Received: 25 August 2021 | Revised: 26 October 2021 | Accepted: 8 November 2021

DOI: 10.1111/cas.15217

Cancer Science WILEY

ORIGINAL ARTICLE

A novel immunodiagnosis panel for hepatocellular carcinoma based on bioinformatics and the autoantibody-antigen system

Jinyu Wu[1,2] | Peng Wang[1,2] | Zhuo Han[1,2] | Tiandong Li[1,2] | Chuncheng Yi[1,2] | Cuipeng Qiu[2,3] | Qian Yang[1,2] | Guiying Sun[1,2] | Liping Dai[2,3] | Jianxiang Shi[2,3] | Keyan Wang[2,3] | Hua Ye[1,2]

[1]College of Public Health, Zhengzhou University, Zhengzhou, China

[2]Henan Key Laboratory of Tumor Epidemiology and State Key Laboratory of Esophageal Cancer Prevention & Treatment, Zhengzhou University, Zhengzhou, China

[3]Henan Institute of Medical and Pharmaceutical Sciences, Zhengzhou University, Zhengzhou, China

Correspondence
Hua Ye, College of Public Health, Department of Epidemiology and Health Statistics, Zhengzhou University, Zhengzhou, Henan, China.
Email: yehua@zzu.edu.cn

Funding information
This research was supported by the National Science and Technology Major Project of China (No.2018ZX10302205), the Key Research Program of Higher Education in Henan Province (No. 22A330003), the Zhengzhou Major Project for Collaborative Innovation (No.18XTZX12007), the Major Project of Science and Technology in Henan Province (No.161100311400), and the Project for International Training of High-level Talents in Henan Province. The funder has no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript

Abstract

Hepatocellular carcinoma (HCC) is a malignancy with a dismal survival rate. The novel autoantibodies panel may provide new insights for the diagnosis of HCC. Biomarkers screened by two methods (bioinformatics and the antigen-antibody system) were taken as candidate tumor-associated antigens (TAAs). Enzyme-linked immunosorbent assay was used to detect the corresponding autoantibodies in 888 samples of verification and validation cohorts. The verification cohort was used to verify the autoantibodies. Samples in the validation cohort were randomly divided into a train set and a test set with the ratio of 6:4. A diagnostic model was established by support vector machines within the train set. The test set further verified the model. Eleven TAAs were selected (AAGAB, C17orf75, CDC37L1, DUSP6, EID3, PDIA2, RGS20, PCNA, TAF7L, TBC1D13, and ZIC2). The titer of six autoantibodies (PCNA, AAGAB, CDC37L1, TAF7L, DUSP6, and ZIC2) had a significant difference in any of the pairwise comparisons among the HCC, liver cirrhosis, and normal control groups. The titer of these autoantibodies had an increasing tendency. Finally, an optimum diagnostic model was constructed with the six autoantibodies. The AUCs were 0.826 in the train set and 0.773 in the test set. The area under the curve (AUC) of this panel for diagnosing early HCC was 0.889. The diagnostic ability of the panel reduced with the progress of HCC. The positive rate of the panel in diagnosing alpha-fetoprotein (AFP)-negative patients was 75.6%. For early HCC, the sensitivity of the combination of AFP with the panel was 90.9% and superior to 53.2% of AFP alone. The novel immunodiagnosis panel combining AFP may be a new approach for the diagnosis of HCC, especially for early-HCC cases.

KEYWORDS
autoantibodies, bioinformatics, hepatocellular carcinoma, immunodiagnosis, panel

Abbreviations: AFP, alpha-fetoprotein; AUC, area under the curve; ELISA, enzyme-linked immunosorbent assay; HCC, hepatocellular carcinoma; LC, liver cirrhosis; NCs, normal controls; SNR, signal-to-noise ratio; SVM, support vector machines; TAAbs, autoantibodies against tumor-associated antigen; TAAs, tumor-associated antigens; WGCNA, weighted gene coexpression network analysis.

Jinyu Wu and Peng Wang contributed equally to this work.

1 | INTRODUCTION

Liver cancer, especially primary liver cancer, is one of the leading causes of cancer-related death worldwide, of which hepatocellular carcinoma (HCC) accounts for 80%-90%.[1] The morbidity and mortality of liver cancer ranked seventh and second in all malignant tumors, respectively.[2] Clinically, only 10%-20% of patients with HCC could be treated by surgical operation due to the difficulty of early diagnosis of HCC.[3,4] Without a specific treatment plan, the median survival time of advanced liver cancer is only 1-2 months.[5,6] Alpha-fetoprotein (AFP), as the only serum biomarker for HCC diagnosis in clinical practice, has been widely used. However, its sensitivity is only about 60%.[7,8] Hence, it is an urgent need to identify effective and noninvasive biomarkers for the early diagnosis of HCC.

A great deal of evidence showed that autoantibodies against tumor-associated antigen (TAAbs) arose in the blood at an early stage of tumorigenesis.[9-11] Compared with other serological markers,[12-14] TAAbs persist in the blood for long periods.[15,16] TAAbs can serve as tumor-diagnostic biomarkers also owing to their easy measurement in serum and immune amplification effect.[17-20] Researchers also found that the diagnostic value of a single TAAb was not particularly ideal in regard to sensitivity and specificity.[16,21,22] Given the problem, scholars mainly focused on combining multiple TAAbs to diagnose tumors.[23,24] Different panels are needed for different types of cancers to enhance diagnostic power.[25]

In recent years, the rapid development of bioinformatics provided many research methods for the screening of tumor markers. Among them, weighted gene coexpression network analysis (WGCNA) is an excellent means of extracting gene modules and correlating them with clinical traits.[26,27] WGCNA has the unique advantage of converting gene expression data into coexpression modules and providing new insights into genes that may be responsible for phenotypic characteristics.[28,29] For these reasons, WGCNA was widely used in several cancers to identify pivot biomarkers for cancer diagnosis or prognosis.[30-33]

Autoantibody-antigen system is used as a common method for disease detection such as chronic hepatitis B[34] and plasmodium.[35] In the same way, the autoantibody-antigen system also can be used to screen disease markers. Among these, the utilization of a human proteomics chip to screen tumor-related markers is a significant method.[36,37] The chip used in this study contains all human recombinant proteins that can be purified at present to comprehensively screen tumor-associated antigens (TAAs) related to HCC.

In this study, bioinformatics and the autoantibody-antigen system were used to screen TAAs related to HCC. After verification and validation, 753 samples in the validation phase were divided into a train set and a test set to evaluate the ability of the anti-TAAs panel as noninvasive markers to diagnose HCC.

2 | MATERIALS AND METHODS

2.1 | Serum samples and research design

A total of 888 serum samples were taken into consideration, including 296 HCC cases, 296 liver cirrhosis (LC) cases, and 296 normal controls (NCs). The subjects among the three groups were age (±2 years) and sex matched according to 1:1 ratio. These samples were acquired from the Tumor Epidemiology Laboratory of Zhengzhou University. All subjects signed informed consent forms, and the study was approved by the Ethics Review Committee of Zhengzhou University (ZZURIB 2019-001). The verification phase including HCC (n = 45), LC (n = 45), and NC (n = 45) patients was used to validate the selected TAA. The remaining samples included in the evaluation phase were randomly divided into a train set and a test set according to the ratio of 6:4. The train set was used to build a diagnostic model, and the test set further verified the model (Figure 1).

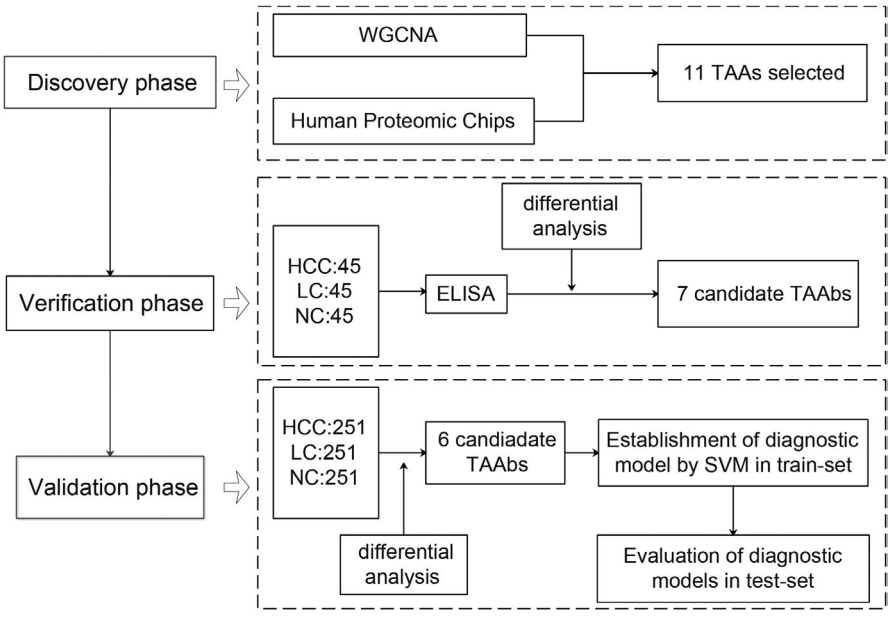


FIGURE 1 Work flow chart of the research. HCC, hepatocellular carcinoma; LC, liver cirrhosis; NC, normal control; TAAs, tumor-associated antigens; TAAbs, autoantibodies against tumor-associated antigens; SVM, support vector machines

2.2 | Data download and preprocessing

Seven microarray datasets, including GSE84005,[38] GSE76297,[39] GSE64041,[40] GSE54236,[39] GSE57957,[41] GSE39791,[42] GSE77314,[43] were downloaded from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) database (Table S1). The expressions of the probes in the chip were normalized by the "normalizeBetweenArrays" function in the "limma" package.[44] Each dataset was annotated with its corresponding annotation platform. When a gene had multiple probes, the maximum value was used as the expression value of the gene. GSM1979485, GSM1979420, and GSM1979407 were removed because they did not match the adjacent tissues in GSE76297. GSM1310594 in GSE54236 was deleted for the same reason. The "Combat" function in the "sva" package[45] was used to perform batch calibration among microarray datasets. The corrected dataset was analyzed.

2.3 | Screening of candidate TAAs

One of the methods used in this study to screen HCC-related markers was WGCNA. The basic theory of the WGCNA algorithm has been presented in a previous study.[46] The "WGCNA" package was used to calculate the gene coexpression network.[26] The "step-by-step" method was used to build the WGCNA network. Cluster analysis was used to remove outliers by the average linkage method. The scale-free topology fit index served as a function of the soft-thresholding power to choose an optimum threshold. The threshold was used to construct adjacency matrix and coexpression similarity. Based on dissimilarity topological overlap, genes were hierarchically clustered to construct geneTree. Dynamic hybrid branch cutting method was used to identified gene modules. The cutoff values of minimum size (minimum cluster size) and sensitivity to cluster splitting were 30 and 2, respectively. The threshold was set at 0.25 to merge modules with similarities. WGCNA has an ability to calculate the correlation between gene modules and clinical phenotypes.[27] The correlation was tested by the Pearson test. Finally, the most relevant module was selected.

Another method for screening candidate TAAs was the human proteomics chip. The human proteomics chip used in the study is currently the highest throughput protein chip. It contains more than 21 000 recombinant proteins (https://cdi.bio/huprot/). Hence, the human proteomics chip is an effective method for comprehensive screening of TAAs.[47] It was applied to detect the TAA concentration in the sera of 10 HCC cases and 10 NCs. The experiment was completed under the company standard procedure (BCBIO Biotechnology).[48] The detected data were normalized based on signal-to-noise ratio (SNR) value. When the SNR of a protein was greater than 4, it was defined as positive protein. On this basis, we set an appropriate cutoff threshold for each protein based on the SNR of the protein in the control group and calculated the positive rate of the protein in the cancer group and the healthy control group. For each protein, we assumed the two sets of samples were from two identical populations. Then, a one-tailed test was performed by Mann-Whitney's *U*, represented by *P*-value. The following criteria were used to screen meaningful TAAs in the human proteomics chip: fold change (FC) > 1 and $P < .05$; positive rate in HCC cases ≥ 50%; and negative rate in the NC group ≥ 90%.

The intersection of the biomarkers screened by WGCNA and the human proteomics chip was treated as candidate TAAs.

2.4 | Indirect enzyme-linked immunosorbent assay (ELISA)

Indirect ELISA was used to detect the concentration of TAAbs in serum. The specific experimental methods and procedures have been introduced in detail in previous studies.[49,50] The 11 recombinant proteins (AAGAB, C17orf75, CDC37L1, DUSP6, EID3, PDIA2, RGS20, PCNA, TAF7L, TBC1D13, and ZIC2) were all diluted in the coating buffer at a concentration of 0.125 μg/mL. HCC, LC, and NC samples were evenly distributed on each plate. Every plate had five replicate sera and three blanks. Different plates were normalized by repeated sera.

2.5 | Diagnostic model construction based on SVM

Support vector machines (SVM) are a set of machine learning methods. Compared with other machine learning methods, SVM is very powerful in discerning nuances in complex data sets,[51] which has been used as a classifier for cancer classification.[52] To build the most suitable SVM learning diagnosis model, the kernel function was selected and the "radial" and gamma was set to 0.1. The model was used to distinguish HCC from NC and applied to the diagnosis of HCC subgroups such as early stage (Barcelona Clinic Liver Cancer [BCLC], stage 0-B) and late stage (BCLC, stage C-D). Electrochemiluminescence immunoassay was used to detect the concentration of AFP. The positive rates of three groups (TAAb panel, AFP, and combination of TAAb panel and AFP) in recognizing HCC were compared. And the positive rates of these three groups were also compared in early and late HCC. According to clinical standards, the threshold of AFP was set to 7 ng/mL. HCC patients were divided into an AFP-positive group (AFP[+]) and an AFP-negative group (AFP[−]). The panel was used to diagnose AFP(+) patients and AFP(−) patients. The model's ability to recognize LC was also evaluated. The diagnostic ability of the models was assessed by receiver operating characteristic curves (ROC) analysis and the area under the ROC curve (AUC).

2.6 | Statistical analysis

The processing of microarray datasets and the realization of WGCNA were carried out in the R 4.0.4 language. Sample

matching and analysis of differences among the three groups were implemented by the IBM SPSS 26.0 software. Before analysis, data were assessed for normal distribution by the Kolmogrov-Simirnov normality test. If the data conformed to the normal distribution, they were analyzed by the ANOVA; otherwise, they were tested by the Kruskal-Wallis test. The statistical difference between the two AUCs were evaluated by the MedCalc 11.0 software. The *P*-values calculated were two tailed, and significant differences were set at 95% level ($P < .05$).

3 | RESULTS

3.1 | Demographic characteristics

This experiment was mainly divided into three phases (Figure 1). In the discovery phase, WGCNA and human proteomic chips were used to screen candidate TAAs. In the verification and validation phase, ELISA was used to detect the titers of 11 corresponding TAAbs in the serum samples, including 296 HCC cases, 296 LC cases, and 296 NCs. The basic information of the participants is shown in detail in Table 1. Statistical analysis showed that there was no difference in the distribution of age and gender among the three groups ($P > .05$).

3.2 | Weighted coexpression network construction and key module identification

A total of 794 samples were included to construct a gene coexpression network. Sample clustering demonstrated that three outliers (GSM979132, GSM979123, and GSM1398656) need to be unlocked (Figure 2A). Then, the remaining 791 samples were used for WGCNA. The thresholding of $\beta = 4$ was selected to determine the scale-free scale of the network (Figure 2B). Through the thresholding power, genes were divided into 17 modules in total (Figure 2C). The turquoise module was obtained because it had the highest correlation with the sample category ($r = .73$, $P < .0001$, Figure 2D). It also showed a high correlation with clinical information (cor = 0.9, $P < .001$ Figure 2E). Therefore, the genes in the turquoise module were chosen for subsequent analyses.

3.3 | Tumor-associated antigen selected in the discovery phase

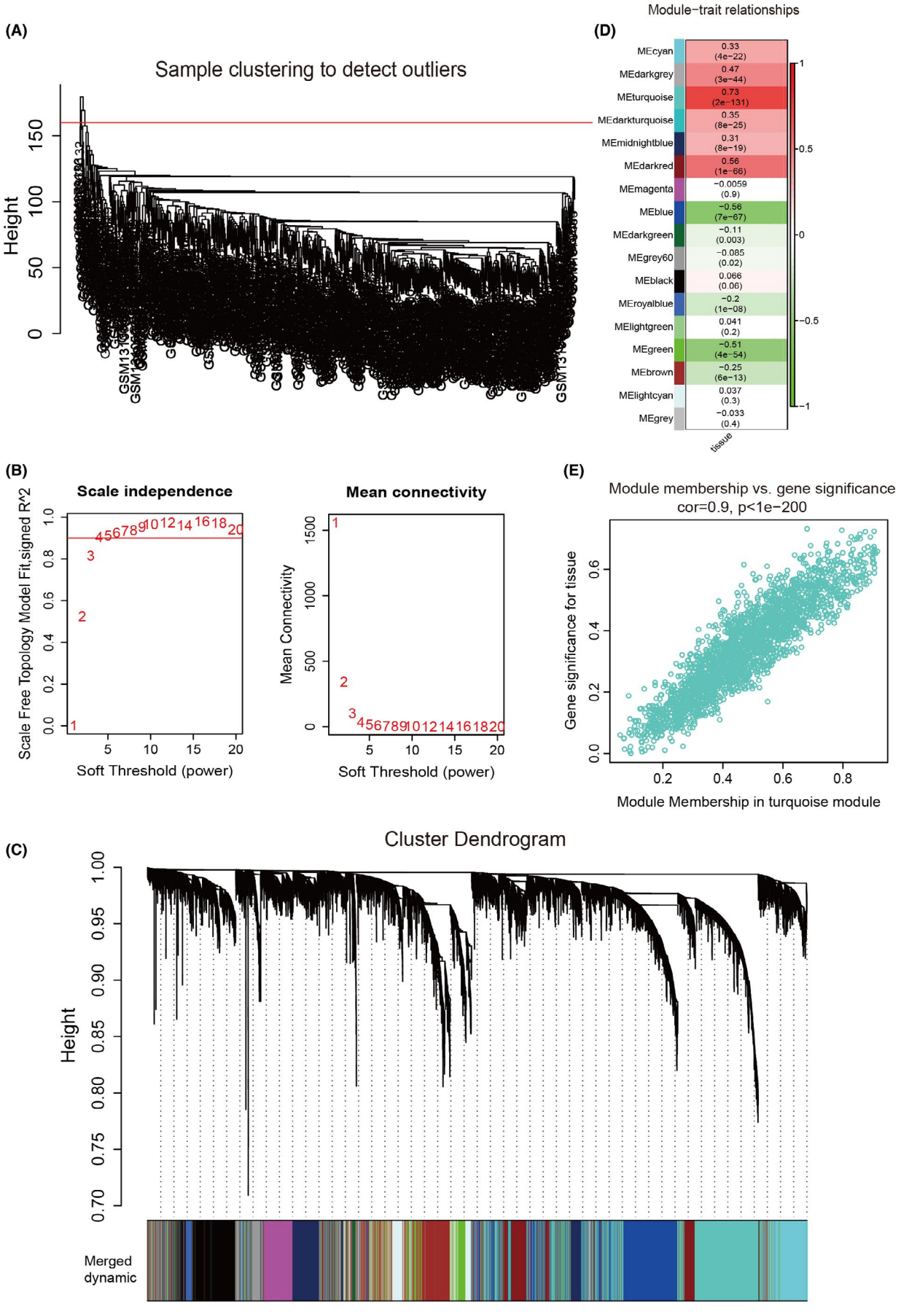
The human proteomics chip data were analyzed according to the standards established in the method. Finally, 11 TAAs were obtained by the two methods, including AAGAB, C17orf75, CDC37L1, DUSP6, EID3, PDIA2, RGS20, PCNA, TAF7L, TBC1D13, and ZIC2. The basic characteristics of the 11 TAAs are shown in Table S2.

TABLE 1 Characteristics of participants

	Verification dataset			Validation dataset		
Characteristics	**HCC (n = 45)**	**LC (n = 45)**	**NC (n = 45)**	**HCC (n = 251)**	**LC (n = 251)**	**NC (n = 251)**
Male n (%)	33 (73.3)	33 (73.3)	33 (73.3)	204 (80.3)	204 (80.3)	204 (80.3)
Age						
Mean ± SD	53.8 ± 9.9	53.5 ± 9.6	53.4 ± 9.8	52.6 ± 10.5	52.4 ± 1.05	52.1 ± 10.3
BCLC, n (%)						
0	0 (0)	IA	IA	4 (1.6)	IA	IA
A	4 (8.8)	IA	IA	40 (15.9)	IA	IA
B	8 (17.8)	IA	IA	35 (13.9)	IA	IA
C	9 (20.0)	IA	IA	55 (21.9)	IA	IA
D	0 (0)	IA	IA	8 (3.2)	IA	IA
NA	24 (53.4)	IA	IA	109 (43.7)	IA	IA
AFP, n (%)						
>7 ng/mL	22	NA	NA	131 (52.2)	NA	NA
<7 ng/mL	18	NA	NA	86 (34.3)	NA	NA
NA	5	NA	NA	34 (13.5)	NA	NA

Abbreviations: AFP, alpha-fetoprotein; BCLC, Barcelona Clinic Liver Cancer; HCC, hepatocellular carcinoma; IA, inapplicable; LC, liver cirrhosis; NA, not available; NC, normal control; SD, standard deviation.

FIGURE 2 Identification of modules related to the clinical features of hepatocellular carcinoma. A, Identification and deletion of abnormal samples. B, Recognition of soft threshold in weighted gene coexpression network analysis (WGCNA). C, Gene clustering dendrogram based on dissimilarity measure (1-TOM). D, Heat map of correlation between the gene clustering module and clinical characteristics of hepatocellular carcinoma. E, Module membership vs gene significance in the turquoise module


(A)
Sample clustering to detect outliers
Height
0
50
100
150
(D)
Module−trait relationships
MEcyan 0.33 (4e−22)
MEdarkgrey 0.47 (3e−44)
MEturquoise 0.73 (2e−131)
MEdarkturquoise 0.35 (8e−25)
MEmidnightblue 0.31 (8e−19)
MEdarkred 0.56 (1e−66)
MEmagenta −0.0059 (0.9)
MEblue −0.56 (7e−67)
MEdarkgreen −0.11 (0.003)
MEgrey60 −0.085 (0.02)
MEblack 0.066 (0.06)
MEroyalblue −0.2 (1e−08)
MElightgreen 0.041 (0.2)
MEgreen −0.51 (4e−54)
MEbrown −0.25 (6e−13)
MElightcyan 0.037 (0.3)
MEgrey −0.033 (0.4)
tissue
1
0.5
0
−0.5
−1
(B)
Scale independence
Scale Free Topology Model Fit,signed R^2
Soft Threshold (power)
Mean connectivity
Mean Connectivity
Soft Threshold (power)
(E)
Module membership vs. gene significance
cor=0.9, p<1e−200
Gene significance for tissue
Module Membership in turquoise module
(C)
Cluster Dendrogram
Height
Merged dynamic

3.4 | Detection of autoantibodies in the verification and validation phase

To determine whether the 11 TAAbs had a potential diagnostic value, their titers were measured in two phases. In the verification phase, ELISA was used to detect the titer of the 11 TAAbs in 135 samples. The scatter plot of the optical density (OD) values of the 11 TAAbs is shown in Figure 3A. The titer of autoantibodies against CDC37L1, ZIC2, DUSP6, PCNA, EID3, TAF7L, and AAGAB had differences among the three groups. Then, the seven TAAbs were examined in the validation cohort. Results showed that the average level of the OD values of six autoantibodies (PCNA, AAGAB, CDC37L1, TAF7L, DUSP6, and ZIC2) had a significant difference in any of the pairwise comparisons between the three groups (Figure 3B). Furthermore, the OD value of the six autoantibodies was the highest in the HCC group and the lowest in the NCs group. Finally, six autoantibodies (PCNA, AAGAB, CDC37L1, TAF7L, DUSP6, and ZIC2) were chosen for further analysis.

3.5 | Establishment of a diagnostic model for distinguishing HCC from NC

Receiver operating characteristic curves analysis was carried out to appraise the diagnostic value of each anti-TAA autoantibody in the validation cohort. The AUCs of the six TAAbs ranged from 0.671 to 0.741 (Figure 4A-F).

The six TAAbs were used as a panel to construct a diagnostic model by the train set. The AUC of the model was 0.826 (95% confidence interval [CI]: 0.779-0.873). Sensitivity, specificity, and accuracy rates were 77.5%, 76.2%, and 77.0%, respectively (Figure 5A, Table 2). The test set was used to verify the model, and a slightly lower AUC (Figure 5B, Table 2) was acquired. There was no statistical difference in AUC between the train set and the test set ($z = 1.299$, $P = .1940$).

The model produced by the train set was also used to diagnose the early stage (BCLC, stage 0-B) and late stage (BCLC, stage C-D) of HCC. The AUC of this panel to diagnose early HCC was 0.889 (Figure 5C and Table 2), and that to diagnose late stage was 0.811

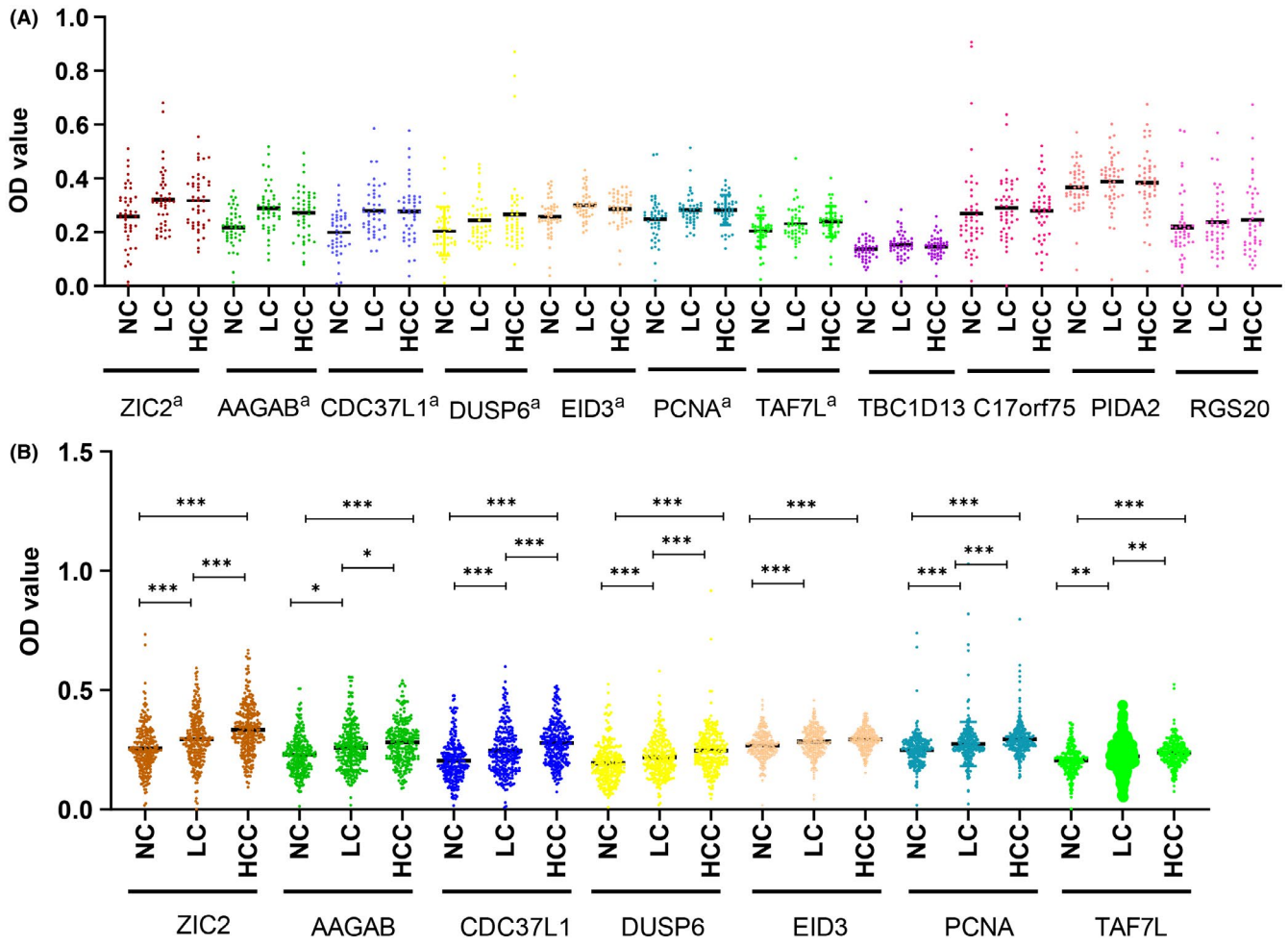


FIGURE 3 Scatter plot of the titers of 11 autoantibodies against tumor-associated antigen in verification phase (A) and validation phase (B). The superscript letter "a" indicates that the OD value of this autoantibody was different among the three groups. OD, optical density

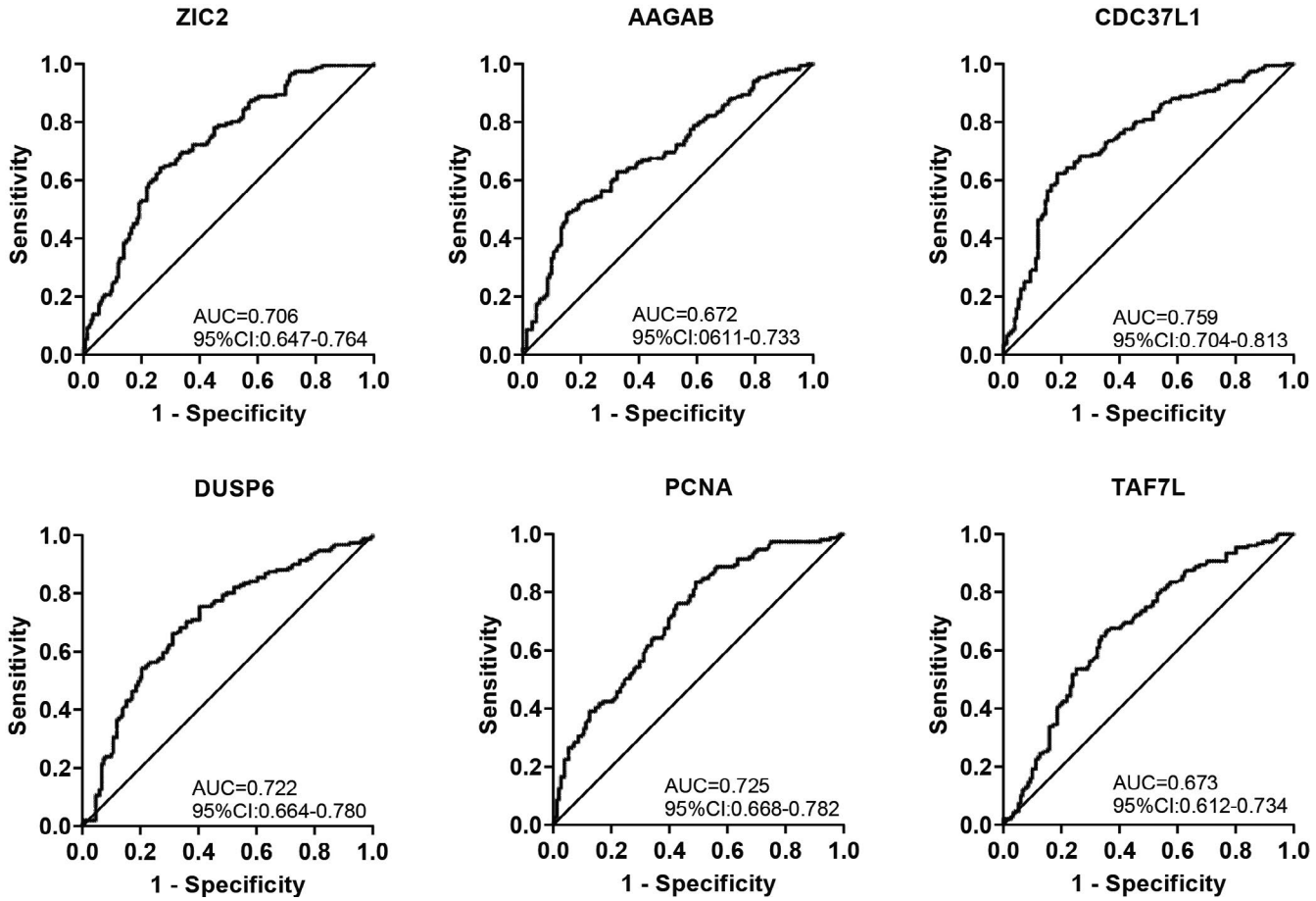


FIGURE 4 The diagnostic capabilities of six autoantibodies in the validation phase. CI, confidence interval; AUC, area under the curve

(Figure 5D and Table 2). The difference between the two AUCs was not statistically significant ($z = 1.673$, $P = .094$).

3.6 | Combination of AFP and autoantibodies

There were 217 HCC patients with AFP information. The positive rate of AFP was only 59.9% (130/217) (Table 3), while the positive rate of the panel was 77.9% (169/217). When the AFP and the panel were connected in parallel, the positive rate reached 90.3% (196/217). There were statistical differences in positive rates among them ($\chi^2 = 55.69$, $P < .001$, Table 3).

According to the clinical cutoff value (7 ng/mL), HCC patients were divided into an AFP(+) group and an AFP(−) group. The model was used to diagnose AFP(+) patients and AFP(−) patients (Figure 5E,F and Table 2). In the AFP(+) group, the AUC value was 0.822, which was slightly higher than that of the AFP(−) group. However, there was no statistical difference between the two groups ($z = 0.923$, $P = .356$).

In early-HCC patients, the sensitivity of AFP was only 53.2% (41/77). Interestingly, 80.5% (29/36) of AFP(−) HCC sera were anti-TAAs positive. If the panel and AFP were combined to diagnose HCC, 90.9% (70/77) of early-HCC patients were correctly diagnosed (Table 4). The change in sensitivity was statistically significant ($\chi^2 = 27.03$, $P < .001$).

In late-HCC patients, the AFP-positive rate was 87.7% (50/57). All AFP-negative patients were diagnosed as HCC cases by the panel. When AFP was used in combination with the panel, all late patients were correctly identified (Table 4). The change in positive rate was statistically significant ($\chi^2 = 5.14$, $P = .023$).

3.7 | Identification of LC using the optimal model

To explore the progression pattern of HCC, the diagnostic ability of the model in the course of the disease was analyzed. First, the model's ability to differentiate HCC from LC was examined. The results showed that the AUC was 0.652 (95% CI: 0.604-0.699). Sensitivity, specificity, and accuracy rates were 58.2%, 64.7%, and 61.4%, respectively (Figure 5G, Table 2). The model's ability to discriminate early HCC from LC was also examined, and a slightly higher AUC was acquired (Figure 5H, Table 3). The AUC of differentiating LC from NC was 0.656 (95% CI: 0.608-0.705). Sensitivity, specificity, and accuracy rates were 55.8%, 74.9%, and 65.4%, respectively (Figure 5G, Table 3).

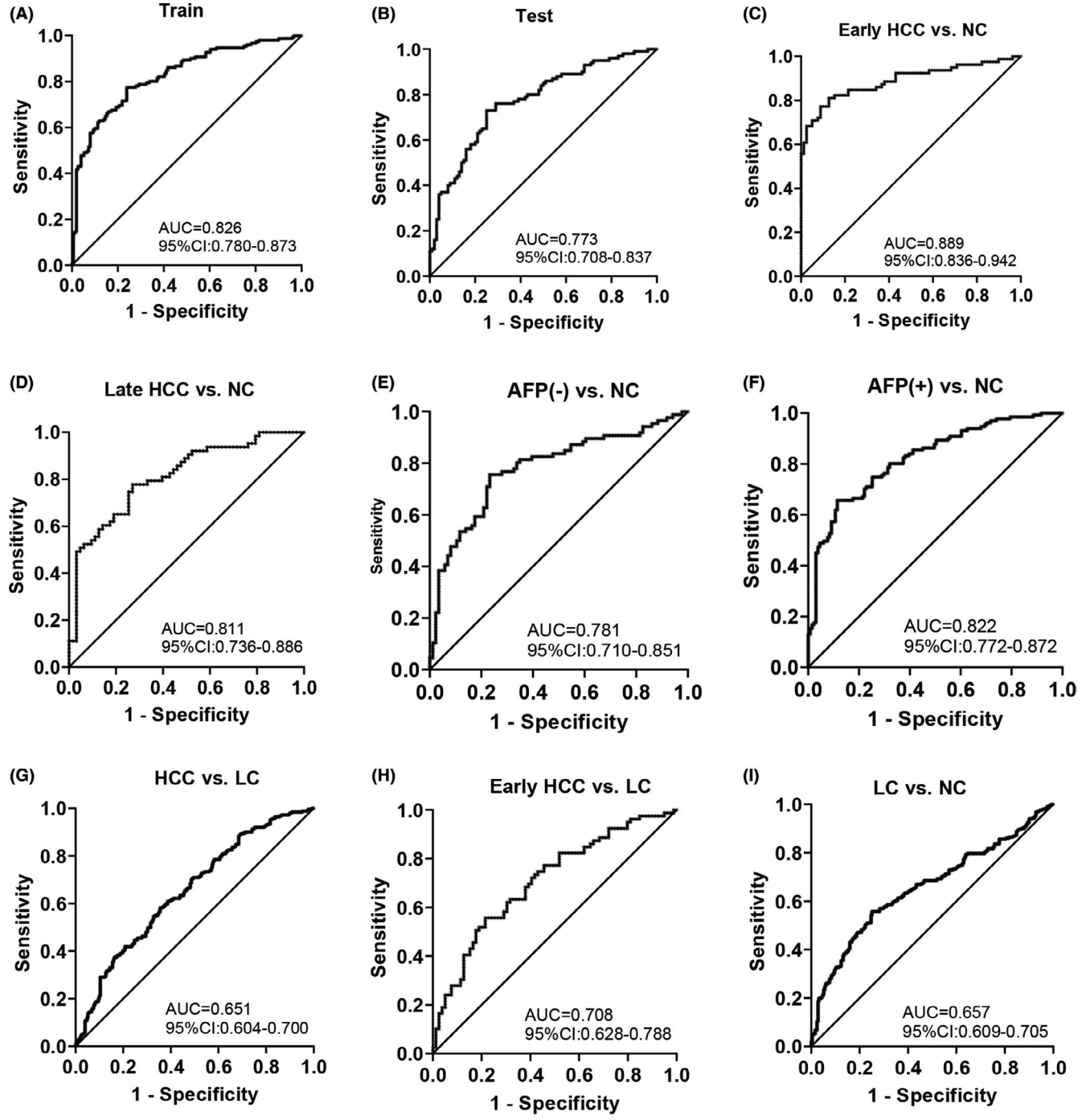


FIGURE 5 Receiver operating characteristic (ROC) curves for the panel by (A) tian, (B) test, (C) early HCC and normal control, (D) late HCC and normal control, (E) AFP(−) HCC and normal control, (F) AFP(+) HCC and normal control, (G) HCC and liver cirrhosis, (H) early HCC and liver cirrhosis, (I) liver cirrhosis and normal control. AFP, alpha-fetoprotein; AFP(−), AFP-negative group; AFP(+), AFP-positive group; AUC, area under the curve; BCLC, Barcelona Clinic Liver Cancer; CI, confidence interval; early HCC, BCLC, stage 0-B; HCC, hepatocellular carcinoma; late HCC, BCLC stage C-D; LC, liver cirrhosis; NC, normal control

4 | DISCUSSION

At present, HCC is still one of the lethal cancers threatening human health in the world.[2] Diagnosis of HCC has always been a research focus. This article mainly used bioinformatics and the autoantibody-antigen system to screen TAAs related to HCC. Then, ELISA was used to detect the titer of the corresponding autoantibodies in the sera, and SVM was used to build a diagnostic model. In the end, we successfully constructed a panel with a good diagnostic performance for detecting HCC.

In this study, bioinformatics and the autoantibody-antigen system were combined to screen the candidate TAAs. Among them,

TABLE 2 Evaluation of the diagnostic value of the SVM model in different datasets

	AUC (95% CI)	Se (%)	Sp (%)	Youden's index	PLR	NLR	PPV (%)	NPV (%)	Accuracy (%)
Train	0.826 (0.779-0.873)	78.0	76.0	0.536	3.250	0.289	76.5	77.6	77.0
Test	0.773 (0.708-0.837)	73.0	75.0	0.480	2.920	0.360	74.5	73.5	74.0
Early HCC vs NC	0.889 (0.836-0.943)	81.0	87.0	0.684	6.231	0.218	86.2	82.1	84.0
Later HCC vs NC	0.811 (0.736-0.886)	78.0	73.0	0.508	2.889	0.301	74.3	76.8	75.5
AFP(+) vs NC	0.822 (0.772-0.872)	65.6	88.5	0.542	5.739	0.384	85.2	72.2	77.3
AFP(−) vs NC	0.781 (0.710-0.851)	75.6	76.7	0.523	3.304	0.312	76.8	76.2	76.5
HCC vs LC	0.652 (0.604-0.699)	58.2	64.5	0.227	1.639	0.648	62.1	60.7	61.4
early HCC vs LC	0.710 (0.628-0.788)	56.0	79.0	0.342	2.667	0.557	72.7	64.2	67.5
LC vs NC	0.656 (0.608-0.705)	55.8	74.9	0.307	2.223	0.590	69.0	62.9	65.4

Abbreviations: AFP(−), AFP negative group; AFP(+), AFP positive group; AUC, area under the receiver operating characteristic curve; CI, confidence interval; early HCC, Barcelona Clinic Liver Cancer 0-B; HCC, hepatocellular carcinoma; later HCC: Barcelona Clinic Liver Cancer C-D; NC, normal control; NLR, negative likelihood ratio; NPV, negative predictive value; PLR, positive likelihood ratio; PPV, positive predictive value; Se, sensitivity; Sp, specificity; SVM, support vector machines.

TABLE 3 Comparison of single and parallel detection of AFP and model

	Panel	AFP	Panel + AFP	χ^2/P
Number of positive	169_a	130_b	196_c	55.69/<.001
Number of negative	48	87	21	

Note: Each subscript letter denotes a subset of work categories whose column proportions do not differ significantly from each other at the 0.05 level.

Abbreviation: AFP, alpha-fetoprotein.

WGCNA is an excellent algorithm in screening the biological markers, and it has been used in many studies.[26] The human proteomics chip contains 21 000 human recombinant proteins in this study, which can comprehensively screen TAAs.[53-56] In summary, bioinformatics can screen for genes that play important roles in the occurrence and development of HCC. Hence, those two methods were used together to screen valuable TAAs as much as possible.

Next, the concentration of autoantibodies against selected TAAs was validated using ELISA. Through the verification and the validation phase, six TAAbs (PCNA, AAGAB, CDC37L1, TAF7L, DUSP6, and ZIC2) had a significant difference in any of the pairwise comparisons between the three groups. Interestingly, we also found the titer of the six autoantibodies had an increasing tendency with the disease stage, which may provide early warning of the onset of HCC.

In recent years, many studies have shown that exploration of a panel is helpful for the diagnosis of cancer. Furthermore, machine learning can provide promising tools; it has been applied in pancreatic cancer[57] and prostate cancer.[58] According to the research of Zhang et al, the diagnostic ability of a single antigen-antibody system was limited, but combining multiple TAAbs can improve the diagnostic performance for cancer detection.[24] Hence, the six TAAbs were used to construct a diagnostic model using SVM. The panel showed good diagnostic performance for detecting HCC. In addition, an interesting phenomenon that caught our attention was that the panel was more valuable in diagnosing early HCC than late HCC. There are similar results in other cancers.[10,59] This decrease may be the result of the loss of antigens to help the tumor escape immune surveillance.[60] In other words, the panel constructed by six autoantibodies may be used for the diagnosis of early HCC. Compared with previous studies,[61-63] our research focuses on systematically and comprehensively screening HCC-related autoantibodies and verifying them in phases. Moreover, we constructed a robust HCC diagnostic panel with large samples, which had not been achieved before either. Review of previous studies showed that our research was consistent with the findings of related studies. For example, Wang et al developed a 22-autoantibody detector to diagnose prostate cancer, with 81.6% sensitivity and 88.2% specificity.[64] These strongly demonstrated that a customized autoantibodies panel is a promising and powerful immunodiagnostic tool for certain types of cancer, such as HCC and prostate cancer.

This study also analyzed the association between AFP and TAAbs in HCC diagnosis. The result showed that there was no statistical difference between the AUCs of the two groups (AFP[+] group and AFP[−] group). It can be concluded that TAAbs had no relationship with AFP. Namely, TAAbs are independent diagnostic markers that can be used to diagnose AFP(−) HCC patients. Some research studies revealed the AFP value was in the normal range in approximately 40% of diagnosed HCC cases.[65,66] The positive rate of AFP in HCC patients is consistent with several studies (Table 3). The positive rate of our model was 78%, which is significantly higher than that of AFP (Table 3). It is gratifying that the positive rate reached 90.3% when the two methods were connected in parallel. In order to further explore the supplementary ability of this panel for AFP to diagnose HCC, we divided patients with AFP information into an early-HCC and a late-HCC group. To our surprise, the positive rate of "panel+AFP" for diagnosis of early HCC was significantly improved, with a sensitivity of 90.9%, while

TABLE 4 Sensitivity of the combination of AFP and the panel to detect early and late HCC

	Serum AFP levels				
	Early HCC		Late HCC		
Anti-TAAs panel	>7 ng/mL	<7 ng/mL	>7 ng/mL	<7 ng/mL	Total
Positive	35	29	40	7	111
Negative	6	7	10	0	23
Total	41	36	50	7	134

Note: Sensitivity (%) of early HCC: (35 + 29 + 6)/(35 + 29 + 6 + 7) = 90.9%; Sensitivity (%) of late HCC: (40 + 7 + 10)/(40 + 7 + 10 + 0) = 100%.

Abbreviations: AFP, alpha-fetoprotein; HCC, hepatocellular carcinoma; TAAs, tumor-associated antigens.

AFP was only 53%. Those results were consistent with the conclusions of previous researches.[7,67] The above research results indicated the immunodiagnostic panel can be used as a marker for the diagnosis of AFP(−) HCC patients, and the panel combined with AFP may improve the sensitivity of diagnosing HCC, especially for early-HCC cases.

The current research has several advantages. First, the method was used to screen HCC-related TAAs should be mentioned. The method of combining bioinformatics and autoantibody-antigen system could be more accurate for screening out TAAs that may play an important role in the development of HCC. Second, multistage verification in large samples was applied, which makes the results more credible. Third, compared with AFP in parallel, the sensitivity of the panel in diagnosing HCC was higher, and the results indicated that the novel immunodiagnosis panel combined with AFP may be a new approach for the diagnosis of HCC, especially for early-HCC cases. However, there are some limitations to our research. First, the human proteomics chip does not include proteins with post-translational modifications and structural changes. Therefore, those related TAAbs cannot be detected. However, the purpose of combining with bioinformatics was to screen out important indicators as accurately as possible. Second, the function of the screened indicators in the occurrence and development of HCC needs further verification.

In summary, our research data showed that the customized autoantibodies panel provides a supplementary method for the diagnosis of HCC, especially for early-HCC and AFP(−) patients. The study also proves that autoantibodies are a promising direction for cancer biomarker research. The ability of this method to diagnose cancer depends on the careful choice of the autoantibodies panel. Our findings may encourage future clinical studies to explore the function of related proteins in the occurrence and progression of HCC.

DISCLOSURE

The authors declare no conflict of interest.

ORCID

Peng Wang https://orcid.org/0000-0003-4666-9706
Hua Ye https://orcid.org/0000-0003-3657-2417

REFERENCES

1. Rimassa L, Pressiani T, Merle P. Systemic treatment options in hepatocellular carcinoma. *Liver Cancer*. 2019;8:427-446.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70:7-30.
3. Marrero JA. Hepatocellular carcinoma. *Curr Opin Gastroenterol*. 2006;22:248-253.
4. Xiao J, Li G, Lin S, et al. Prognostic factors of hepatocellular carcinoma patients treated by transarterial chemoembolization. *Int J Clin Exp Pathol*. 2014;7:1114-1123.
5. Allaire M, Nault JC. Advances in management of hepatocellular carcinoma. *Curr Opin Oncol*. 2017;29:288-295.
6. Dutkowski P, Linecker M, DeOliveira ML, Müllhaupt B, Clavien PA. Challenges to liver transplantation and strategies to improve outcomes. *Gastroenterology*. 2015;148:307-323.
7. Wang K, Li M, Qin J, et al. Serological biomarkers for early detection of hepatocellular carcinoma: a focus on autoantibodies against Tumor-associated antigens encoded by cancer driver genes. *Cancers*. 2020;12:1271.
8. Xu RH, Wei W, Krawczyk M, et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat Mater*. 2017;16:1155-1161.
9. Zhu Q, Liu M, Dai L, et al. Using immunoproteomics to identify tumor-associated antigens (TAAs) as biomarkers in cancer immunodiagnosis. *Autoimmun Rev*. 2013;12:1123-1128.
10. Tan EM, Zhang J. Autoantibodies to tumor-associated antigens: reporters from the immune system. *Immunol Rev*. 2008;222:328-340.
11. Paradis V, Bedossa P. In the new area of noninvasive markers of hepatocellular carcinoma. *J Hepatol*. 2007;46:9-11.
12. Yan L, Chen Y, Zhou J, Zhao H, Zhang H, Wang G. Diagnostic value of circulating cell-free DNA levels for hepatocellular carcinoma. *Int J Infect Dis*. 2018;67:92-97.
13. Fu Y, Xu X, Huang D, et al. Plasma heat shock protein 90alpha as a biomarker for the diagnosis of liver cancer: an official, large-scale, and multicenter clinical trial. *EBioMedicine*. 2017;24:56-63.
14. Sartorius K, Sartorius B, Kramvis A, et al. Circulating microRNA's as a diagnostic tool for hepatocellular carcinoma in a hyper endemic HIV setting, KwaZulu-Natal, South Africa: a case control study protocol focusing on viral etiology. *BMC Cancer*. 2017;17:894.
15. Caron M, Choquet-Kastylevsky G, Joubert-Caron R. Cancer immunomics using autoantibody signatures for biomarker discovery. *Mol Cell Proteomics*. 2007;6:1115-1122.
16. Wu J, Li X, Song W, et al. The roles and applications of autoantibodies in progression, diagnosis, treatment and prognosis of human malignant tumours. *Autoimmun Rev*. 2017;16:1270-1281.
17. Katchman BA, Chowell D, Wallstrom G, et al. Autoantibody biomarkers for the detection of serous ovarian cancer. *Gynecol Oncol*. 2017;146:129-136.

18. Pedersen JW, Gentry-Maharaj A, Nøstdal A, et al. Cancer-associated autoantibodies to MUC1 and MUC4–a blinded case–control study of colorectal cancer in UK collaborative trial of ovarian cancer screening. *Int J Cancer*. 2014;134:2180-2188.
19. Sun G, Ye H, Wang X, et al. Autoantibodies against tumor-associated antigens combined with microRNAs in detecting esophageal squamous cell carcinoma. *Cancer Med*. 2020;9:1173-1182.
20. Chapman CJ, Thorpe AJ, Murray A, et al. Immunobiomarkers in small cell lung cancer: potential early cancer signals. *Clin Cancer Res*. 2011;17:1474-1480.
21. Zaenker P, Ziman MR. Serologic autoantibodies as diagnostic cancer biomarkers–a review. *Cancer Epidemiol Biomarkers Prev*. 2013;22:2161-2181.
22. Dai L, Tsay JC, Li J, et al. Autoantibodies against tumor-associated antigens in the early detection of lung cancer. *Lung Cancer*. 2016;99:172-179.
23. Zayakin P, Ancāns G, Siliņa K, et al. Tumor-associated autoantibody signature for the early detection of gastric cancer. *Int J Cancer*. 2013;132:137-147.
24. Zhang JY, Casiano CA, Peng XX, Koziol JA, Chan EK, Tan EM. Enhancement of antibody detection in cancer using panel of recombinant tumor-associated antigens. *Cancer Epidemiol Biomarkers Prev*. 2003;12:136-143.
25. Zhou JH, Zhang B, Kernstine KH, Zhong L. Autoantibodies against MMP-7 as a novel diagnostic biomarker in esophageal squamous cell carcinoma. *World J Gastroenterol*. 2011;17:1373-1378.
26. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
27. Farber CR. Identification of a gene module associated with BMD through the integration of network analysis and genome-wide association data. *J Bone Miner Res*. 2010;25:2359-2367.
28. Udyavar AR, Hoeksema MD, Clark JE, et al. Co-expression network analysis identifies Spleen Tyrosine Kinase (SYK) as a candidate oncogenic driver in a subset of small-cell lung cancer. *BMC Syst Biol*. 2013;7(Suppl 5):S1.
29. Shi Z, Derow CK, Zhang B. Co-expression module analysis reveals biological processes, genomic gain, and regulatory mechanisms associated with breast cancer progression. *BMC Syst Biol*. 2010;4:74.
30. Tian Z, He W, Tang J, et al. Identification of important modules and biomarkers in breast cancer based on WGCNA. *Onco Targets Ther*. 2020;13:6805-6817.
31. Wan Q, Tang J, Han Y, Wang D. Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma. *Exp Eye Res*. 2018;166:13-20.
32. Nangraj AS, Selvaraj G, Kaliamurthi S, Kaushik AC, Cho WC, Wei DQ. Integrated PPI- and WGCNA-retrieval of Hub gene signatures shared between Barrett's esophagus and esophageal adenocarcinoma. *Front Pharmacol*. 2020;11:881.
33. Ma Z, Wang J, Ding L, Chen Y. Identification of novel biomarkers correlated with prostate cancer progression by an integrated bioinformatic analysis. *Medicine (Baltimore)*. 2020;99:e21158.
34. Tiwari AK, Pabbi S, Aggarwal G, et al. Application of sequential serological testing strategy for detection of Hepatitis B surface antigen (HBsAg) for diagnosing HBV infection. *J Virol Methods*. 2019;274:113726.
35. Wanja EW, Kuya N, Moranga C, et al. Field evaluation of diagnostic performance of malaria rapid diagnostic tests in western Kenya. *Malar J*. 2016;15:456.
36. Huang Y, Zhu H. Protein array-based approaches for biomarker discovery in cancer. *Genomics Proteomics Bioinformatics*. 2017;15:73-81.
37. Wang X, Zhang Y, Sun L, et al. Evaluation of the clinical application of multiple tumor marker protein chip in the diagnostic of lung cancer. *J Clin Lab Anal*. 2018;32:e22565.
38. Zhang Q, Sun S, Zhu C, et al. Prediction and analysis of weighted genes in hepatocellular carcinoma using bioinformatics analysis. *Mol Med Rep*. 2019;19:2479-2488.
39. Chaisaingmongkol J, Budhu A, Dang H, et al. Common molecular subtypes among Asian hepatocellular carcinoma and cholangiocarcinoma. *Cancer Cell*. 2017;32:57-70.e53.
40. Makowska Z, Boldanova T, Adametz D, et al. Gene expression analysis of biopsy samples reveals critical limitations of transcriptome-based molecular classifications of hepatocellular carcinoma. *J Pathol Clin Res*. 2016;2:80-92.
41. Mah WC, Thurnherr T, Chow PK, et al. Methylation profiles reveal distinct subgroup of hepatocellular carcinoma patients with poor prognosis. *PLoS One*. 2014;9:e104158.
42. Kim JH, Sohn BH, Lee HS, et al. Genomic predictors for recurrence patterns of hepatocellular carcinoma: model derivation and validation. *PLoS Med*. 2014;11:e1001770.
43. Liu G, Hou G, Li L, Li Y, Zhou W, Liu L. Potential diagnostic and prognostic marker dimethylglycine dehydrogenase (DMGDH) suppresses hepatocellular carcinoma metastasis in vitro and in vivo. *Oncotarget*. 2016;7:32607-32616.
44. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:e47.
45. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28:882-883.
46. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4:45.
47. Figeys D, Pinto D. Proteomics on a chip: promising developments. *Electrophoresis*. 2001;22:208-216.
48. Tao SC, Zhu H. Protein chip fabrication by capture of nascent polypeptides. *Nat Biotechnol*. 2006;24:1253-1254.
49. Wang S, Qin J, Ye H, et al. Using a panel of multiple tumor-associated antigens to enhance autoantibody detection for immunodiagnosis of gastric cancer. *Oncoimmunology*. 2018;7:e1452582.
50. Qin J, Wang S, Wang P, et al. Autoantibody against 14-3-3 zeta: a serological marker in detection of gastric cancer. *J Cancer Res Clin Oncol*. 2019;145:1253-1262.
51. Aruna S, Rajagopalan SJIjoca. A novel SVM based CSSFFS feature selection algorithm for detecting breast cancer. 2011;31.
52. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics*. 2018;15:41-51.
53. Cui C, Duan Y, Qiu C, et al. Identification of novel autoantibodies based on the human proteomic chips and evaluation of their performance in the detection of gastric cancer. *Front Oncol*. 2021;11:637871.
54. Yang Q, Qin J, Sun G, et al. Discovery and validation of serum autoantibodies against tumor-associated antigens as biomarkers in gastric adenocarcinoma based on the focused protein arrays. *Clin Transl Gastroenterol*. 2020;12:e00284.
55. Ma Y, Wang X, Qiu C, et al. Using protein microarray to identify and evaluate autoantibodies to tumor-associated antigens in ovarian cancer. *Cancer Sci*. 2021;112:537-549.
56. Qiu C, Wang P, Wang B, et al. Establishment and validation of an immunodiagnostic model for prediction of breast cancer. *Oncoimmunology*. 2020;9:1682382.
57. Ye H, Li T, Wang H, et al. TSPAN1, TMPRSS4, SDR16C5, and CTSE as novel panel for pancreatic cancer: a bioinformatics analysis and experiments validation. *Front Immunol*. 2021;12:649551.
58. Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol*. 2019;16:391-403.
59. Comtesse N, Zippel A, Walle S, et al. Complex humoral immune response against a benign tumor: frequent antibody response against

specific antigens as diagnostic targets. *Proc Natl Acad Sci USA*. 2005;102:9601-9606.
60. Khong HT, Restifo NP. Natural selection of tumor variants in the generation of "tumor escape" phenotypes. *Nat Immunol*. 2002;3:999-1005.
61. Dai L, Ren P, Liu M, Imai H, Tan EM, Zhang JY. Using immunomic approach to enhance tumor-associated autoantibody detection in diagnosis of hepatocellular carcinoma. *Clin Immunol*. 2014;152:127-139.
62. Chen Y, Zhou Y, Qiu S, et al. Autoantibodies to tumor-associated antigens combined with abnormal alpha-fetoprotein enhance immunodiagnosis of hepatocellular carcinoma. *Cancer Lett*. 2010;289:32-39.
63. Okada R, Otsuka Y, Wakabayashi T, et al. Six autoantibodies as potential serum biomarkers of hepatocellular carcinoma: a prospective multicenter study. *Int J Cancer*. 2020;147:2578-2586.
64. Wang X, Yu J, Sreekumar A, et al. Autoantibody signatures in prostate cancer. *N Engl J Med*. 2005;353:1224-1235.
65. Luo P, Yin P, Hua R, et al. A Large-scale, multicenter serum metabolite biomarker identification study for the early detection of hepatocellular carcinoma. *Hepatology*. 2018;67:662-675.
66. Forner A, Reig M, Bruix J. Alpha-fetoprotein for hepatocellular carcinoma diagnosis: the demise of a brilliant star. *Gastroenterology*. 2009;137:26-29.
67. Wang T, Liu M, Zheng SJ, et al. Tumor-associated autoantibodies are useful biomarkers in immunodiagnosis of α-fetoprotein-negative hepatocellular carcinoma. *World J Gastroenterol*. 2017;23:3496-3504.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Wu J, Wang P, Han Z, et al. A novel immunodiagnosis panel for hepatocellular carcinoma based on bioinformatics and the autoantibody-antigen system. *Cancer Sci*. 2022;113:411–422. doi:10.1111/cas.15217