cell inductively coupled plasma mass spectrometry (SC-ICP-MS). Linear discriminant analysis (LDA) identifies seven specific glycan-linkages (α-2,3-Sia, α-Gal, α-1,2-Fuc, α-1,6-Fuc, α-2,6-Sia, α-GalNAc, and Gal-β-1,3-GalNAc) as biomarkers. Our methodology is initially validated at the cellular level with 100% accuracy in discriminating between hepatic carcinoma HepG2 cells and their normal HL7702 cells. We apply this approach to quantify and classify glycan-linkages on the surfaces of 55 clinical surgical HCC specimens. Leveraging these seven glycan-linkages as biomarkers, we achieve precise differentiation between 8 normal hepatic specimens, 40 early HCC specimens, and 7 colorectal metastasis HCC specimens. This pioneering work represents the first instance of employing single-cell glycan-linkages as biomarkers promising for the early diagnosis of HCC with a remarkable 100% predictive accuracy rate, which holds immense potential for enhancing the feasibility and precision of HCC diagnosis in clinical practice.

1. Introduction

Hepatocellular carcinoma (HCC) represents a formidable global health challenge, ranking as the second most common cause of cancer-related mortality. It predominantly afflicts individuals with a history of cirrhosis, often stemming from chronic alcohol abuse, non-alcoholic fatty liver disease, or infection with hepatitis B or C viruses (HBV or HCV) [1,2]. Regrettably, the clinical manifestation of HCC in its early stages is often subtle, leading to a dearth of accurate early diagnostic biomarkers. Consequently, a significant proportion of HCC cases are diagnosed at an advanced and often incurable stage, with a disheartening 5-year survival rate of merely 10% [3]. Current diagnostic strategies for HCC encompass liver imaging modalities such as computed tomography (CT) and magnetic resonance imaging (MRI), as well as the measurement of serum α-fetoprotein (AFP) levels [4–6]. Nevertheless, distinguishing small HCC lesions from cirrhotic nodules remains a substantial challenge when relying solely on imaging techniques [7,8]. Furthermore, the diagnostic utility of AFP, with its reported sensitivity

* Corresponding author. Department of Chemistry & the MOE Key Laboratory of Spectrochemical Analysis and Instrumentation, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen, 361005, China.
** Corresponding author.
*** Corresponding author.
E-mail addresses: liangyong@gzucm.edu.cn (Y. Liang), pgliu@xmu.edu.cn (P. Liu), qqwang@xmu.edu.cn (Q. Wang).

https://doi.org/10.1016/j.talanta.2024.126056
Received 11 January 2024; Received in revised form 1 April 2024; Accepted 4 April 2024
Available online 5 April 2024



Scheme 1. Overall frame of HCC early diagnosis by glycan-linkages machinery learning analysis. α-2,3-Sia, α-Gal, α-1,2-Fuc, α-1,6-Fuc, α-GalNAc, α-2,6-Sia and Gal-β-1,3-GalNAc are encoded by Nd, Sm, Eu, Gd, Dy, Er and Yb; lectin = MAL-II, GSL-I-B4, UEA-I, AAL, DBL, SNA and ABL using a MS2 signal amplifier and then monitored the isotopes of ^{142}Nd, ^{152}Sm, ^{153}Eu, ^{160}Gd, ^{162}Dy, ^{164}Er, and ^{174}Yb with single cell inductively coupled plasma mass spectrometry. The measured quantity of glycan-linkages of paracancerous HL7702 cell, cancerous HepG2 cell, normal surgically tissues and early HCC tissues as well as hepatocellular metastatic carcinoma tissues are performed with linear discriminate analysis.

ranging from 41% to 65% and specificity from 80% to 94%, is compromised by its limited ability to discriminate early-stage HCC from other hepatobiliary disorders, including acute hepatitis and cirrhosis [9,10]. This deficiency in sensitivity often results in the detection of HCC only in its advanced stages. Therefore, enhancing the overall survival rates of individuals afflicted by HCC compels us to seek more efficacious early diagnostic biomarkers and explore novel therapeutic targets.

The realm of glycosylation post-translational modification research has witnessed remarkable advancements, especially in the characterization of glycan structures and quantification, with a particular focus on glycomics studies related to carcinoma [11,12]. Extensive investigation into aberrant glycosylation in Hepatocellular Carcinoma (HCC)

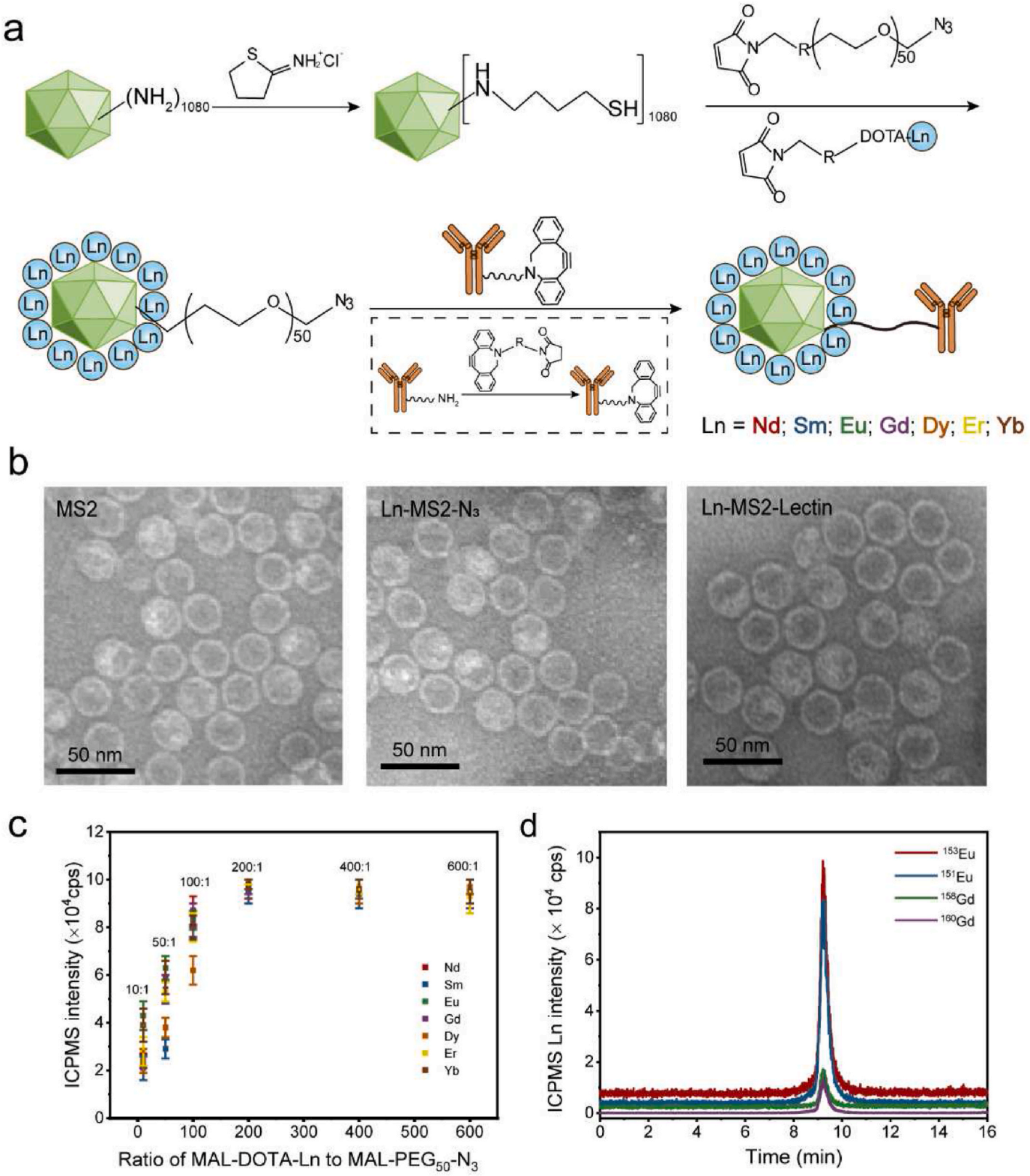


Fig. 1. (a) The preparation of Ln-MS2-lectin signal amplification platform; (b) Transmission electron microscopy imaging of MS2, Ln-MS2 and Ln-MS2-Lectin; (c) Evaluation of reactants ratio within Mal-DOTA-Ln and Mal-PEG_{50}-N_3 determined using ICP-MS; (d) Size-exclusion chromatography isotope dilution ICP-MS spiked using ^{153}Eu and ^{158}Gd for accurately quantifying the number of lanthanide number tagged on MS2 and the number of azide group which was re-clickable by DBCO-DOTA-Gd.

has unveiled its pivotal role in HCC tumorigenesis and invasion [13–15]. An exemplar of such progress is the identification of core-fucosylated AFP, which exhibits superior specificity in HCC detection compared to AFP alone and has gained approval from the United States Food and Drug Administration [15,16]. Additionally, glycoproteins like GP73 and haptoglobin are currently under scrutiny as potential HCC biomarkers within the glycomics framework [17]. Furthermore, several glycan structures have been scrutinized for their predictive value in HCC development [18]. Studies have underscored the close association between heightened levels of α-2,6-linked sialylation and tetra-antennary N-glycan structures with HCC progression [19,20]. Notably, an escalation in tetra-antennary glycan levels has been observed in HCC tissues relative to surrounding non-tumor liver tissues. Complex N-glycan structures, such as GlcNAc-β-1-6 branching, have been posited to wield a pivotal influence on cancer metastasis [21]. These insights into glycan-linkages hold the promise of providing valuable, complementary information for the precise differentiation of HCC and, consequently, enhancing its treatment outcomes [22]. However, despite the potential wealth of information offered by glycan-linkage analysis, the clinical application of measuring diverse glycan-linkage modifications expressed on the surface of a single cell as a guide for early HCC diagnosis remains elusive. This can be attributed to concerns about potential disruptions to the glycosylation machinery resulting from alterations in biochemical environments.

Given the intricate pathology and inherent heterogeneity of Hepatocellular Carcinoma (HCC), existing data suggests that achieving optimal sensitivity and specificity for HCC detection, especially in its early stages, is unlikely with a single biomarker alone [23]. HCC arises from a single aberrant hepatocellular cell and exhibits significant diversity even within the same cell populations. Hence, the precise and sensitive differentiation of biomarkers at the single-cell level holds paramount importance for early HCC diagnosis. Alterations in glycan linkages and their upregulated quantities offer the potential to furnish more precise indicators of impending HCC onset [24]. Consequently, a comprehensive and global analysis is imperative to investigate glycosylation changes during tumor development and progression, thus complementing each other and bolstering early diagnostic capabilities.

In this context, we introduce a novel approach for HCC discovery that contributed to detect the earliest abnormal HCC cells with single-cell sensitivity through the mapping of multiple glycan linkages. Unlike the commonly studied immuno-histochemical staining biomarkers, our method leverages the conversion of “hard” glycan linkages on the surface of single cells [25–27], into “hard” lanthanides using single-cell inductively coupled plasma mass spectrometry (SC-ICP-MS) [28]. These transformed glycan linkages are subsequently classified using linear discriminant analysis (LDA) to distinguish early HCC. To achieve single-cell sensitivity, we employ virus-like bacteriophage nanoparticles, specifically MS2, as signal amplifiers. These nanoparticles are linked with glycan-linkage specific lectins, including *Maackia Amurensis* Lectin II (MAL-II), *Griffonia Simplicifolia* Lectin I isolectin B4 (GSL-I-B4), *Ulex Europaeus* Agglutinin I (UEA-I), *Aleuria Aurantia* Lectin (AAL), *Dolichos Biflorus* Lectin (DBL), *Sambucus nigra* Agglutinin from elderberry bark (SNA), and *Agaricus Bisporus* Lectin (ABL), forming signal amplifiers such as Nd-MS2-PEG_{50}-MAL-II, Sm-MS2-PEG_{50}-GSL-I-B4, Eu-MS2-PEG_{50}-UEA-I, Gd-MS2-PEG_{50}-AAL, Dy-MS2-PEG_{50}-DBL, Er-MS2-PEG_{50}-SNA, and Yb-MS2-PEG_{50}-ABL. These amplifiers selectively recognize cell surface glycan linkages, including α-2,3-Sia, α-Gal, α-1,2-Fuc, α-1,6-Fuc, α-GalNAc, α-2,6-Sia, and Gal-β-1,3-GalNAc (T antigen) [29,30]. In contrast to conventional fluorescent techniques and “soft” ionization mass spectrometry [25–27], our method converts these “soft” glycan linkages into “hard” elemental mass readouts using ICP-MS. This instrument is renowned for its capability to quantify inorganic elements and isotopes, demonstrating clear advantages in biomolecule quantification [31–34]. By harnessing the power of lanthanides encoded within MS2, we achieve nearly three orders of magnitude signal amplification [35]. This technological breakthrough enables the analysis of glycan linkages on a single-cell level with unparalleled sensitivity. Furthermore, quantitative data on glycan linkages at the single-cell surface, derived from laboratory-cultivated hepatocellular cells and surgically resected HCC neoplasm specimens, are subjected to linear discriminant analysis to facilitate early HCC diagnosis (Scheme 1).

2. Results and discussions

One of the central challenges in single-cell glycan-linkage analysis is achieving the necessary sensitivity due to the minuscule quantities of glycans present on individual cells. To address this challenge, we employ bacteriophage nanoparticles, specifically MS2, as signal amplifiers in this study. It’s worth noting that our laboratory pioneered the use of MS2-based signal amplification [35]. In this work, we take a novel approach by first converting the active –NH_2 groups on the MS2 surface to active –SH groups. These –SH groups are then linked to maleimide-lanthanides (MAL-Ln, where Ln represents Nd, Sm, Eu, Gd, Dy, Er, and Yb) and a maleimide-PEG_{50}-N_3 (MAL-PEG_{50}-N_3) linker using cysteine-maleimide chemistry (Figs. S1 and S2). Subsequently, we conjugate Ln-MS2-PEG50-N_3 with dibenzocyclooctyne-lectin (DBCO-Lectin, lectin types: MAL-II, GSL-I-B4, UEA-I, AAL, DBL, SNA, and ABL) through an N_3-DBCO copper-free click reaction (Fig. 1a). Transmission electron microscopy (TEM) confirms that Ln-MS2-N_3 and Ln-MS2-lectin maintain excellent dispersibility and dimensional homogeneity in aqueous buffer when compared to native MS2 nanoparticles (Fig. 1b). To optimize the lanthanide signal tagging on MS2, we systematically varied the molecular ratio between MAL-DOTA-Ln and MAL-PEG_{50}-N_3 at 10:1, 50:1, 100:1, 200:1, 400:1, and 600:1. Our results indicated that the lanthanide intensity tagged on MS2 reached a plateau and achieved the highest lanthanide tagging efficiency at a ratio of 200:1 (Fig. 1c). Furthermore, we quantified the number of SH-MMA conjugated lanthanide atoms and azide groups attached to each MS2 particle using double isotope dilution ICP-MS (^{153}Eu and ^{160}Gd used here). The results of ^{153}Eu isotope dilution ICP-MS confirmed that each MS2 nanoparticle was tagged with 986 lanthanide atoms (MMA-DOTA-Eu exampled) (Fig. 1d), while the azide groups were re-clicked with DBCO-DOTA-Gd via N_3-cyclooctyne copper free click reaction and thus determined almost 90 azide groups linked per MS2 (Fig. 1d, Fig. S3). Such results were consistent with the total 1080 –SH groups per MS2 particle. Before anchoring lectins onto Ln-MS2-N_3, we prepared DBCO-modified lectins through NHS and NH_2 conjugation of commercially purchased lectins (Table S1, Fig. S4). Given the larger size and molecular mass of lectins (up to 140 kDa), which may introduce steric hindrance when anchored onto MS2, we determined the actual number of lectins attached to Ln-MS2-lectin using spectral absorption methods. Our results (Fig. S5) indicate an average of 6 DBL, 5 SNA, 10 UEA, 3 GSL-I-B4, 5 ABL, 4 MAL, and 2 AAL molecules linked to each MS2. Thus, we successfully obtained Nd_{986}-MS2-PEG_{50}-MAL-II_4, Sm_{986}-MS2-PEG_{50}-GSL-I-$B4_3$, Eu_{986}-MS2-PEG_{50}-UEA-I_{10}, Gd_{986}-MS2-PEG_{50}-AAL_2, Dy_{986}-MS2-PEG_{50}-DBL_6, Er_{986}-MS2-PEG_{50}-SNA_6, and Yb_{986}-MS2-PEG_{50}-ABL_6 conjugates.

To verify the binding specificity of Ln_{986}-MS2-lectin conjugates, we conducted glycan-linkage binding assays, despite manufacturer declarations. These assays included standard microliter plate glycan binding screenings (Table S2). The results revealed that the lectins exhibited a binding specificity of at least 95%, confirming their accessibility to glycan targets (Fig. S6). Our aim was to enable early diagnosis of hepatocellular carcinoma (HCC) using our signal amplifiers. To initiate this research, we performed cell model experiments using laboratory-cultured HepG2 cells (hepatocellular carcinoma) and their paired paracancerous HL7702 cells. These cells were treated with Ln_{986}-MS2-PEG_{50}-lectin conjugates at optimized working concentrations. Specifically, we used 10 nM Gd_{986}-MS2-PEG_{50}-AAL_2 for α-1,6-Fuc targeting, 100 nM Sm_{986}-MS2-PEG_{50}-GSL-I-$B4_3$ for α-Gal targeting, 200 nM Yb_{986}-MS2-PEG_{50}-ABL_5 for Gal-β-1,3-GalNAc targeting, 100 nM Nd_{986}-MS2-PEG_{50}-MAL-II_4 for α-2,3-Sia targeting, 50 nM Eu_{986}-MS2-PEG_{50}-UEA-I_{10} for α-1,2-Fuc targeting, Er_{986}-MS2-PEG_{50}-SNA_6 for α-2,6-Sia targeting, and 100 nM Dy_{986}-MS2-PEG_{50}-DBL_6 for α-GalNAc targeting. These

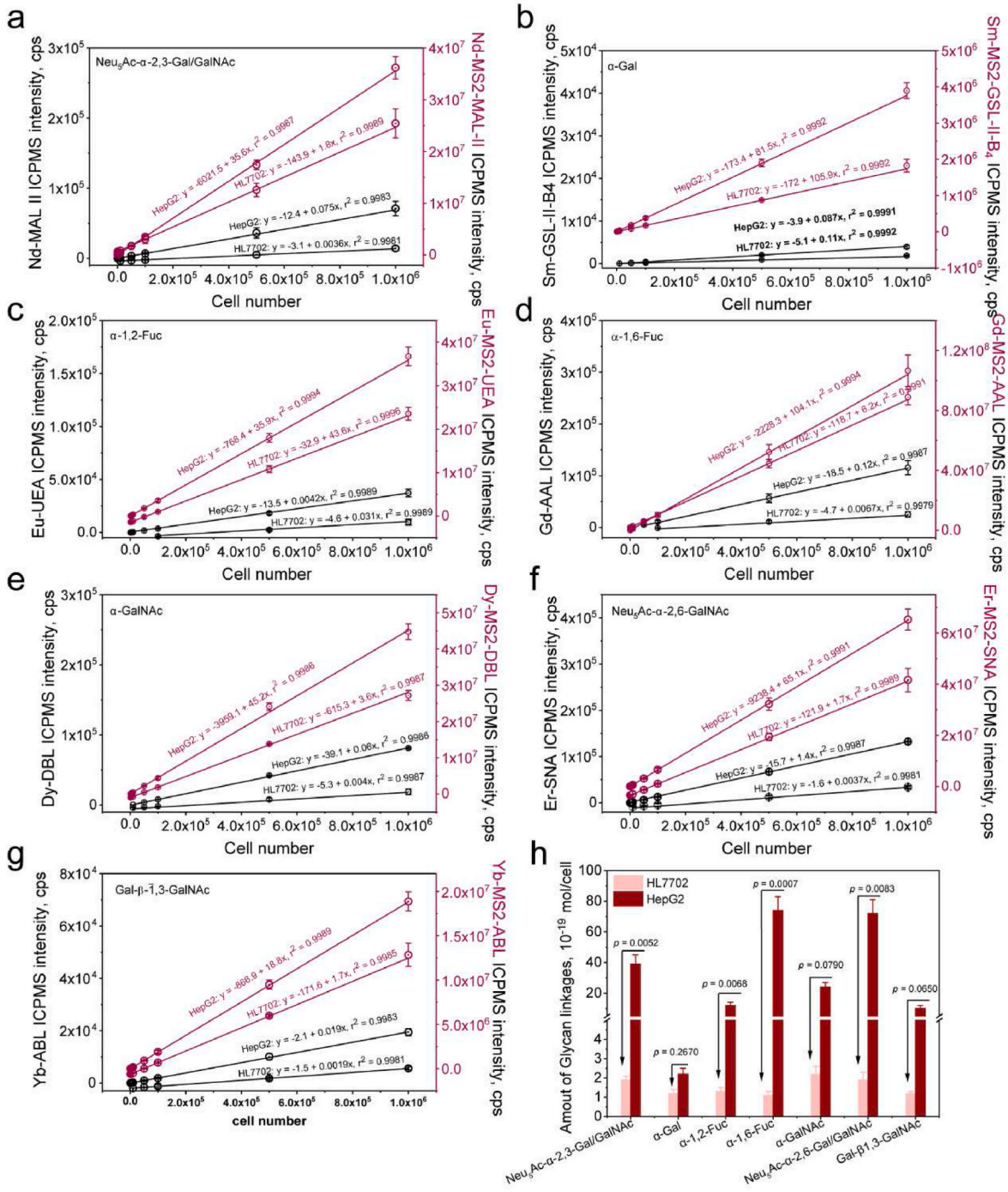


Fig. 2. (a–g) Calibration linear curve of HL7702 and HepG2 cells against Ln-lectin and Ln_{986}-MS2-PEG_{50}-lectin ICP-MS intensity by monitoring ^{142}Nd (a), ^{152}Sm (b), ^{153}Eu (c), ^{160}Gd (d), ^{164}Dy (e), ^{166}Er (f) and ^{174}Yb (g) to determine the average content of glycan-linkages. The curve is colored with black by Ln-lectin and blue by Ln_{986}-MS2-PEG_{50}-lectin. The standard deviation is presented as error bar from 7 repeated runs for each glycan linkages (n = 5); (h) Column chart of expression difference for these seven glycan linkages in HL7702 and HepG2 cells. For α-1,2-Fuc, α-2,3-sialic acid and α-2,6-sialic acid, there're very significant difference between HL7702 (colored green) and HepG2 cells (colored red) ($p^{**} < 0.01$, N = 21), and there's extremely significant difference in α-1,6 Fuc ($p^{***} < 0.001$, N = 21) between HL7702 and HepG2 cells.

concentrations were chosen to prevent cell aggregation (Fig. S7). Comparing the results with bare Ln-DOTA-lectin probes, we achieved nearly a 1000-fold enhancement in ICP-MS readouts when Ln_{986}-MS2-PEG_{50}-lectin conjugates were used for single-cell glycan-linkage quantification. Additionally, we observed better linear calibration correlations ($R^2 > 0.9990$) (Fig. 2). Notably, the detection limit reached as low as 10 fmol (Eu was exemplified with a dwell time of 100 ms) using single-cell ICP-MS (SC-ICP-MS) for all seven targeted glycan linkages.

To determine the quantities of glycan-linkages (α-2,3-Sia encoded with Nd, α-Gal with Sm, α-1,2-Fuc with Eu, α-1,6-Fuc with Gd, α-2,6-Sia with Er, α-GalNAc with Dy, and Gal-β-1,3-GalNAc or T antigen with Yb) using isotope dilution ICP-MS. For the para-cancerous HL7702 cells, we found that the content of α-Gal and α-GalNAc was (1.3 ± 0.2) and (2.2 ± 0.4) × 10^{-19} mol/cell, respectively (Fig. 2a and **b**) (n = 5). These results indicate that the acetylation of galactose was not significantly up-regulated in para-cancerous liver cells. Moreover, Gal-β-1,3-GalNAc,

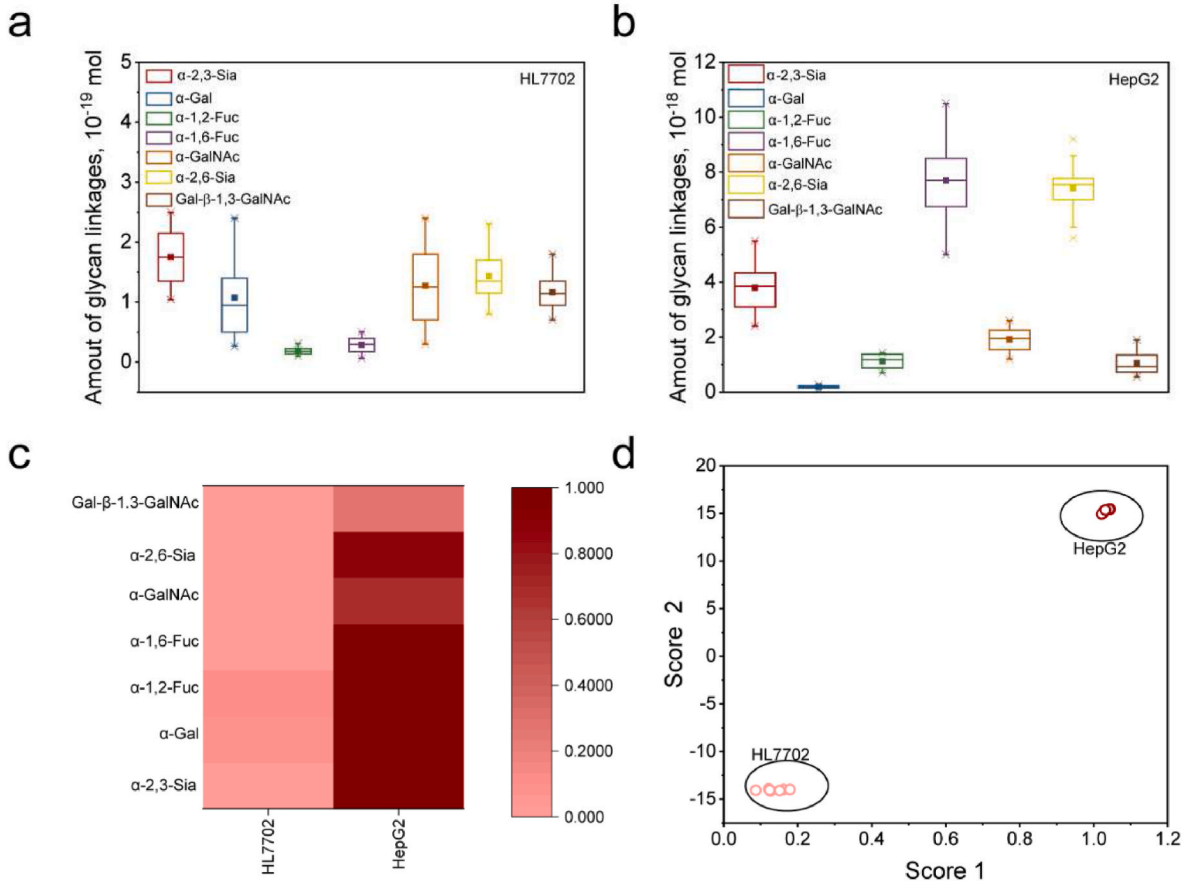


Fig. 3. Box chart of α-Gal, α-GalNAc, α-1,2-Fuc, α-1,6-Fuc, α-2,3-Sia, α-2,6-Sia and Gal-β-1,3-GalNAc distribution of (a) HL7702 and (b) HepG2 (n = 3). (c) Heat map of normalized glycan-linkages of HL7702 and HepG2, the quantity of α-Gal, α-GalNAc, α-1,2-Fuc, α-1,6-Fuc, α-2,3-Sia, α-2,6-sicalic acid and Gal-β-1,3-GalNAc are colored as the right scale. (d) Linear discrimination analysis to distinguish HL7702 and HepG2 ($p < 0.001$) scored from α-Gal, α-GalNAc, α-1,2-Fuc, α-1,6-Fuc, α-2,3-Sia, α-2,6-Sia and Gal-β1,3-GalNAc, each sell species with seven samples are set as training set and other four unknown cell samples as test set. The machinery learning platform depends on IBM Statistics SPSS, Edition 26).

often considered a carcinoma metastasis marker, was found at a lower expression level of $(1.2 \pm 0.1) \times 10^{-19}$ mol/cell (Fig. 2g) (n = 5).

The commonly associated terminal sialic acids (Sia) that are indicative of tumorigenesis showed almost the same content for α-2,3-Sia ($(1.9 \pm 0.2) \times 10^{-19}$) and α-2,6-Sia ($(1.9 \pm 0.4) \times 10^{-19}$) mol/cell on HL7702 cells (Fig. 2e and Fig. **S8f**) (n = 5), suggesting minimal differential expression among Sia linkages in para-cancerous cells. Additionally, the quantities determined for α-1,2-Fuc and α-1,6-Fuc, which play crucial roles in tumorigenesis and invasion, were $(1.3 \pm 0.2) \times 10^{-19}$ and $(1.1 \pm 0.2) \times 10^{-19}$ mol/cell, respectively (Fig. 2c and Fig. S2d) (n = 5). In the case of HepG2 cells, the average content of α-Gal was $(2.2 \pm 0.3) \times 10^{-19}$ mol/cell, without significant up-regulation compared to para-cancerous HL7702 (Fig. 2a). However, the prominently up-regulated acetylated galactose α-GalNAc was $(2.4 \pm 0.3) \times 10^{-19}$ mol/cell in HepG2, a substantial increase compared to HL7702, indicating that α-GalNAc is often over-expressed in most carcinomas, not limited to HCC (Fig. 2b). Furthermore, α-2,3-Sia and α-2,6-Sia in HepG2 cells were found to be $(3.9 \pm 0.6) \times 10^{-18}$ and $(7.2 \pm 0.9) \times 10^{-18}$ mol/cell (Fig. 2e and Fig. S2f), which almost twenty and forty times higher than HL7702 just at the level of 1.9×10^{-19} mol/cell (Fig. 2e and Fig. S8f). Similarly, α-1,2-Fuc and core α-1,6-Fuc increased to $(1.2 \pm 0.2) \times 10^{-18}$ and $(7.1 \pm 0.9) \times 10^{-18}$ mol/cell in HepG2 (Fig. 2c and Fig. S2d) (n = 5), with α-1,6-Fuc per HepG2 cell being almost 70 times higher than in para-cancerous cells. Lastly, the average Gal-β-1,3-GalNAc linkage per HepG2 cell was $(1.0 \pm 0.2) \times 10^{-18}$ mol (n = 5), significantly higher than in HL7702 cells (Fig. 2g). These findings highlight minimal differences in α-Gal expression between the two cell lines, while significant differences were observed in α-1,2-Fuc, α-2,3-Sia, and α-2,6-Sia expression, with the latter showing an extremely significant difference ($p^{***} < 0.001$, n = 5) between HL7702 and HepG2. Importantly, with the assistance of Ln_{986}-MS2-PEG_{50}-N_3, only 89 HL7702 cells were required for detection using the most abundant α-GalNAc-^{164}Dy ($2.2 \pm 0.4 \times 10^{-19}$ mol/cell) (Fig. 2b), and merely 3 HepG2 cells were required for detection using the most abundant α-2,6-Sia-^{166}Er ($(7.2 \pm 0.9) \times 10^{-18}$ mol/cell) (Fig. 2f). The determined quantities of glycan-linkages for HL7702 and HepG2, along with their respective Ln-MS2-PEG50-lectin signal amplifier, are summarized in Table S3.

Following the quantification of glycan-linkages by bulk model ICP-MS, we extended our analysis to observe the distribution of glycan-linkages around the surface of single cells in both HL7702 and HepG2 cells. Mass distribution histograms of cell events under the single-cell model ICP-MS are presented in Figs. S8 and S9. These histograms reveal the cell-to-cell variabilities of glycan-linkage information for HL7702 (Fig. 3a) and HepG2 (Fig. 3b). As expected, the average single-cell glycan-linkage quantities in these two model cell lines, as determined by the bulk model, were consistent with the mean values obtained from the statistical results of 2000 valid single-cell events. The glycan-linkage content of HL7702 and HepG2 cells is displayed in a heat map (Fig. 3c). Apart from α-Gal and T antigen (metastasis marker), the quantities of five other glycan-linkages in HepG2 cells were significantly distinct from those in HL7702 cells ($p < 0.01$). Among these, α-1,6-Fuc (encoded with Gd) and α-2,6-Sia (Er) showed extremely significant differences between these two cell species ($p < 0.001$). These results

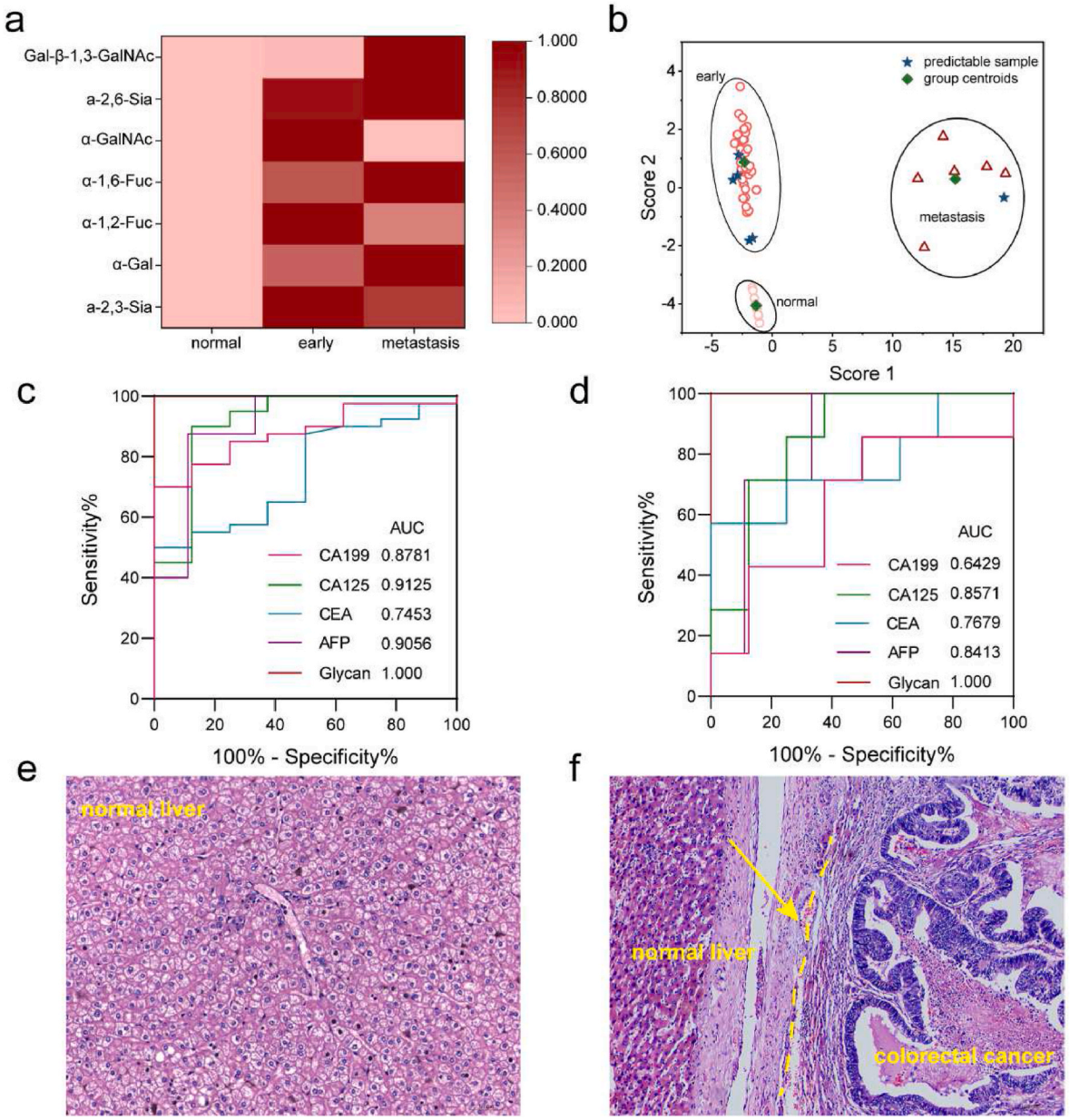


Fig. 4. (a) Normalized heat map of α-Gal, α-1.2-Fuc, α-1,6-Fuc, α-2,3-Sia, α-2,6-Sia, Gal-β-1,3-GalNAc in normal hepatic specimens, early HCC and colorectal metastasis specimens; (b) linear discriminate analysis of α-Gal, α-1.2-Fuc, α-1,6-Fuc, α-2,3-Sia, α-2,6-Sia, Gal-β-1,3-GalNAc as seven variables to distinguish and predict normal hepatic specimens, early HCC and colorectal metastasis specimens using the score of canonical discriminated functions. (c) ROC curve of CA199, CA125, CEA, AFP and glycan-linkages to distinguish 40 early HCC from 8; (d) ROC curve of CA199, CA125, CEA, AFP and glycan-linkages to distinguish 8 normal liver tissue from 7 colorectal metastasis hepatic specimens; (e) HE staining of normal liver tissues and (f) the boundary between colorectal metastasis and normal liver tissue.

align with the mass frequency distribution observed in SC-ICP-MS (Figs. S8 and S9) and suggest potential biomarkers for distinguishing between paracancerous and cancerous cells.

To further enhance the accuracy of distinguishing HCC cancerous cells, we applied linear discriminant analysis (LDA) using machine learning methods in this work. Specifically, seven HL7702 and seven HepG2 cell samples were used as the training set, and an additional four cell samples were used as the test set to predict their cell species. LDA was performed using IBM Statistics SPSS (edition 26), with the quantities of the seven glycan-linkages (N) as variables for distinguishing and predicting HL7702 and HepG2. After the LDA process, two LDA functions were generated, incorporating the quantities of these seven glycan-linkages to discriminate between HL7702 and HepG2 (see supporting information, **Function S1 and S2**). The scores of HL7702 and HepG2 derived from Function 1 and Function 2 are plotted in Fig. 3d, showing that all fourteen cell samples (100%) were accurately classified and distinguished between HL7702 and HepG2 cell species, with an extremely significant difference ($p < 0.0001$). These results suggest that the quantities of glycan-linkages on single cells, after LDA, can effectively differentiate cancerous HCC cells from their normal counterparts. Moreover, with the advantage of accurately distinguishing HL7702 and HepG2, a normalized canonical discriminant function was derived for predicting other test species ($n = 4$, including both HL7702 and HepG2). The normalized canonical discriminant function is defined as

$$\text{score } 3 = 0.026 \times N_{\alpha-\text{Gal}} + 0.127 \times N_{\alpha-\text{GalNAc}} + 0.126 \times N_{\alpha-1,2-\text{fucose}} + 0.19 \times N_{\alpha-1,6-\text{fucose}} + 0.205 \times N_{\alpha-2,3-\text{sialic aicd}} + 0.237 \times N_{\alpha-2,6-\text{sialic aicd}} + 0.087 \times N_{\text{Gal}-\beta-1,3-\text{GalNAc}}$$

From this equation, it's evident that all seven glycan-linkages contribute to cell discrimination, which is more reliable than relying on a single marker. Specifically, α-GalNAc (prefix coefficient 0.127), α-1,2-Fuc (0.126), α-1,6-Fuc (0.19), α-2,3-Sia (0.205), and α-2,6-Sia (0.237) predominantly contribute to hepatocellular tumorigenesis. These results align with our previous findings from bulk and single-cell ICP-MS determinations of glycan-linkages. In unclassified test cell samples, scores of the seven glycan-linkages from the canonical discriminant function <0 indicate HL7702, while scores >0 indicate HepG2. Using this function, discrimination accuracy reaches 100% in these samples, underscoring the contribution of the seven glycan-linkages. It's worth noting that using just one or two glycan-linkages, such as α-1,2 Fuc or α-2,6-Sia, results in discrimination accuracies of 91.6% and 95.6%, respectively. However, if α-Gal is employed as a biomarker (prefix coefficient), it's challenging to discriminate between HL7702 and HepG2, with an accuracy of no more than 50%. These results emphasize the advantage of using multiple glycan-linkages over a single biomarker. Therefore, our lanthanide-encoded signal amplification-based single-cell glycan-linkages linear discriminant analysis enables the analysis of multiple variables. This approach holds great promise for accurately distinguishing HCC cancerous cells from their normal counterparts in clinical HCC specimens, enabling early diagnosis.

Building on the successful single-cell sensitivity demonstrated in HL7702 and HepG2 cell experiments for glycan-linkages quantification using Ln-MS2-PEG_{50}-lectin and SC-ICP-MS, we proceeded to validate this approach for HCC early diagnosis using clinical specimens. We collected 55 surgically resected specimens, including 8 non-carcinoma normal hepatocellular tissue specimens, 40 early HCC specimens, and 7 cases of colorectal metastases to the liver (Fig. S10 and Table S4). After encoding with Ln_{986}-MS2-PEG_{50}-lectin, the glycan-linkages in non-carcinoma normal hepatocellular tissues were quantified. The average quantities of glycan-linkages were as follows: α-Gal $(1.5 \pm 0.2) \times 10^{-19}$ mol/cell, α-GalNAc $(1.9 \pm 0.1) \times 10^{-19}$ mol/cell, α-1,2-Fuc $(2.4 \pm 0.2) \times 10^{-20}$, α-1,6-Fuc $(2.1 \pm 0.2) \times 10^{-20}$, α-2,3-Sia $(1.4 \pm 0.2) \times 10^{-19}$, α-2,6-Sia $(1.8 \pm 0.2) \times 10^{-19}$, and Gal-β-1,3-GalNAc $(7.4 \pm 0.4) \times 10^{-20}$ mol/cell (n = 3). For the 40 early hepatic specimens, which included 16 high, 11 moderate, and 13 poorly differentiated hepatic specimens, the glycan-linkages were determined as follows: α-Gal $(1.9 \pm 0.2) \times 10^{-19}$ mol/cell, α-GalNAc $(1.3 \pm 0.1) \times 10^{-18}$ mol/cell, α-1,2-Fuc $(5.5 \pm 0.4) \times 10^{-19}$, α-1,6-Fuc $(3.1 \pm 0.3) \times 10^{-18}$, α-2,3-Sia $(2.6 \pm 0.3) \times 10^{-18}$, α-2,6-Sia $(4.4 \pm 0.5) \times 10^{-18}$, and Gal-β-1,3-GalNAc $(8.6 \pm 0.7) \times 10^{-19}$ mol/cell (Table S5). In the seven colorectal metastasis hepatic specimens, Gal-β-1,3-GalNAc was significantly up-regulated compared to normal and early HCC specimens. The differential quantities of glycan-linkages within normal, early, and metastasis hepatic specimens are visually displayed in a heat map (Fig. 4a). Abnormal expression of α-GalNAc, α-1,2-Fuc, α-1,6-Fuc, α-2,3-Sia, and α-2,6-Sia was significantly up-regulated in early HCC ($p < 0.001$).

Before conducting LDA analysis, the data on glycan-linkages were divided into three regions: region 1 (8 normal hepatocellular specimens), region 2 (33 early HCC specimens), and region 3 (6 colorectal metastasis hepatic specimens), serving as the training group. The remaining 5 early HCC and 1 colorectal metastasis hepatic specimens were set as the test group. Following the same LDA procedure as applied to HL7702 and HepG2, two normalized canonical discriminant functions were derived, incorporating glycan-linkages as seven variables to discriminate the three regions.

$$\text{score } 1 = 0.848 \times N_{\alpha-\text{Gal}} - 0.078 \times N_{\alpha-\text{GalNAc}} + 0.035 \times N_{\alpha-1,2-\text{Fuc}} - 0.013 \times N_{\alpha-1,6-\text{Fuc}} - 0.068 \times N_{\alpha-2,3-\text{Sia}} + 0.010 N_{\alpha-2,6-\text{Sia}} + 0.121 N_{\text{Gal}-\beta-1,3-\text{GalNAc}} - 2.461$$

$$\text{score } 2 = -1.552 \times N_{\alpha-\text{Gal}} + 0.022 \times N_{\alpha-\text{GalNAc}} + 0.020 \times N_{\alpha-1,2-\text{Fuc}} + 0.111 \times N_{\alpha-1,6-\text{Fuc}} + 0.255 \times N_{\alpha-2,3-\text{Sia}} - 0.108 N_{\alpha-2,6-\text{Sia}} + 0.002 N_{\text{Gal}-\beta-1,3-\text{GalNAc}} - 2.007$$

Based on the scores from these two canonical discriminant functions, region 1 (normal hepatocellular specimens), region 2 (early HCC specimens), and region 3 (colorectal metastasis HCC specimens) were effectively discriminated (Fig. 4b). The predictive accuracy was 100% for distinguishing early HCC from normal hepatocellular specimens and also for identifying metastasis HCC ($p < 0.001$). Notably, the biomarker Gal-β-1,3-GalNAc (prefix coefficient 0.121 in Function 1) made the most significant contribution to distinguishing colorectal metastasis from primary liver. All seven glycan-linkages contributed to the discrimination, serving as seven variables in IBM Statistic SPSS 26.

We next assessed the diagnostic performance of glycan-linkages and compared them with traditional clinical biomarkers by plotting a Receiver Operating Characteristic (ROC) curve. The ROC curve provides valuable insights into the accuracy of these diagnostic markers. As depicted in Fig. 4c, traditional clinical biomarkers, with the exception of CEA antigen, demonstrated notable diagnostic performance in distinguishing HCC from normal controls ($AUC_{AFP} = 0.9056$, $p < 0.001$; $AUC_{CEA} = 0.7453$, $p < 0.05$; $AUC_{CA125} = 0.9125$, $p < 0.001$; $AUC_{CA\ 199} = 0.8781$, $p < 0.001$). However, the sensitivity of these markers ranged from 20% to 65% at the chosen cutoff value, consistent with previous studies [10]. In stark contrast, glycan-linkages exhibited superior diagnostic ability compared to these traditional clinical biomarkers ($AUC_{\text{glycan-linkages}} = 1.000$, $p < 0.0001$). To discriminate HCC from normal controls, the optimal cutoff values were determined as follows: 7.9×10^{-19} mol/cell α-2,3-Sia, 1.9×10^{-19} mol/cell α-Gal, 1.2×10^{-19} mol/cell α-1,2-Fuc, 3.4×10^{-19} mol/cell α-1,6-Fuc, 2.4×10^{-19} mol/cell α-GalNAc, 4.7×10^{-19} mol/cell α-2,6-Sia, and 1.9×10^{-19} mol/cell Gal-β-1,3-GalNAc, resulting in a sensitivity of 100% and specificity of 100%.

Furthermore, we evaluated the diagnostic value of these biomarkers and glycan-linkages for distinguishing metastasis HCC from normal controls (Fig. 4d). The AUC of the glycan-linkages score for HCC detection was 1.000 ($p < 0.0001$), indicating higher AUC values, greater sensitivity, and specificity compared to AFP, CEA, CA125, and CA199. Importantly, unlike normal and early HCC specimens, Gal-β-1,3-GalNAc emerged as a distinctive marker for discriminating metastasis HCC in clinical therapy, offering enhanced accuracy and complementarity with immunohistochemistry (Fig. 4e and f).

In our investigation, we further explored the discrimination of 40 early HCC specimens, which had been histologically categorized as high (16 samples), moderate (11), and poor differentiation HCC (13) through immunohistochemistry (Table S6). Utilizing the same procedure, two canonical functions were derived from the glycan-linkages as seven variables.

$$\text{score } 1 = 0.903 \times N_{\alpha-\text{Gal}} + 0.164 \times N_{\alpha-\text{GalNAc}} + 0.332 \times N_{\alpha-1,2-\text{Fuc}} + 0.117 \times N_{\alpha-1,6-\text{Fuc}} + 0.032 \times N_{\alpha-2,3-\text{Sia}} + 0.045 \times N_{\alpha-2,6-\text{Sia}} + 0.228 \times N_{\text{Gal}-\beta-1,3-\text{GalNAc}} - 14.69$$

$$\text{score } 2 = 0.490 \times N_{\alpha-\text{Gal}} + 0.139 \times N_{\alpha-\text{GalNAc}} + 0.291 \times N_{\alpha-1,2-\text{Fuc}} - 0.182 \times N_{\alpha-1,6-\text{Fuc}} + 0.056 \times N_{\alpha-2,3-\text{Sia}} - 0.118 \times N_{\alpha-2,6-\text{Sia}} + 0.449 N_{\text{Gal}-\beta-1,3-\text{GalNAc}} + 1.169$$

As shown in Fig. 4a, when plotted with the scores from these two functions, the LDA results successfully discriminated the 16 high, 11

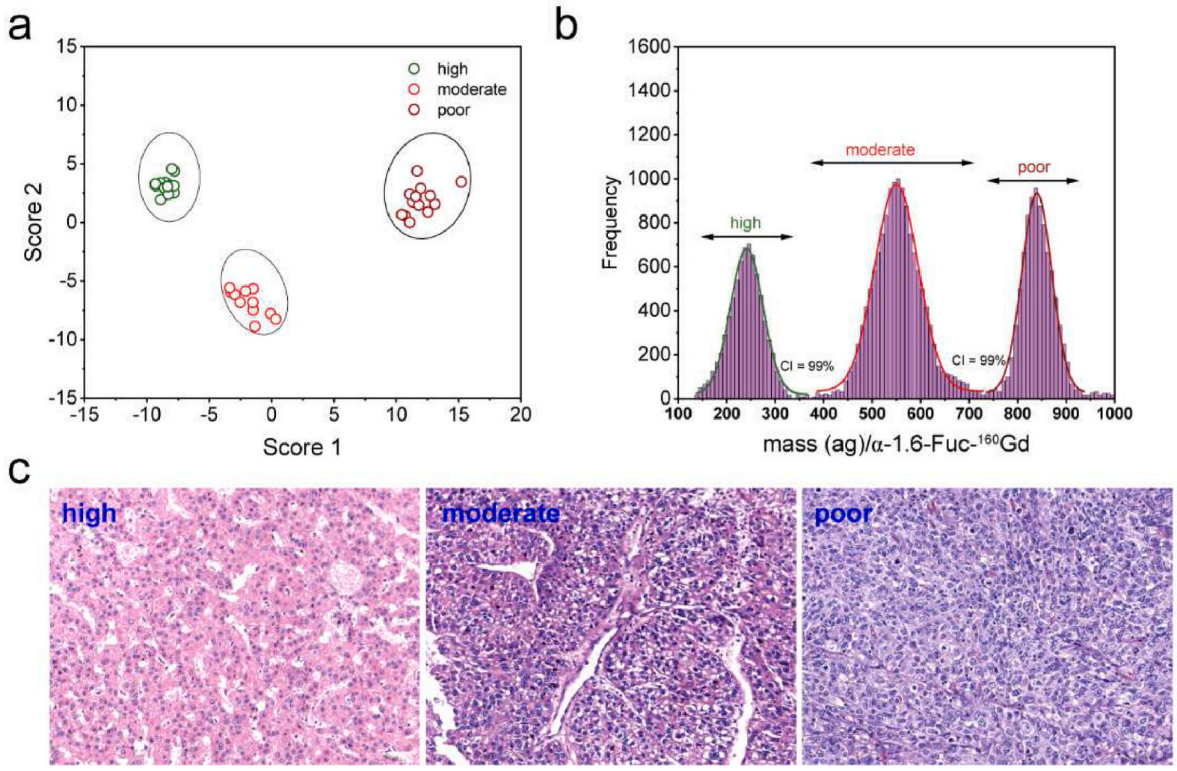


Fig. 5. (a) LDA analysis to discriminate high, moderate and poor differentiation HCC specimens ($p < 0.001$); (b) mass distribution of α-1,6-Fuc-^{160}Gd *vs* frequency to discriminate high, moderate and poor differentiation HCC in a single SC-ICP-MS run; (c) HE staining of high, moderate and poor differentiation HCC.

moderate, and 13 poor differentiation HCC specimens ($p < 0.001$). These glycan-linkages-based LDA findings were more distinctive than the results obtained through hematoxylin and eosin (HE) staining (Fig. 5a and c). The LDA results were also consistent with the mass distribution observed in SC-ICP-MS analysis, where the high, moderate, and poor differentiation HCC specimens were distinguishable when α-1.6-Fuc-^{160}Gd was used, with quantities of $(8.1 \pm 0.5) \times 10^{-19}$ mol/cell for high, $(4.5 \pm 0.5) \times 10^{-18}$ mol/cell for moderate, and $(5.3 \pm 0.9) \times 10^{-18}$ mol/cell for poor, demonstrating significant differences among them ($p < 0.01$) (Fig. 5b). To provide a clearer visualization, the mass (ag) quantified by SC-ICP-MS from α-1.6-Fuc-^{160}Gd was normalized into three distinct mass ranges (X-axis) versus the corresponding frequency (Y-axis). The high HCC specimens fell within the range of 100–400 ag/cell, moderate HCC specimens within 400–750 ag/cell, and poor HCC specimens within 750–950 ag/cell, and they could be effectively differentiated with a confidence incidence (CI) exceeding 98%. These findings underscore that our approach not only achieves single-cell sensitivity for glycan-linkages but also permits the classification of multidimensional data through LDA, enabling the discrimination of abnormal HCC cells using glycan-linkages as variables.

3. Conclusions

In this study, we have presented a novel approach which contributed to the early diagnosis of hepatocellular carcinoma (HCC) using single-cell glycan-linkages quantification through SC-ICP-MS, bolstered by the use of MS2 signal amplifiers. The incorporation of lanthanide-encoded MS2 has enabled an impressive signal amplification of nearly 1000 times, facilitating the quantification of glycan linkages expressed at extremely low levels on the surface of single cells (in the order of 10^{-19} mol/cell). We have further enhanced the discriminatory power of our approach by employing linear discriminant analysis (LDA) based on machine learning techniques. This analytical strategy yielded two Fisher canonical discriminant functions that employ the quantity of glycan-linkages as seven distinct biomarkers. When scores derived from these functions were plotted, early HCC specimens were effectively distinguished from normal cells, as well as metastatic HCC specimens. This level of single-cell sensitivity surpasses the capabilities of "soft" ionization mass spectrometry methods like ESI and MALDI. Moreover, it offers the significant advantage of processing multidimensional data through machine learning-driven linear discriminant analysis. It is noteworthy that our approach outperforms traditional methods such as the examination of cell morphology (e.g., cell boundary) and nuclear/cytoplasm ratio, as well as a series of immunohistochemically staining techniques. This superior accuracy in detecting single abnormal early HCC cells demonstrates the potential of our single-cell glycan linkage quantification method in clinical applications [36]. We anticipate that with the inclusion of more cell surface glycan linkage data and an expanded dataset of clinical patient specimens, the approach developed in this study will greatly enhance the feasibility and accuracy of early HCC diagnosis in clinical practice.

CRediT authorship contribution statement

Yong Liang: Writing – original draft, Visualization, Investigation, Data curation, Conceptualization. **Zhen Liu:** Writing – original draft, Software, Formal analysis. **Dongliang Zuo:** Validation, Resources. **Shi Chen:** Software, Formal analysis. **Jianbin Chen:** Software, Formal analysis. **Xiaowen Yan:** Formal analysis. **Pingguo Liu:** Project administration. **Qiuquan Wang:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We thank the financial supports from the National Natural Science Foundation of China (Grants 21904110), the National Key Research and Development Program of China (2022YFF0710200) and Key Laboratory of Spectrochemical Analysis & Instrumentation (Xiamen University), Ministry of Education-SCAI2004.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.talanta.2024.126056.

References

[1] A. Forner, J.M. Llovet, J. Bruix, Lancet 379 (2012) 1245–1255.
[2] J.D. Yang, P. Hainaut, G.J. Gores, A. Amadou, A. Plymoth, L. Roberts, Nat. Rev. Gastroenterol. Hepatol. 16 (2019) 589–604.
[3] H.B. El-Serag, L.K. Rudolph, Gastroenterology 132 (2007) 2557–2576.
[4] J.M. Llovet, J. Zucman-Rossi, E. Pikarsky, B. Sangro, M. Schwartz, M. Sherman, G. Gores, Nat. Rev. Dis. Prim. 2 (2016) 16018.
[5] K.C. Han, J.H. Kim, World J. Gastroenterol. 21 (2015) 10327–10335.
[6] T. Couri, A. Pillai, Hepatol. Int. 13 (2019) 125–137.
[7] N. Tsuchiya, Y. Sawada, I. Endo, K. Saito, Y. Uemura, T. Nakatsura, World J. Gastroenterol. 21 (2015) 10573–10583.
[8] W.Y. Wang, C. Wei, Genes Dis. 7 (2020) 308–319.
[9] T. Ichikawa, K. Sano, H. Morisaka, Liver Cancer 3 (2014) 97–107.
[10] B.I. Choi, J.M. Lee, T.K. Kim, M.D. Burgio, V. Vilgrain, AJR Am. J. Roentgenol. 1 (2015) 10–21.
[11] T. Wang, K.H. Zhang, Front. Oncol. 10 (2020) 1316.
[12] Z.J. Liu, Y.W. Pu, Y.X. Bao, S. He, Int. J. Gen. Med. 14 (2021) 4369–4380.
[13] N.P. Chalasani, T.S. Ramasubramanian, A. Bhattacharya, M.C. Olson, V.D. K. Edwards, L.R. Roberts, J.B. Kisiel, K.R. Reddy, G.P. Lidgard, S.C. Johnson, J. Bruinsma, J. Clin. Gastroenterol Hepatol. 19 (2021) 2597–2605.
[14] P. Kotwani, W. Chan, F. Yao, N. Metha, Clin, Gastroenterol. Hepatol. 20 (2022) 701–703.
[15] Y.X. Pan, X.Q. Sun, Z.L. Hu, W. Xie, K.X. Nie, A.P. Fang, Y.Y. Zhang, Y.Z. Fu, J. B. Chen, J.C. Wang, X. Wang, Y.J. Zhang, D.D. Hu, M.S.J. Chen, Hepatocell Carcinoma 8 (2021) 657–670.
[16] Y.K. Wen, S.S. Jeong, Q. Xia, X.N. Kong, Int J. Biol. Sci. 12 (2016) 1121–1128.
[17] Y.N. Wang, Y.J.Y. Wan, Liver Res. 4 (2020) 161–167.
[18] A. Gil-Gómez, Á. Rojas, C.H. Liu, R. Gallego-Duran, R. Muńoz-Hernandez, G. Fassina, P. Pontisso, J. Ampuero, M. Romero-Gómez, World J. Gastroenterol. 27 (2021) 8343–8356.
[19] N. Ei-Abd, A. Fawzy, T. Elbaz, S. Hamdy, Tumour Biol. 37 (2016) 211–216.
[20] D.D. Li, J.W. Zhang, J.M. Li, Clin. Chim. Acta 500 (2020) 10–19.
[21] S. Zhang, X.Y. Cao, Q. Gao, Y.K. Liu, Cancer Lett. 406 (2017) 64–70.
[22] M. Schneider, E. Al-Shareffi, R.S. Haltiwanger, Glycobiology 27 (2017) 601–618.
[23] O.J. Adams, M.A. Stanczak, S.V. Gunten, H. Läubli, Glycobiology 28 (2018) 640–647.
[24] V.D. Guruaribam, T.J. Sarumathi, Cancer Res. Ther. 16 (2020) 401–404.
[25] K.K. Palaniappan, C.R. Bertozzi, Chem. Rev. 116 (2016) 14277–14306.
[26] S. Gaunitz, G. Nagy, N.L.B. Pohl, M.V. Novotny, Anal. Chem. 89 (1) (2017) 389–413.
[27] R.L. Ruhaak, G.G. Xu, Q.Y. Li, E. Goonatilleke, C.B. Lebrilla, Chem. Rev. 118 (2018) 7886–7930.
[28] Y. Liang, L.M. Yang, Q.Q. Wang, Appl. Spectrosc. Rev. 51 (2016) 117–128.
[29] W.I. Weis, K. Anna Drickamer, Rev. Biochem. 65 (1996) 441–473.
[30] J. Hirabayashi, Glycoconj. J. 21 (2004) 35–40.
[31] R. Liu, S.X. Zhang, C. Wei, Z. Xing, S.C. Zhang, X.R. Zhang, Acc. Chem. Res. 49 (5) (2016) 775–783.
[32] Z.R. Liu, X.T. Li, G.Y. Xiao, B.B. Chen, B. Hu, Anal. Chim. Acta 1137 (2017) 191–207.
[33] X. Wei, Y. Lu, X. Zhang, M.L. Chen, J.H. Wang, TrAC, Trends Anal. Chem. 127 (2020) 115886.
[34] Sanz-Medel, A. Anal. Bioanal. Chem. 408, 5393–5395.
[35] R. Yuan, F.C. Ge, Y. Liang, Y. Zhou, L.M. Yang, Q.Q. Wang, Anal. Chem. 91 (2019) 4948–4952.
[36] Y.S. Kim, M.J. Jo, X.L. Luo, T.P. Prakash, T.Y. Zhou, S. Klein, X.K. Xiao, N. Post, Z. F. Yin, A.R. MacLeod, Mol. Ther. 27 (2019) 1547–1557.