Prognosis prediction of Hepatocellular Carcinoma After Surgical Resection based on Serum Metabolic Profiling from Gas Chromatography-Mass Spectrometry

Chengnan Fang
CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China

Benzhe Su
School of Computer Science & Technology, Dalian University of Technology, Dalian 116024, China.

Tianyi Jiang
International Cooperation Laboratory on Signal Transduction, Eastern Hepatobiliary Surgery Institute, The Second Military Medical University, Shanghai 200438, China

Chao Li
School of Computer Science & Technology, Dalian University of Technology, Dalian 116024, China

Yexiong Tan
International Cooperation Laboratory on Signal Transduction, Eastern Hepatobiliary Surgery Institute, The Second Military Medical University, Shanghai 200438, China.

Qingqing Wang
CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China.

Liwei Dong
International Cooperation Laboratory on Signal Transduction, Eastern Hepatobiliary Surgery Institute, The Second Military Medical University, Shanghai 200438, China.

Xinyu Liu (liuxy2012@dicp.ac.cn)
Dalian Institute of Chemical Physics https://orcid.org/0000-0002-0564-7167

Xiaohui Lin
School of Computer Science & Technology, Dalian University of Technology, Dalian 116024, China

Guowang Xu
CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China

Research

Keywords: Metabolomics, Hepatocellular carcinoma, Prognosis, Network-based metabolic feature selection, Risk score

Posted Date: September 1st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-63025/v1

Abstract

Background In recent years, clinical diagnosis and surgical techniques have developed rapidly, however the long-term survival of hepatocellular carcinoma (HCC) has not been significantly enhanced. Comprehensive prediction of the prognostic risk of surgical treatment is particularly important for guiding clinical decisions and improving survival. The objective of our study was to build prognostic models based on serum metabolomics data, and assess the prognostic risk of HCC within five years after surgical resection.

Methods A pseudotargeted gas chromatography-mass spectrometry (GC-MS) based metabolomics method was applied to analyze serum profiling of 78 patients with HCC. Important metabolic features with discriminant ability for postoperative outcome were identified by a novel network-based metabolic feature selection method based on composite significance index (N-CSI). The critical metabolites defined by N-CSI were further evaluated by Cox regression analyses to identify metabolite panel associated with prognosis of HCC.

Results Phenylalanine and galactose were identified to be relevant with mortality, and while galactose and tyrosine were associated with recurrence and metastasis. These two panels of metabolites were considered to respectively generate models to predict risk of mortality (Risk score of overall survival; RS_{OS}) and risk of recurrence and metastasis (Risk score of disease-free survival; RS_{DFS}). Low-risk and high-risk groups for overall survival and disease-free survival were both clearly stratified (p<0.05). Multivariate analysis showed that RS_{OS} (p=0.002, HR=2.57) and RS_{DFS} (p=0.015, HR=2.33) were independent prognostic factors for predicting postoperative overall survival and disease-free survival respectively. The performance of models was not only internally validated by bootstrapping with 1000 resampling, but also validated in an external independent cohort.

Conclusions The risk scores developed were independent prognostic factors of HCC, which present favorable ability to predict overall survival and disease-free survival respectively, especially when combined with clinical staging system. This study demonstrated that metabolomics is a powerful tool for risk screening of HCC prognosis.

Background

Hepatocellular carcinoma (HCC) is the most common type of primary liver cancer, comprising 75%-85% of cases. New cases and deaths of liver cancer accounted for 4.7% and 8.2% of the total number of cancer cases in the world in 2018, respectively [1]. Clinical therapeutics and diagnostics have been rapidly developed in recent years. At present, surgical resection is still the best method for the treatment of patients with early HCC. In addition, transcatheter arterial chemoembolization, liver transplantation, stereotactic radiotherapy, radiofrequency ablation and other therapies also provide more options for treatment [2–4]. However, the long-term survival of HCC has not been significantly enhanced [4]. The high recurrence and metastasis rates seriously affect the surgical efficacy and survival of patients, which is an important factor causing poor prognosis of surgical treatment [5, 6]. Therefore, effective prediction of the prognostic risk of surgical treatment is particularly important and in urgent need for guiding clinical decisions and improving survival.

Metabolomics performs analysis of all small molecule metabolites with the goal of discovering similarities and differences of metabolic characteristics among biological samples. And it is widely used in disease marker screening and mechanisms exploring [7–9]. Some studies demonstrated that metabolites play important roles in tumorigenesis and development, and are associated with cancer prognosis [6, 10–14], although the detailed prognostic mechanism of HCC has not been completely elucidated. Multiple clinical staging systems, such as tumor-node-metastasis (TNM) and the Barcelona Clinic Liver Cancer (BCLC), have been established for guiding prognostication of HCC, but there is no consensus regarding the optimal risk stratification tool. Hence, it is undoubtedly worth exploring metabolic markers to provide more precisely personalized treatment and comprehensive risk monitoring for patients after HCC surgery.

Mining the crucial information from metabolomics data is important, and many feature selection methods have been developed in metabolomics study. Network biology provides a novel perspective for understanding the molecule interactions for complex diseases at the system level [15]. It is well known that selecting hub nodes has been a common and popular way in network-based disease data analysis [16, 17]. However, molecules do not exist independently in life activities. They interact with each other to reflect the physiological and pathological changes of the body. Selecting hub nodes which ignores the partners that interact with them may lose important information. Hence, searching the interesting sub-networks based on the important nodes which integrates the hub nodes and node interactions together may be more efficient to define valuable information for disease studies.

In this study, serum metabolic profiling was firstly investigated by using gas chromatography-mass spectrometry (GC-MS) based pseudotargeted metabolomics method, then a novel network-based metabolic feature selection method based on composite significance index (N-CSI) was developed to define metabolic features associated with postoperative outcome. Finally, two models that could predict the prognostic risk of HCC surgery were built to predict the overall survival and disease-free survival within five years. The performance of models was not only internally validated by bootstrapping with 1000 resampling, but also validated in an external independent cohort. The result of this study could thereby provide a valuable guidance for individualized clinical treatments.

Materials And Methods

Study design

This study aimed at identifying metabolic risk factors in serum for prognosis of HCC. According to the outcome of patients, two events were focused, namely survival or not (Survivors *vs* Non-survivors, S *vs* NS), and recurrence/metastasis or not (No recurrence/metastasis *vs* Recurrence/metastasis, NR *vs* R). Crucial metabolic features associated with prognostic outcome were identified by a novel method named as N-CSI. Then the critical metabolites defined by N-CSI were further evaluated by Cox regression analyses to define metabolite panel associated with prognosis of HCC. Two prognostic models were built and subsequently subjected to internally and externally validation. The flow chart of analysis process is shown in Fig. 1.

Patients and clinical specimens

A total of 78 preoperative serum samples were collected from patients undergoing curative resection of HCC at the Eastern Hepatobiliary Surgery Institute of the Second Military Medical University (Shanghai, China) from June 2012 to July 2012, and then stored at -80 °C until used for scientific research. Likewise, another 91 serum samples collected from July 2013 to June 2014 were enrolled as an external validation cohort. All patients were followed up for 5 years after surgery. At the end of follow up, 38 (48.7%) and 37 (40.7%) patients died in the discovery and validation cohorts, respectively, and overall survival was defined as the time from the date of curative resection to the time of death, while the latest follow-up time was counted as the endpoint of overall survival for subjects who were not dead during the follow-up period. Sixty-three patients with specific information of recurrence and metastasis in the discovery cohort, and all 91 samples in the validation cohort were enrolled to study the disease-free survival. Among these patients, recurrence and metastasis occurred in 34 (54.0%) and 13 (14.3%) cases, respectively. Disease-free survival was defined as the time from the date of curative resection to the first diagnosis of recurrence or metastasis. For the patients without recurrence or metastasis, the latest follow-up time was considered as the endpoint of disease-free survival.

The baseline clinical and pathological characteristics of all patients are described in details in Table 1, supplementary Table S1 and Table S2. Tumor stages were defined according to the 8th edition tumor-node-metastasis (TNM) classification of the American Joint Committee on Cancer (AJCC)/Union for International Cancer Control (UICC), and the Barcelona Clinic Liver Cancer (BCLC) staging system as well. The overall and disease-free survival curves of all patients in discovery and validation cohorts can be seen in supplementary Fig. S1.

Table 1

Baseline characteristics of all patients. [a, b, c]

Characteristics	Discovery cohort (n = 78)	Validation cohort (n = 91)
Age, years	51 (10.2)	49 (10.9)
Gender, Male/Female	70 (89.7)/8 (10.3)	77 (84.6)/14 (15.4)
Smoking, Yes/No	34 (43.6)/44 (56.4)	45 (49.5)/46 (50.5)
Preoperative HBV-DNA level, copies/mL, $>10^3/<10^3$	35 (44.9)/41 (52.6), n = 76	37 (41.6)/52 (58.4), n = 89
Hepatitis B, Positive/Negative	70 (89.7)/8 (10.3)	74 (81.3)/17 (18.7)
Cirrhosis, Presence/Absence	44 (56.4)/34 (43.6)	21 (23.1)/70 (76.9)***
Preoperative AFP level, μg/L, > 400/<400	30 (38.5)/48 (61.5)	37 (40.7)/54 (59.3)
Preoperative ALP level, U/L	84 (37–335), n = 77	79.5 (26–911), n = 88
Preoperative GGT level, U/L	77 (16–457), n = 77	50.5 (14–438), n = 88
Preoperative Bilirubin, μmol/L	13.3 (5.8–49.1)	13.2 (6.1–37.5)
Preoperative Albumin, g/L	41.8 (5.3)	41.8 (3.8)
TNM Stage, Stage I/Stage II/Stage IIIa/Stage IIIb	28 (36.4)/23 (29.9)/9 (11.7)/17 (22.1), n = 77	48 (52.7)/16 (17.6)/6 (6.6)/21 (23.1)
BCLC Stage, Stage 0/Stage A/Stage B/Stage C	2 (2.7)/42 (56.0)/14 (18.7)/17 (22.7), n = 75	9 (9.9)/54 (59.3)/7 (7.7)/21 (23.1)
Maximum tumor diameter, cm	4.9 (1.2–17.6), n = 77	5.1 (1.1–17.8)
Tumor number, ≥ 2/1	56 (72.7)/21 (27.3)	11 (12.1)/80 (87.9)***
Microvascular invasion, Presence/Absence	36 (46.2)/42 (53.8)	34 (37.4)/57 (62.6)

[a] Age and preoperative albumin are expressed as average (SD). Preoperative GGT level, preoperative ALP level, preoperative bilirubin and largest tumor diameter are expressed as median (range). Other characteristics are expressed as number (proportion %). [b] *** represent *p* value of less than 0.001. [c] n is as indicated in the column headings unless otherwise state. AFP, alpha-fetoprotein; GGT, gamma-glutamyltransferase; ALP, alkaline phosphatase; SD, standard deviation.

Metabolomics analysis

Serum metabolomics analysis was performed by using gas chromatography-mass spectrometry (GC-MS) based pseudotargeted metabolomics method [14, 18–20]. Fourty μL of serum sample was mixed with 160 μL methanol containing 5 μg/mL internal standards (valine-d8, succinic acid-d4, phenylalanine-d5, tridecanoic acid and citric acid-d3) and then vortexed for 30 s. Supernatants were lyophilized followed by oximation and silylation reactions for subsequent GC-MS analysis. QC samples generated by pooling real samples were

processed with the same method as the real samples, and inserted into the analytical batch every eight samples to evaluate the data quality.

Metabolic profiling was acquired on a QP 2010 plus GC-MS system coupled with AOC-20i automatic sampler (Shimadzu, Japan), and used column was DB-5MS fused-silica capillary column (30 m × 0.25 mm × 0.25 μm, Agilent Technologies, Santa Clara, CA, USA). Detailed process and system parameters of the pseudotargeted metabolomics method has been published in our previous studies [14, 18–20]. A total of 163 metabolic characteristic ions from 134 metabolites were quantified and corrected by internal standards. The system parameter settings have been described previously [19]. Metabolite identification was performed based on commercial standard library (NIST, Wiley, Fiehn and Mainlib) and an internal metabolite library.

Network-based metabolic feature selection

Feature ratios could be considered as the presumptive pathway reactions from reactants to products [16]. The drastic changes of feature ratios between two conditions could characterize the changes of pathway reactions during the disease development. N-CSI constructed the network based on the feature ratios with significant differences ($p < 0.05$, *t*-test) between two groups. The nodes were evaluated by the composite significance index (CSI) which combined the distinguishing ability of a node itself and the interactions between the node and its adjacent nodes in the network. Top scored nodes were selected as seed nodes, and then N-CSI searched subnetworks with crucial distinguishing information from each seed node. The subnetwork with the highest AUC in the searching process was defined as the key informative subnetwork. The workflow of N-CSI is shown in Supplementary Fig. S2. And the detail of network construction and the key informative subnetwork identification are given in supplementary materials.

Statistical analyses

Principal components analysis (PCA) was performed by SIMCA-P 11.0 program (Umetrics, Umea, Sweden) to observe clustering degree of QCs and evaluate the quality of the data. The developed method N-CSI was employed to select crucial metabolic features with a discriminant ability for postoperative outcome. Then the metabolites defined by N-CSI were further evaluated by Cox regression analyses to identify metabolite panels associated with prognosis of HCC, and construct prognostic risk models. The predicted overall survival rate and disease-free survival rate were visualized by nomograms, and the Harrell's concordance index (C-index) was used to evaluate their ability to discriminate outcomes of patients. Discrimination and calibration of the models were assessed via time-dependent receiver operating characteristic (tdROC) curves and calibration curves. Above statistical analyses were implemented by R 3.6.2 version. Kaplan-Meier curves and Logrank test results were generated by MedCalc 19.0.4 Software (MedCalc Software, Ostend, Belgium) to evaluate the impact of the metabolic features on the overall survival and disease-free survival. For categorical variables, the proportional hazards assumption was verified by examining the plots of cumulative hazard function against time. For continuous variables, the proportional hazards assumption was tested by examining the Schoenfeld residual scatter plot, further testing the correlation between Schoenfeld residuals and time rank. The test of proportional hazards assumption and Cox regression of risk scores with clinical parameters were completed by SPSS 25.0 software (IBM SPSS, Chicago, Illinois, USA).

Results

Global metabolic profiling analyzed by GC-MS

The serum metabolic profiling of HCC patients was obtained by the pseudotargeted metabolomics method. The PCA score scatter plot based on 163 characteristic ions from 134 metabolites was constructed, which showed that QC samples were tightly clustered (see Supplementary Fig S3A). In QC samples, 88.8% of metabolites had RSD% less than 30%, the corresponding peak area accounted for 91.3% of the total peak area (see Supplementary Fig S3B), and the RSD% of the total peak area in all QC samples was only 2.4%, which showed that the data quality of metabolic profiling was good. Metabolites with RSD% less than 30% in QC samples were selected for further analysis.

Construction of risk prediction models for overall survival and disease-free survival

N-CSI was employed to identify the crucial discrimination information of two predefined events (S *vs* NS, NR *vs* R) in the fifty times repeated five-fold cross-validation, respectively. The sum of occurrences of each metabolite in all subnetworks generated by cross-validation was counted as the frequency. The metabolites with frequency greater than 30 were subjected to univariable Cox proportional hazards regression analysis to identify potential prognostic related metabolites. Among them, phenylalanine and galactose were found to be relevant with mortality, while galactose and tyrosine were associated with recurrence and metastasis ($p < 0.05$). These two panels of metabolites were subjected to generating risk prediction models based on the coefficients by the stepwise multivariable Cox regression analysis respectively (Table 2, Table 3). The risk score for predicting risk of mortality was constructed:

Table 2

Cox regression analyses of metabolites associated with overall survival. [a]

Variable	Univariate			Multivariate		
	Coefficient	HR (95% CI)	p value	Coefficient	HR (95% CI)	p value
Phenylalanine*	0.436	1.55 (1.06–2.25)	0.023	0.396	1.49 (1.04–2.12)	0.029
Galactose*	0.297	1.35 (1.08–1.68)	0.009	0.292	1.34 (1.06–1.68)	0.013

[a] * represents *p* value of less than 0.05. HR, Hazard ratio. CI, confidence interval.

Table 3

Cox regression analyses of metabolites associated with disease-free survival. [a]

Variable	Univariate			Multivariate		
	Coefficient	HR (95% CI)	p value	Coefficient	HR (95% CI)	p value
Galactose*	0.295	1.34 (1.05–1.72)	0.019	0.322	1.38 (1.07–1.77)	0.012
Tyrosine*	0.246	1.28 (1.00-1.63)	0.047	0.278	1.32 (1.03–1.70)	0.030

[a] * represents *p* value of less than 0.05. HR, Hazard ratio. CI, confidence interval.

RS_{OS} = (0.396 × relative concentration of Phenylalanine) + (0.292 × relative concentration of Galactose).

The risk score for predicting risk of recurrence and metastasis was constructed:

RS_{DFS} = (0.322 × relative concentration of Galactose) + (0.278 × relative concentration of Tyrosine).

Prognostic performance of the models

The Risk Score (RS) of overall survival and disease-free survival was calculated according to above models, respectively. Using the median value of RS_{OS} (2.33) as the cut-off value, low-risk (n = 39) and high-risk (n = 39) groups of mortality were clearly stratified ($p < 0.05$), which indicated that overall survival of the low-risk group was significantly better than that of the high-risk group (Fig. 2A). The 1, 2 and 5-year overall survival rates for patients with low-RS_{OS} were 84.6%, 76.9%, 61.5% respectively, and 69.2%, 56.4%, 41.0% respectively for patients with high-RS_{OS}. According to the mortality risk prediction model, the overall survival rate of an individual patient could be quantified by combining two related metabolites as shown in the nomogram (Fig. 3A). The C-index of nomogram obtained by bootstrapping with 1000 resampling to predict 5-year overall survival rate was 0.623 (95%CI 0.576–0.670), and the 1, 2 and 5-year time-dependent AUCs (tdAUCs) of RS_{OS} for overall survival prediction were 0.611, 0.637 and 0.614, respectively (Fig. 2C).

Using the median value of RS_{DFS} (2.15) as the cut-off value, low-risk (n = 32) and high-risk (n = 31) groups of recurrence and metastasis were clearly stratified ($p < 0.05$), which suggested that disease-free survival of the low-risk group was significantly better than that of the high-risk group (Fig. 2B). Besides, the 1, 2 and 5-year disease-free survival rates for patients with low-RS_{DFS} were 75.0%, 68.8%, 59.4% respectively, and 45.2%, 38.7%, 32.3% respectively for patients with high-RS_{DFS}. According to the recurrence and metastasis risk prediction model, the disease-free survival rate of an individual patient could be quantified by combining two related metabolites as shown in the nomogram (Fig. 3B). The C-index of nomogram obtained by bootstrapping with 1000 resampling to predict 5-year disease-free survival rate was 0.660 (95%CI: 0.605–0.716), and the 1, 2 and 5-year tdAUCs of RS_{DFS} for disease-free survival prediction were 0.624, 0.640 and 0.660 respectively (Fig. 2D).

Models were calibrated by comparing probability of actual survival versus nomogram-predicted survival via calibration curves, after bootstrapping with 1000 resampling for internal verification (Fig S4). The calibration curves were close to the baseline, showing a good prediction agreement.

Relationship between risk scores and clinical pathological parameters

To investigate the correlation of risk scores and clinical pathological parameters with prognostic survivals, univariate and multivariable Cox regression analyses were performed.

Univariate analysis showed that overall survival was significantly relevant with RS_{OS}, alpha-fetoprotein (AFP), alkaline phosphatase (ALP), gamma-glutamyltransferase (GGT), albumin, maximum tumor diameter, tumor number, microvascular invasion, TNM stage and BCLC stage ($p < 0.05$). After all relevant clinical variables were incorporated into multivariate Cox regression, RS_{OS} remained an independent prognostic factor for overall survival ($p = 0.002$, HR = 2.57, 95%CI: 1.41–4.67). In addition, AFP ($p = 0.021$, HR = 2.21, 95%CI: 1.12–4.36) and TNM stage ($p < 0.001$) were also significant mortality risk predictors (Table 4).

Table 4
Univariate and multivariate Cox regression analyses of risk scores and clinical pathological parameters associated with overall survival and disease-free survival. [a, b]

Variable	Overall survival (n = 78)				Disease-free survival (n = 63)			
	Univariate		Multivariate		Univariate		Multivariate	
	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value
Risk scores	2.74 (1.54, 4.87)	0.001	2.57 (1.41, 4.67)	0.002	2.53 (1.35, 4.71)	0.004	2.33 (1.18, 4.60)	0.015
Cirrhosis, Presence/Absence	-	-	-	-	3.88 (1.74, 8.67)	0.001	2.99 (1.27, 7.06)	0.012
Preoperative AFP level, μg/L, > 400/<400	2.50 (1.31, 4.78)	0.006	2.21 (1.12, 4.36)	0.021	2.36 (1.18, 4.71)	0.015	-	-
Preoperative ALP level, U/L	1.01 (1.01, 1.02)	< 0.001	-	-	1.01 (1.00, 1.02)	0.002	-	-
Preoperative GGT level, U/L	1.00 (1.00, 1.01)	0.016	-	-	1.00 (1.00, 1.01)	0.015	-	-
Preoperative albumin, g/L	0.91 (0.86, 0.97)	< 0.001	-	-	0.86 (0.79, 0.93)	< 0.001	-	-
Maximum tumor diameter, cm	1.18 (1.09, 1.28)	< 0.001	-	-	1.20 (1.06, 1.36)	0.004	-	-
Tumor number, ≥ 2/1	1.96 (1.01, 3.83)	0.048	-	-	2.25 (1.11, 4.56)	0.025	-	-
Microvascular invasion, Presence/Absence	2.40 (1.24, 4.65)	0.009	-	-	2.89 (1.44, 5.82)	0.003	-	-
TNM Stage[c]	-	< 0.001	-	< 0.001	-	< 0.001	-	< 0.001

[a] Four clinicopathological parameters, namely gender, smoking, preoperative HBV-DNA level and preoperative bilirubin were excluded because they did not meet proportional hazards assumption. Two parameters namely Age and hepatitis C were not significantly associated with overall survival and disease-free survival. [b] AFP, alpha-fetoprotein; GGT, gamma-glutamyltransferase; ALP, alkaline phosphatase; HR, Hazard ratio; CI, confidence interval. [c] Grade I was used as the reference group. [d] Grade 0 was used as the reference group.

Variable	Overall survival (n = 78)				Disease-free survival (n = 63)			
	Univariate		Multivariate		Univariate		Multivariate	
	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value
TNM Stage II	2.32 (0.82, 6.51)	0.111	2.02 (0.71, 5.73)	0.186	1.82 (0.64, 5.18)	0.264	1.55 (0.53, 4.57)	0.427
TNM Stage IIIa	8.53 (2.90, 25.04)	< 0.001	6.65 (2.21, 20.07)	0.001	7.13 (2.33, 21.83)	0.001	8.15 (2.50, 26.54)	< 0.001
TNM Stage IIIb	8.23 (3.10, 21.85)	< 0.001	5.99 (2.17, 16.51)	0.001	14.36 (5.03, 40.99)	< 0.001	15.22 (5.00, 46.37)	< 0.001
BCLC Stage[d]	-	< 0.001	-	-	-	< 0.001	-	-
BCLC Stage A	0.40 (0.05, 3.09)	0.380	-	-	0.39 (0.05, 3.06)	0.374	-	-
BCLC Stage B	1.49 (0.19, 11.68)	0.703	-	-	1.65 (0.21, 13.29)	0.637	-	-
BCLC Stage C	2.32 (0.30, 17.72)	0.418	-	-	4.84 (0.60, 39.00)	0.138	-	-

[a] Four clinicopathological parameters, namely gender, smoking, preoperative HBV-DNA level and preoperative bilirubin were excluded because they did not meet proportional hazards assumption. Two parameters namely Age and hepatitis C were not significantly associated with overall survival and disease-free survival. [b] AFP, alpha-fetoprotein; GGT, gamma-glutamyltransferase; ALP, alkaline phosphatase; HR, Hazard ratio; CI, confidence interval. [c] Grade I was used as the reference group. [d] Grade 0 was used as the reference group.

RS_{OS} was combined with clinical classification systems including TNM stage and BCLC stage to test whether RS_{OS} could improve the discrimination capability. The 1, 2 and 5-year tdAUCs of the combined model (TNM and RS_{OS}) were 0.793, 0.795, 0.817 respectively, and 0.764, 0.740, 0.787 respectively for the combined model (BCLC and RS_{OS}) (Fig. 4A-C), which proved that RS_{OS} could improve the performance of clinical classification systems and provide good predictive ability for overall survival of HCC. Similar results were obtained in the validation cohort. The 1, 2 and 5-year tdAUCs of the combined model (TNM and RS_{OS}) were 0.774, 0.769, 0.757 respectively, and 0.774, 0.765, 0.723 respectively for the combined model (BCLC and RS_{OS}) (Fig. 4G-I).

Univariate analysis showed that disease-free survival was significantly relevant with RS_{DFS}, cirrhosis, AFP, ALP, GGT, albumin, maximum tumor diameter, tumor number, microvascular invasion, TNM stage and BCLC stage

($p < 0.05$). After all relevant clinical variables were incorporated into multivariate cox regression, RS_{DFS} remained an independent prognostic factor for disease-free survival ($p = 0.015$, HR = 2.33, 95%CI: 1.18–4.60). In addition, Cirrhosis (Presence/Absence) ($p = 0.012$, HR = 2.99, 95%CI: 1.27–7.06), and TNM stage ($p < 0.001$) were also significant recurrence and metastasis risk predictors (Table 4).

Similarly, RS_{DFS} was combined with clinical classification systems including TNM stage and BCLC stage to test whether RS_{DFS} could improve the discrimination capability. The 1, 2 and 5-year tdAUCs of the combined model (TNM and RS_{DFS}) were 0.888, 0.909, 0.850 respectively, and 0.901, 0.912, 0.840 respectively for the combined model (BCLC and RS_{DFS}) (Fig. 4D-F), which proved that RS_{DFS} could improve the performance of predictive power for HCC prognosis. In validation cohort, the 1, 2 and 5-year tdAUCs of the combined model (TNM and RS_{DFS}) were 0.622, 0.622, 0.817 respectively, and 0.662, 0.662, 0.809 respectively for the combined model (BCLC and RS_{DFS}) (Fig. 4J-L).

Discussion

In present study, two predictive models were built and validated based on preoperative serum metabolic profiling of HCC patients to assess surgical prognosis. It could be known from survival analysis that high RS_{OS} and RS_{DFS} were associated with poor prognosis of HCC.

Increasing evidences have revealed great potential of metabolites identified based on metabolomics serve as biomarkers for cancer diagnosis, prognosis and treatment. For example, it has been reported that the panel of retinol and retinal was identified as an independent predictor of survival time of HCC [13]. According to our results, phenylalanine (HR = 1.55, 95%CI = 1.06–2.25) and galactose (HR = 1.35, 95%CI = 1.08–1.68) were shown to be hazard factors for death after HCC surgery, while galactose (HR = 1.34, 95%CI = 1.05–1.72), and tyrosine (HR = 1.28, 95%CI = 1.00-1.63) were hazard factors for recurrence and metastasis. Although it was incapable to speculate the specific molecular mechanism of prognosis by these metabolites, existing research has confirmed that they specifically concerned the occurrence and development of HCC. Ye et al. [14] defined phenylalanine and 4 other metabolites (ethanolamine, acotinic acid, lactic acid, and ribose) from the urinary metabolic profiling favorably to predict early recurrence of HCC. Wang et al. [22] generated a prognostic metabolite panel (phenylalanine and choline) for overall survival assessment. Additionally, it was reported that galactose in the urine of patients with HCC was significantly up-regulated [23]. As a specific binding protein of beta-galactose, galectin-1 (Gal-1) was overexpressed in HCC, which has been confirmed to be associated with aggressive clinical pathological features and poorer survival [24]. It was speculated that the up-regulated galactose may have an adverse effect on the prognosis by affecting Gal-1 overexpression. Some studies revealed an increase of tyrosine in blood, in addition, tyrosine, phenylalanine and other 12 metabolites (glutamate, leucine, etc.) were further determined to be significantly associated with HCC risk [25–27].

In this study, RS_{OS} and RS_{DFS} were proved to be independent and significant predictors for overall survival and disease-free survival respectively. Higher AUCs meant that combining RS_{OS} or RS_{DFS} with clinical stages could improve the ability of distinguishing prognosis, thereby improving the individualized treatment of patients. Currently, clinical classification systems such as BCLC stage and TNM stage used to make clinical treatment decisions are mainly based on clinical data. However, these staging systems provide little information about

the small molecule metabolic characteristics of HCC, which is critical for future customized treatments. Importantly, the risk scores established in this study could offer clues for the characteristics of tumor metabolism and prognosis. Therefore, it could provide an opportunity for the development of clinical trials to supplement the current staging systems based on the metabolic characteristics of tumors.

The defined two panels of metabolites have promising potential in predicting overall survival and disease-free survival, respectively. Although currently there is no standard of care for the adjuvant treatment for surgically treated HCC patients, clinicians could provide potentially feasible clinical adjuvant trials as early as possible by appropriately monitoring high-risk patients based on metabolic characteristics. However, it is necessary to conduct a broader range of clinical follow-up study and select multi-center samples for further external verification and prospective research, to explore the applicability and reliability of the models.

Conclusions

In conclusion, metabolomics technique is a potential tool for screening prognostic risk factors of HCC. The risk scores (RS_{OS} and RS_{DFS}) developed in this study were independent prognostic factors for predicting overall survival and disease-free survival of HCC patients after liver resection. Our findings provide an opportunity of complementing the effective metabolic prognostic biomarkers for HCC, which could assist in the design of adjuvant therapy trials and future research on the molecular mechanism of prognosis.

Abbreviations

HCC: hepatocellular carcinoma

GC-MS: gas chromatography-mass spectrometry

N-CSI: network-based metabolic feature selection method based on composite significance index

OS: overall survival

DFS: disease-free survival

HR: Hazard Ratio

CI: confidence interval

TNM: tumor-node-metastasis

BCLC: Barcelona Clinic Liver Cancer

TACE: transcatheter arterial chemoembolization

S: survivors

NS: non-survivors

NR: no recurrence or metastasis

R: recurrence or metastasis

AJCC: American Joint Committee on Cancer

UICC: Union for International Cancer Control

QC: quality control

PCA: Principal components analysis

C-index: Harrell's concordance index

tdROC: time-dependent receiver operating characteristic

RSD: relative standard deviation

AUC: area under curve

RS: risk score

HBV: hepatitis B virus

AFP: alpha-fetoprotein

ALP: alkaline phosphatase

GGT: gamma-glutamyltransferase

Declarations

Ethics approval and consent to participate

The study was approved by the ethics committee of the Eastern Hepatobiliary Surgery Hospital of the Second Military Medical University and was conducted in accordance with the guidelines of the 1964 Helsinki declaration, as amended in 1975. The written informed consent from all subjects were obtained.

Consent for publication

All authors approved the submission.

Availability of data and materials

All data and materials in this study are available from the corresponding author upon reasonable request.

Competing interests

We certify this manuscript has not been published in whole or in part nor is it being considered for publication elsewhere. The authors declare that they have no competing interests.

Funding

This work was supported by the foundation (No. 21876169 and No. 21705147) from the National Natural Science Foundation of China, the National Key Research and Development Program of China (2017YFC0906900).

Authors' contributions

CNF, BZS, CL, QQW, XYL, XHL and GWX participated in study design, data analysis and interpretation. TYJ, YXT and LWD collected samples and clinical information. CNF performed acquisition of serum metabolomics. CNF and BZS drafted the manuscript. XYL, XHL and GWX revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank all participants in the study.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.** *CA Cancer J Clin* 2018, **68:**394-424.
2. Yang JD, Larson JJ, Watt KD, Allen AM, Wiesner RH, Gores GJ, Roberts LR, Heimbach JA, Leise MD: **Hepatocellular Carcinoma Is the Most Common Indication for Liver Transplantation and Placement on the Waitlist in the United States.** *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* 2017, **15**.
3. Rajyaguru DJ, Borgert AJ, Smith AL, Thomes RM, Conway PD, Halfdanarson TR, Truty MJ, Kurup AN, Go RS: **Radiofrequency Ablation Versus Stereotactic Body Radiotherapy for Localized Hepatocellular Carcinoma in Nonsurgically Managed Patients: Analysis of the National Cancer Database.** *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2018, **36:**600-608.
4. Tsurusaki M, Murakami T: **Surgical and Locoregional Therapy of HCC: TACE.** *Liver Cancer* 2015, **4:**165-175.
5. Chan AWH, Zhong J, Berhane S, Toyoda H, Cucchetti A, Shi K, Tada T, Chong CCN, Xiang BD, Li LQ, et al: **Development of pre and post-operative models to predict early recurrence of hepatocellular carcinoma after surgical resection.** *J Hepatol* 2018, **69:**1284-1293.
6. Zaimenko I, Jaeger C, Brenner H, Chang-Claude J, Hoffmeister M, Grötzinger C, Detjen K, Burock S, Schmitt CA, Stein U, Lisec J: **Non-invasive metastasis prognosis from plasma metabolites in stage II colorectal cancer patients: The DACHS study.** *International journal of cancer* 2019, **145:**221-231.
7. Johnson CH, Ivanisevic J, Siuzdak G: **Metabolomics: beyond biomarkers and towards mechanisms.** *Nature reviews Molecular cell biology* 2016, **17:**451-459.
8. Chen P-H, Cai L, Huffman K, Yang C, Kim J, Faubert B, Boroughs L, Ko B, Sudderth J, McMillan EA, et al: **Metabolic Diversity in Human Non-Small Cell Lung Cancer Cells.** *Molecular cell* 2019, **76**.

9. Satriano L, Lewinska M, Rodrigues PM, Banales JM, Andersen JB: **Metabolic rearrangements in primary liver cancers: cause and consequences.** *Nature reviews Gastroenterology & hepatology* 2019, **16:**748-766.
10. Luo P, Yin P, Hua R, Tan Y, Li Z, Qiu G, Yin Z, Xie X, Wang X, Chen W, et al: **A Large-scale, multicenter serum metabolite biomarker identification study for the early detection of hepatocellular carcinoma.** *Hepatology* 2018, **67:**662-675.
11. Wang X, Zhang A, Sun H: **Power of metabolomics in diagnosis and biomarker discovery of hepatocellular carcinoma.** *Hepatology (Baltimore, Md)* 2013, **57:**2072-2077.
12. Kowalik MA, Puliga E, Cabras L, Sulas P, Petrelli A, Perra A, Ledda-Columbano GM, Morandi A, Merlin S, Orrù C, et al: **Thyroid hormone inhibits hepatocellular carcinoma progression via induction of differentiation and metabolic reprogramming.** *Journal of hepatology* 2020.
13. Han J, Han Ml, Xing H, Li Zl, Yuan Dy, Wu H, Zhang H, Wang Md, Li C, Liang L, et al: **Tissue and serum metabolomic phenotyping for diagnosis and prognosis of hepatocellular carcinoma.** *International Journal of Cancer* 2019, **146:**1741-1753.
14. Ye G, Liu Y, Yin P, Zeng Z, Huang Q, Kong H, Lu X, Zhong L, Zhang Z, Xu G: **Study of induction chemotherapy efficacy in oral squamous cell carcinoma using pseudotargeted metabolomics.** *Journal of proteome research* 2014, **13:**1994-2004.
15. Wang Z-T, Tan C-C, Tan L, Yu J-T: **Systems biology and gene networks in Alzheimer's disease.** *Neuroscience and biobehavioral reviews* 2019, **96:**31-44.
16. Netzer M, Weinberger KM, Handler M, Seger M, Fang X, Kugler KG, Graber A, Baumgartner C: **Profiling the human response to physical exercise: a computational strategy for the identification and kinetic analysis of metabolic biomarkers.** *Journal of clinical bioinformatics* 2011, **1:**34.
17. Baumgartner C, Lewis GD, Netzer M, Pfeifer B, Gerszten RE: **A new data mining approach for profiling and categorizing kinetic patterns of metabolic biomarkers after myocardial injury.** *Bioinformatics (Oxford, England)* 2010, **26:**1745-1751.
18. Li Y, Ruan Q, Li Y, Ye G, Lu X, Lin X, Xu G: **A novel approach to transforming a non-targeted metabolic profiling method to a pseudo-targeted method using the retention time locking gas chromatography/mass spectrometry-selected ions monitoring.** *Journal of chromatography A* 2012, **1255:**228-236.
19. Zhou Y, Song R, Zhang Z, Lu X, Zeng Z, Hu C, Liu X, Li Y, Hou J, Sun Y, et al: **The development of plasma pseudotargeted GC-MS metabolic profiling and its application in bladder cancer.** *Analytical and bioanalytical chemistry* 2016, **408:**6741-6749.
20. Zheng F, Zhao X, Zeng Z, Wang L, Lv W, Wang Q, Xu G: **Development of a plasma pseudotargeted metabolomics method based on ultra-high-performance liquid chromatography-mass spectrometry.** *Nature protocols* 2020.
21. Ye G, Zhu B, Yao Z, Yin P, Lu X, Kong H, Fan F, Jiao B, Xu G: **Analysis of Urinary Metabolic Signatures of Early Hepatocellular Carcinoma Recurrence after Surgical Removal Using Gas Chromatography–Mass Spectrometry.** *Journal of Proteome Research* 2012, **11:**4361-4372.
22. Wang Q, Su B, Dong L, Jiang T, Tan Y, Lu X, Liu X, Lin X, Xu G: **Liquid Chromatography-Mass Spectrometry-Based Nontargeted Metabolomics Predicts Prognosis of Hepatocellular Carcinoma after Curative Resection.** *J Proteome Res* 2020.

23. Ladep NG, Dona AC, Lewis MR, Crossey MM, Lemoine M, Okeke E, Shimakawa Y, Duguru M, Njai HF, Fye HK, et al: **Discovery and validation of urinary metabotypes for the diagnosis of hepatocellular carcinoma in West Africans.** *Hepatology* 2014, **60:**1291-1301.
24. Leung Z, Ko FCF, Tey SK, Kwong EML, Mao X, Liu BHM, Ma APY, Fung YME, Che C-M, Wong DKH, et al: **Galectin-1 promotes hepatocellular carcinoma and the combined therapeutic effect of OTX008 galectin-1 inhibitor and sorafenib in tumor cells.** *Journal of experimental & clinical cancer research : CR* 2019, **38:**423.
25. Stepien M, Duarte-Salles T, Fedirko V, Floegel A, Barupal DK, Rinaldi S, Achaintre D, Assi N, Tjonneland A, Overvad K, et al: **Alteration of amino acid and biogenic amine metabolism in hepatobiliary cancers: Findings from a prospective cohort study.** *Int J Cancer* 2016, **138:**348-360.
26. Beyoglu D, Idle JR: **The metabolomic window into hepatobiliary disease.** *J Hepatol* 2013, **59:**842-858.
27. Yang F, Li J, Deng H, Wang Y, Lei C, Wang Q, Xiang J, Liang L, Xia J, Pan X, et al: **GSTZ1-1 Deficiency Activates NRF2/IGF1R Axis in HCC via Accumulation of Oncometabolite Succinylacetone.** *EMBO J* 2019, **38:**e101964.
28. Mostafa H, Amin AM, Teh CH, Murugaiyah V, Arif NH, Ibrahim B: **Metabolic phenotyping of urine for discriminating alcohol-dependent from social drinkers and alcohol-naive subjects.** *Drug Alcohol Depend* 2016, **169:**80-84.

Figures

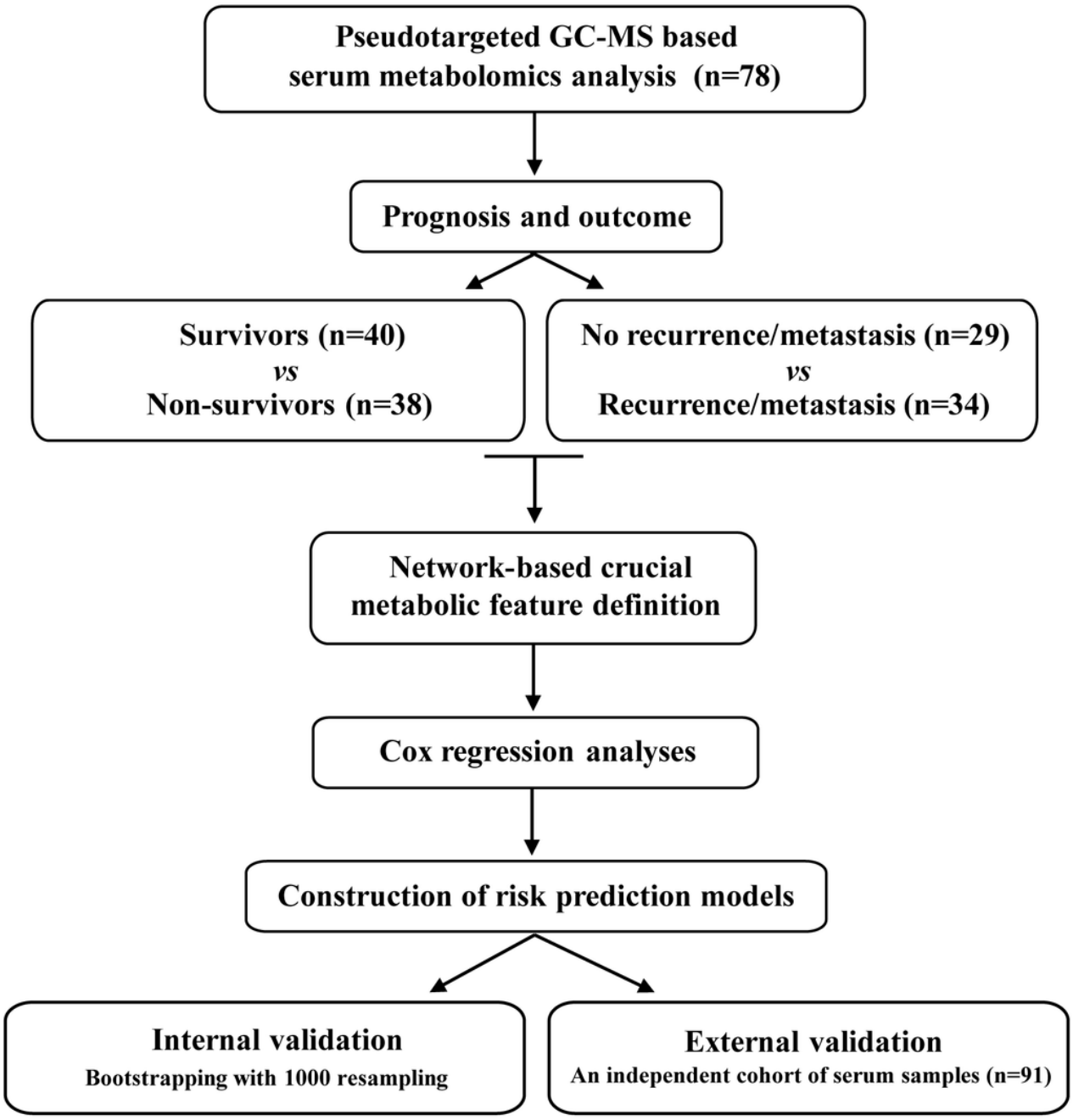

Figure 1

Flow chart of analysis process.

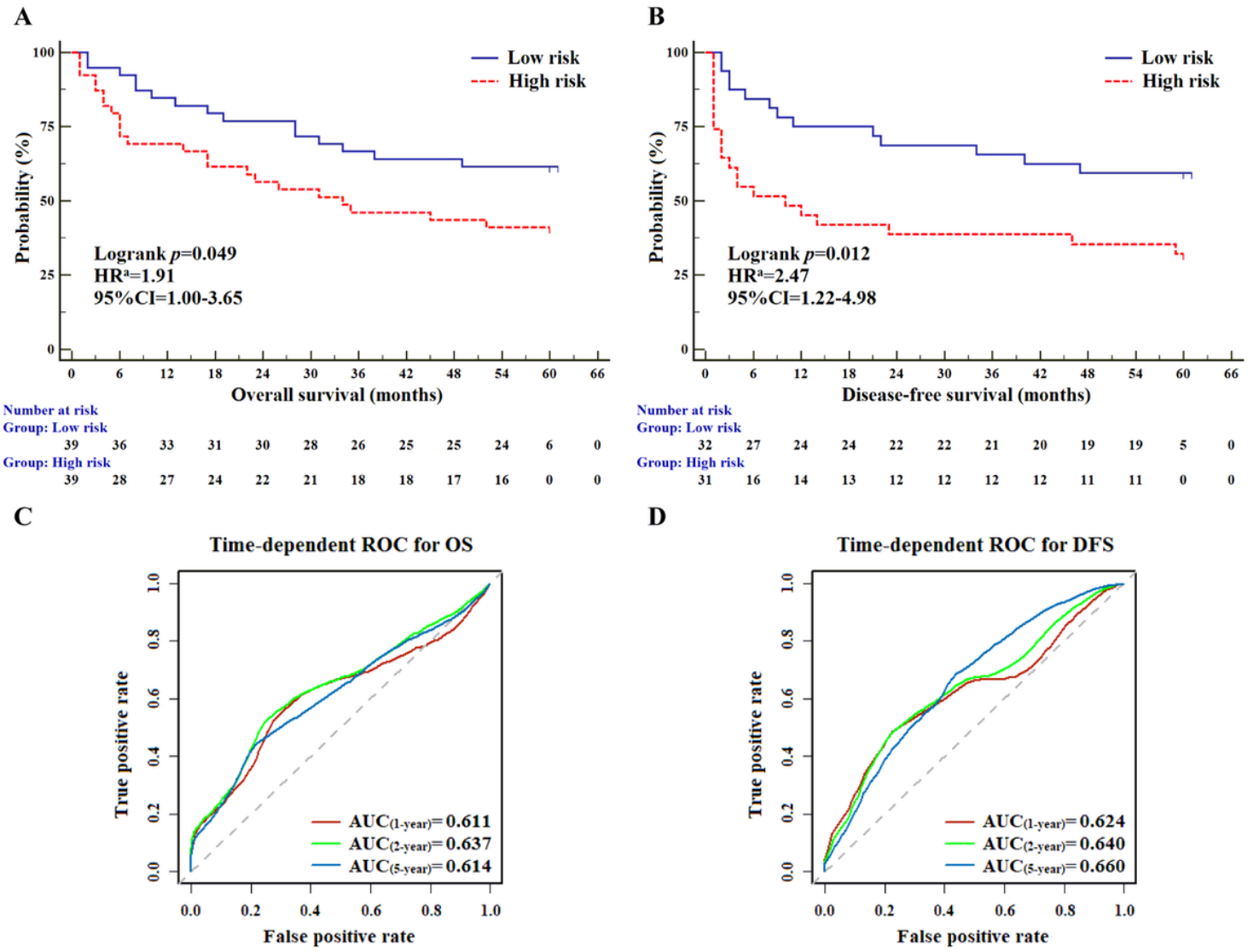

Figure 2

Prognostic performance of two risk score models. Kaplan-Meier plots for (A) overall survival and (B) disease-free survival in the low and high-risk groups; The time-independent ROC analysis of the risk scores to predict the (C) overall survival and (D) disease-free survival. a High risk/Low risk.

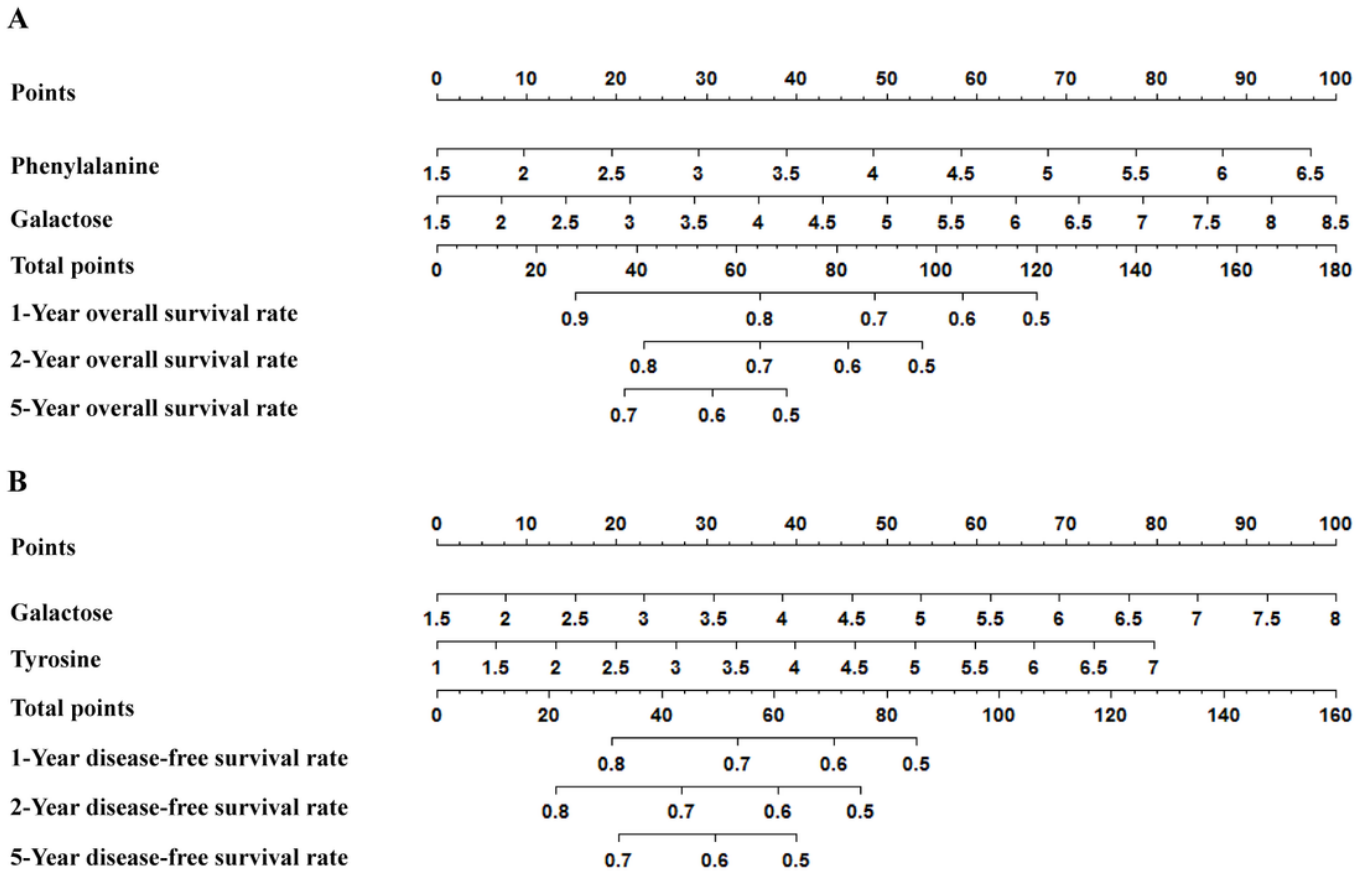

Figure 3

Nomograms for predicting (A) overall survival rate and (B) disease-free survival rate. The relative concentration of metabolites was obtained by multiplying the concentration of internal standards after internal standard correction.

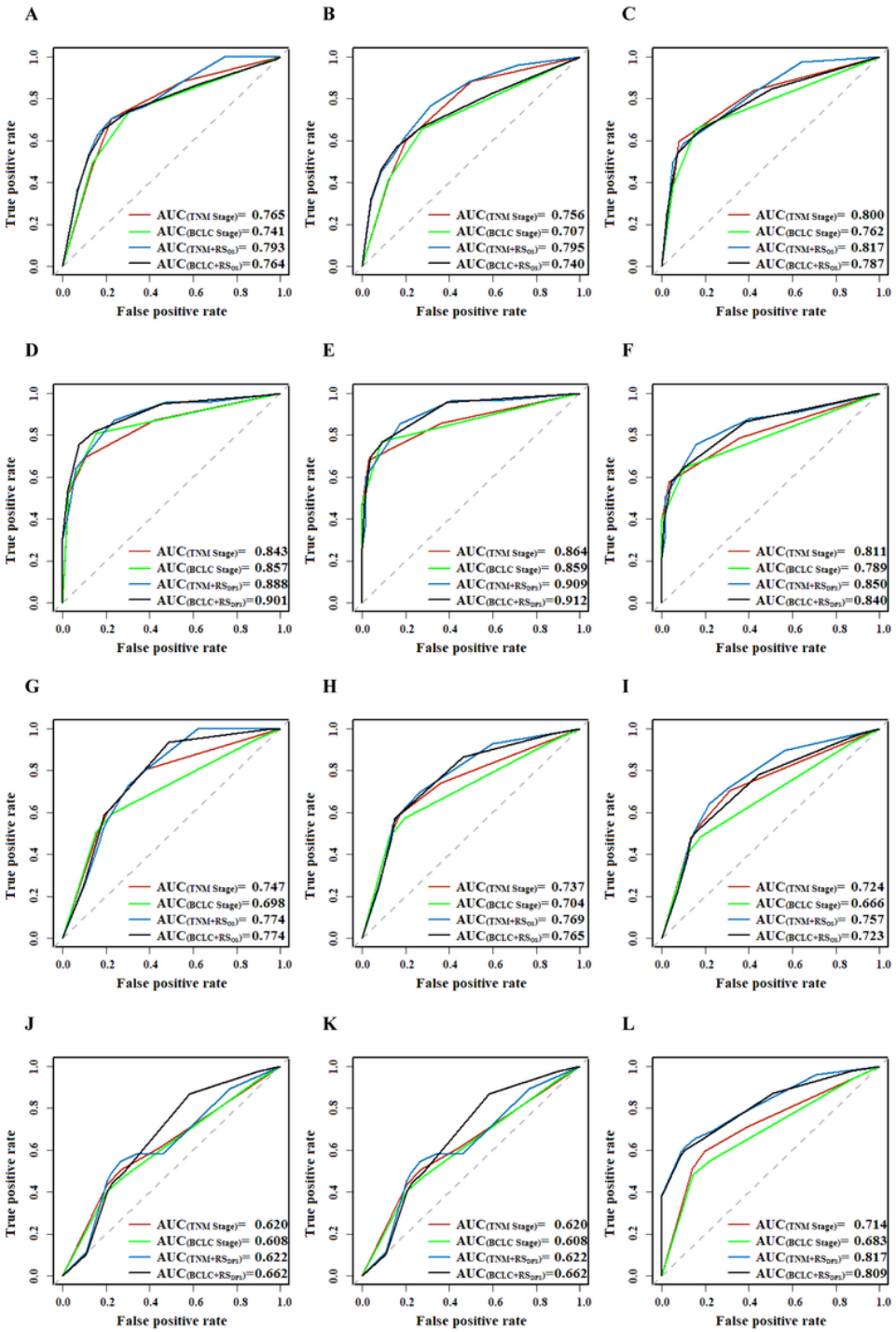

Figure 4

ROC curves of clinical staging systems and the combined models in discovery and validation cohorts. In discovery cohort: (A) 1-Year overall survival; (B) 2-Year overall survival; (C) 5-Year overall survival; (D) 1-Year disease-free survival; (E) 2-Year disease-free survival; and (F) 5-Year disease-free survival. In validation cohort: (G) 1-Year overall survival; (H) 2-Year overall survival; (I) 5-Year overall survival; (J) 1-Year disease-free survival; (K) 2-Year disease-free survival; and (L) 5-Year disease-free survival.

Supplementary Files

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

- AdditionalMaterial.pdf