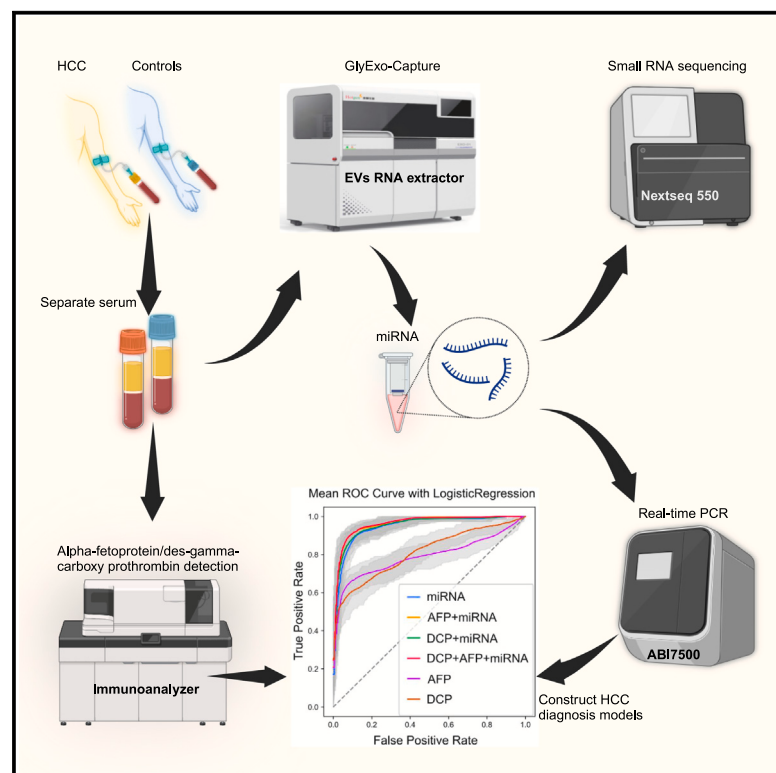


# Five miRNAs identified in fucosylated extracellular vesicles as non-invasive diagnostic signatures for hepatocellular carcinoma

## Graphical abstract



## Authors

Boan Li, Kun Hao, Mengyang Li, ..., Zhiming Zhao, Ronghua Jin, Ruifu Yang

## Correspondence

gaoqi@youngen.net.cn (Q.G.),  
 zhimingzhao616@126.com (Z.Z.),  
 ronghuajin@ccmu.edu.cn (R.J.),  
 ruifuyang@gmail.com (R.Y.)

## In brief

Li and colleagues present the GlyExo-Capture method for isolating fucosylated extracellular vesicles (Fu-EVs) and identify five miRNAs for early HCC diagnosis. The method effectively detects early-stage HCC and differentiates it from other cancers, offering a high-throughput, rapid, and non-invasive screening approach.

## Highlights

- Introduction of the GlyExo-Capture approach for isolating extracellular vesicles
- Identification and validation of five miRNAs with strong diagnostic performance for HCC
- Effective detection of early-stage HCC and differentiation from other malignancies
- Development of a high-throughput, rapid, and non-invasive method for early HCC diagnosis

## Article

# Five miRNAs identified in fucosylated extracellular vesicles as non-invasive diagnostic signatures for hepatocellular carcinoma

Boan Li,<sup>1,8</sup> Kun Hao,<sup>2,7,8</sup> Mengyang Li,<sup>3,8</sup> Ailan Wang,<sup>2</sup> Huixue Tang,<sup>2</sup> Lida Xu,<sup>2</sup> Cuidie Ma,<sup>2</sup> Wenqian Du,<sup>2</sup> Lijuan Sun,<sup>4</sup> Xufeng Hou,<sup>4</sup> Tianye Jia,<sup>1</sup> Aixia Liu,<sup>1</sup> Qi Gao,<sup>2,4,9,\*</sup> Zhiming Zhao,<sup>3,\*</sup> Ronghua Jin,<sup>5,\*</sup> and Ruifu Yang<sup>6,7,\*</sup>

<sup>1</sup>Department of Laboratory Medicine, The Fifth Medical Center of Chinese PLA General Hospital, Beijing, China

<sup>2</sup>Beijing Hotgen Biotech Co., Ltd., Beijing, China

<sup>3</sup>The Faculty of Hepatopancreatobiliary Surgery, The First Medical Center of Chinese PLA General Hospital, Beijing, China

<sup>4</sup>Beijing Youngen Technology Co., Ltd., Beijing, China

<sup>5</sup>Capital Medical University Affiliated Beijing Ditan Hospital, Beijing, China

<sup>6</sup>Beijing Key Laboratory of POCT for Bioemergency and Clinic (No. BZ0329), Beijing, China

<sup>7</sup>State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, China

<sup>8</sup>These authors contributed equally

<sup>9</sup>Lead contact

\*Correspondence: [gaoqi@youngen.net.cn](mailto:gaoqi@youngen.net.cn) (Q.G.), [zhimingzhao616@126.com](mailto:zhimingzhao616@126.com) (Z.Z.), [ronghuajin@ccmu.edu.cn](mailto:ronghuajin@ccmu.edu.cn) (R.J.), [ruifuyang@gmail.com](mailto:ruifuyang@gmail.com) (R.Y.)

<https://doi.org/10.1016/j.xcrm.2024.101716>

## SUMMARY

Hepatocellular carcinoma (HCC) is a prevalent and aggressive cancer that presents significant challenges for early detection. This study introduces the GlyExo-Capture method for isolating fucosylated extracellular vesicles (Fu-EVs) from serum. We analyze microRNA (miRNA) profiles from Fu-EVs in 88 HCC patients and 179 non-HCC controls using next-generation sequencing (NGS) and identify five miRNAs (hsa-let-7a, hsa-miR-21, hsa-miR-125a, hsa-miR-200a, and hsa-miR-150) as biomarkers for HCC diagnosis. The five-miRNA panel demonstrates exceptional HCC diagnostic performance, with a sensitivity of 0.90 and specificity of 0.92 in a combined cohort of 194 HCC and 412 non-HCC controls, significantly surpassing the performance of alpha-fetoprotein (AFP) and des-gamma-carboxy prothrombin (DCP). Notably, the miRNA model achieves recall rates of 85.7% and 90.8% for stage 0 and stage A early-stage HCC, respectively, identifies 88.1% of AFP-negative HCC cases, and effectively differentiates HCC from other cancers. This study provides a high-throughput, rapid, and non-invasive approach for early HCC detection.

## INTRODUCTION

Liver cancer is a leading cause of cancer-related death globally, estimated to cause over one million fatalities by 2025.<sup>1</sup> Hepatocellular carcinoma (HCC) is the most prevalent class of liver cancer, accounting for approximately 85% of all primary hepatic malignancies.<sup>2</sup> The survival rate for HCC patients is strikingly low, with only 18% surviving beyond five years of diagnosis. This high mortality rate is largely attributed to the lack of effective early diagnosis strategies.<sup>3</sup> The prognosis for HCC depends on the stage of diagnosis. For patients with advanced-stage disease, the 5-year survival rate is below 5%, while it exceeds 70% for those diagnosed early.<sup>4</sup> The current standard approach for HCC screening and surveillance in patients with cirrhosis and subgroups with chronic hepatitis B virus infection involves abdominal ultrasound with or without serum alpha-fetoprotein (AFP) measurements.<sup>5</sup> In a meta-analysis of cohort investigations, ultrasound alone was found to have a sensitivity of only 45% for detecting early-stage HCC, rising to 63% alongside

AFP.<sup>6</sup> Furthermore, the utilization of HCC surveillance remains low, with under 20% of at-risk patients participating in early detection programs due to poor adherence to ultrasound screening.<sup>6</sup> Therefore, there is an urgent need for a minimally invasive and highly accurate approach for detecting HCC.

Liquid biopsy has been identified as a promising approach for early detection of HCC patients. It encompasses analyzing tumor components, primarily nucleic acids and tumor cells, which are released into the bloodstream by tumors.<sup>7</sup> In the realm of nucleic acids, microRNAs (miRNAs) are a category of single-stranded, small non-coding RNAs, typically 18–22 nucleotides long.<sup>8</sup> miRNAs, frequently dysregulated in multiple cancers, including HCC, are involved in various cancer biology processes, promoting tumor growth, apoptosis, progression, metastasis, immune evasion, and drug resistance.<sup>9–12</sup> Current research indicates that miRNAs show promise as potential candidates for liquid biopsy in the detection of human cancers.<sup>13,14</sup> In addition, circulating miRNAs have been improved to be promising molecules for early detection of HCC.<sup>15</sup> In another study, four miRNAs

were identified to be significantly differentially expressed between individuals with HCC and non-cancerous healthy individuals, indicating their potential as biomarkers for HCC detection. However, the specificity of miRNA-based biomarkers is challenged by the observation that cell-free miRNAs (cf-miRNAs) can originate from various sources, like apoptotic and immune cells, not only tumor cells.<sup>16</sup> Therefore, we infer that the potential heterogeneity associated with the origins of cf-miRNAs may limit the specificity of miRNA as a biomarker.

Another promising liquid biopsy approach involves examining of extracellular vesicles (EVs), which are lipid bilayer membrane-enclosed structures 40–160 nm in size, excreted by the majority of cells, and stably circulated in body fluid.<sup>17,18</sup> EVs have diverse biochemical signals, including genetic material, proteins, and lipids, from their cells of origin. Various studies suggest that EVs participate in cell-to-cell communication by delivering molecular cargo to target cells in the tumor microenvironment, promoting tumorigenesis and growth.<sup>19</sup> Furthermore, EVs have inherent stability due to their resilient lipid bilayers, allowing sustained circulation within the demanding tumor microenvironment and under physiological conditions.<sup>20–22</sup> This exceptional biological stability ensures reliable EV isolation and detection while allowing for extended specimen storage.<sup>23</sup> These advantages make EVs a reliable source of biomarkers for cancers like HCC.<sup>24–26</sup>

According to previous studies, we inferred that miRNAs within EVs are highly promising candidate biomarkers for the initial screening of liver cancer. However, challenges in efficiently isolating and specifically detecting EVs due to their small size and heterogeneity pose obstacles.<sup>27,28</sup> As cancer-derived EVs make up only a small fraction of the total EVs in bodily fluids, achieving ultra-sensitive and specific identification is imperative for advancing EV-based cancer diagnostics. Various methods have been developed for EV isolation and nucleic acids detection demonstrating progress.<sup>29–31</sup> However, challenges persist, including limited sensitivity and specificity, as well as issues linked to low purity and throughput, influencing both academic research and practical applications.<sup>32,33</sup>

Therefore, developing an effective approach to extracting and enriching tumor-derived EVs from serum is the pivotal challenge in leveraging EV cargo contents for tumor screening. Fortunately, tumor cells frequently produce surface glycans with distinct structures and expression levels compared to normal counterparts.<sup>34</sup> Therefore, the membrane surface of tumor-derived EVs should exhibit a significant abundance of fucosylated proteins, linked to the shared similarity in glycosylation structures between EVs and their secreted parent cells. Moreover, the excessive expression of “core” fucosylation is a crucial event in the development and progression of human HCC.<sup>35</sup> In this study, we developed an innovative method for EV isolation, termed GlyExo-Capture technology, leveraging the affinity of lectins for fucose on the EV membranes. This method can enrich fucosylated EVs (Fu-EVs) from liquid samples with high throughput and rapid turnaround times, enhancing the potential of EVs-based diagnostics in cancer. We performed an extensive small RNA transcriptomic profiling on a substantial quantity of clinical samples from patients with HCC, cirrhosis, and hepatitis and healthy controls. The primary objective was to identify an EV-

miRNA signature capable of distinguishing HCC from non-HCC conditions. Subsequent to the identification of potential biomarkers, we confirmed our classifier model across multiple independent clinical cohorts, advancing HCC detection methods.

## RESULTS

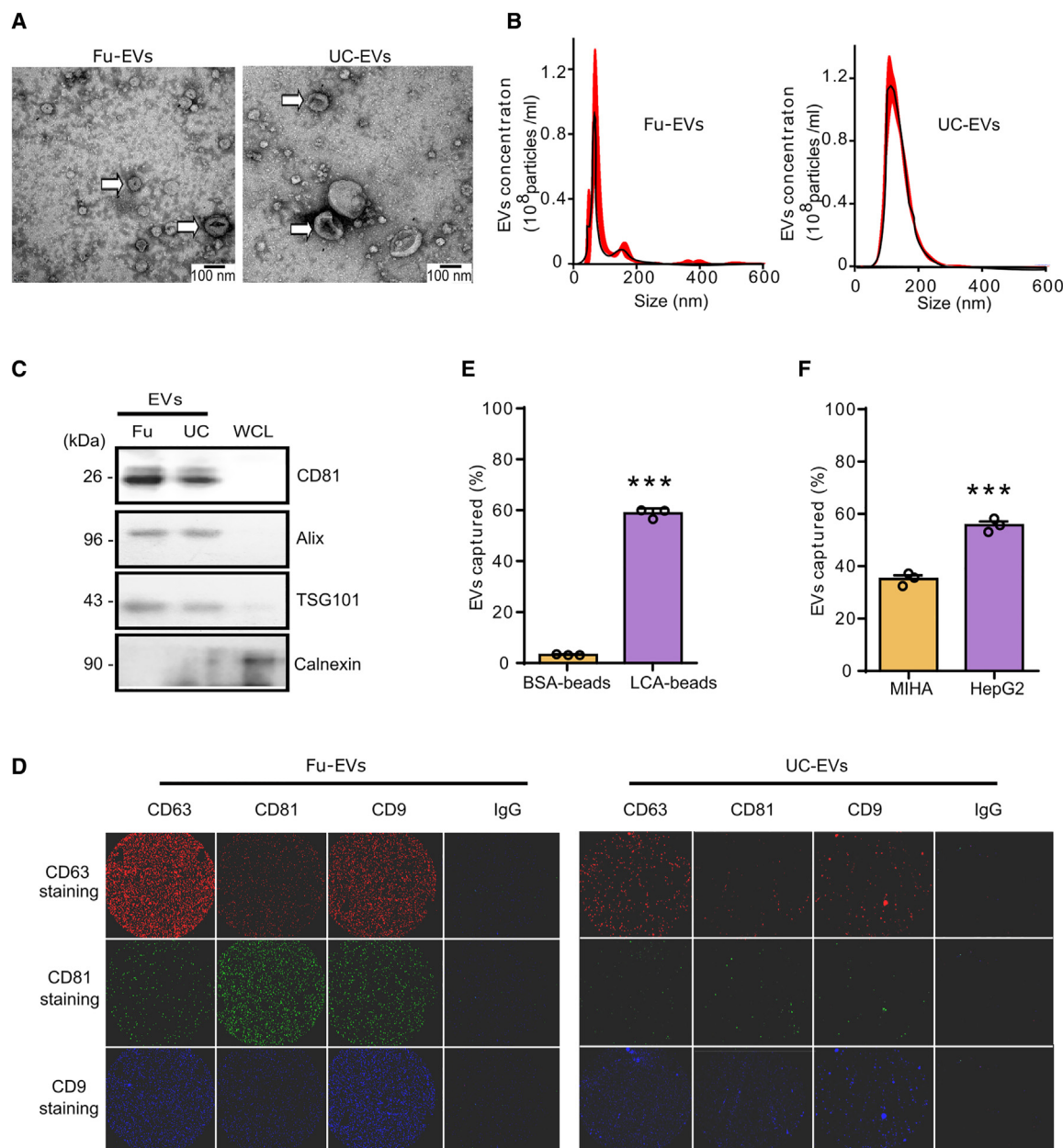
### The GlyExo-Capture approach substantially enriches Fu-EVs

We developed a GlyExo-Capture method to specifically capture Fu-EVs from liquid samples, utilizing lectin immobilized on hydroxyl macromolecular magnetic beads (Figure S1). Notably, this method enabled the entire process of preparing 96 samples to be completed in 11 min. We conducted a comprehensive evaluation of Fu-EVs isolated by GlyExo-Capture with EVs isolated by ultracentrifugation (UC) (referred to as UC-EVs). Transmission electron microscopy analysis uncovered that both Fu-EVs and UC-EVs exhibited oval or bowl-like shapes (Figure 1A). Fu-EVs exhibited a smaller mean size ( $106.5 \pm 6.7$  nm) compared to UC-EVs ( $137.5 \pm 0.4$  nm) yet had a broader size distribution (Figure 1B). Further validation through immunoblot and ExoView analyses confirmed the presence of EV membrane protein markers, including CD81, CD63, and CD9, as well as tetraspansins like TSG101 and ALIX (Figures 1C and 1D). These findings collectively confirm the successful isolation of Fu-EVs. To further characterize Fu-EVs, we evaluated their uptake dynamics using flow cytometry. Both Fu-EVs and UC-EVs were labeled with 1,1'-dioctadecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate (DiD). As illustrated, Fu-EVs exhibited time- and dose-dependent uptake patterns (Figures S2A and S2B). Notably, Fu-EVs exhibited increased uptake efficiency compared to UC-EVs (Figures S2C–S2F). Consistent with this, the enhanced uptake efficiency of Fu-EVs was reduced to levels similar to UC-EVs upon enzymatic removal of N-glycans using peptide-N-glycosidase F (PNGase F).

To assess the efficacy of lectin-immobilized beads, UC-EVs were separately incubated with lectin- or BSA-immobilized beads. Further quantitative analysis of the unbound fractions revealed a 40% reduction in the number of UC-EVs particles in the unbound fraction following incubation with lectin-immobilized beads (Figure 1E). In addition, the enrichment of Fu-EVs secreted by HepG2 cells was significantly higher relative to those obtained from non-cancerous immortalized human hepatocyte cell line (MIHA) (Figure 1F). Overall, the GlyExo-Capture method is a rapid and reliable approach for the high-throughput isolation of Fu-EVs.

### EV-miRNA expression profiles of the discovery cohort identify five HCC-specific miRNA signatures

Given the increased presence of glycans on EVs originating from cancer cells, there is substantial promise in identifying miRNA cargos within enriched Fu-EVs for discovering more specific and sensitive cancer biomarkers. To assess this potential, we obtained a cohort of 88 patients diagnosed with HCC and 179 non-HCC controls, including 49 healthy individuals, 62 hepatitis, 54 cirrhosis, and 14 benign hepatic tumor patients. Next-generation sequencing (NGS) was employed to compare the Fu-EVs miRNA expression files across 88 HCC patients and



**Figure 1. Characterization and comparative analysis of EVs isolated using GlyExo and UC methods**

(A) Wide-field electron microscopy images of EVs isolated by GlyExo and UC methods, with representative EVs highlighted by arrows (scale bar: 100 nm).

(B) Size distribution measurements of EVs using nanoparticle tracking analysis (NTA).

(C) Western blot analysis showing Fu-EVs, UC-EVs, and whole-cell lysates (WCLs).

(D) Analysis of EVs through ExoView using the specified antibodies.

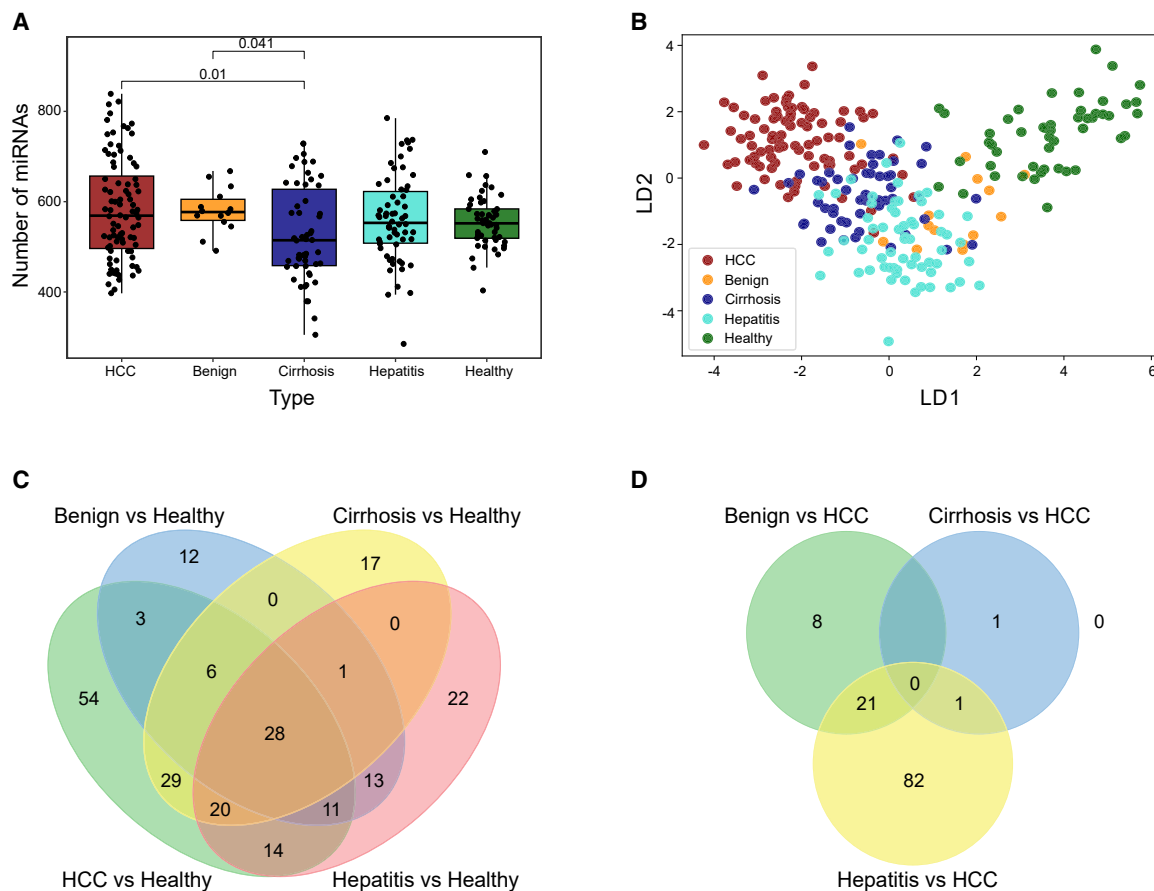
(E) Comparison of EV capture approaches using lectin- or BSA-immobilized beads; the captured EVs are quantified as the bound fraction relative to the UC-EVs loaded.

(F) Comparison between EVs isolated from the cancer cell line HepG2 and the non-cancerous MIHA cell line. Experiments were performed in triplicate, with statistical significance determined using two-tailed Student's t tests. \*,  $p < 0.05$ ; \*\*,  $p < 0.01$ ; \*\*\*,  $p < 0.001$ .

179 non-HCC controls. The quality of the NGS data is displayed in Figures S3A and S3B.

A total of 2,278 known miRNAs were identified (Figure 2A). Differential analysis between the HCC group and the non-HCC

group was conducted using the R software package DESeq2, leading to the identification of 112 miRNAs with significant differences (Figure S4). Moreover, linear discriminant analysis was conducted using the 112 differentially expressed miRNAs



**Figure 2. Analysis of EVs-miRNA profiles in various cohorts**

(A) Distribution of annotated miRNA counts in healthy, hepatitis, and HCC groups. The values 0.01 and 0.041 within the figure represent the *p* values from the Wilcoxon's rank-sum test

(B) Linear discriminant analysis of significantly different miRNAs among five populations.

(C) Venn diagram showing differentially expressed miRNAs across disease groups compared to the healthy cohort.

(D) Venn diagram illustrating differentially expressed miRNAs in benign liver disease compared to the HCC group.

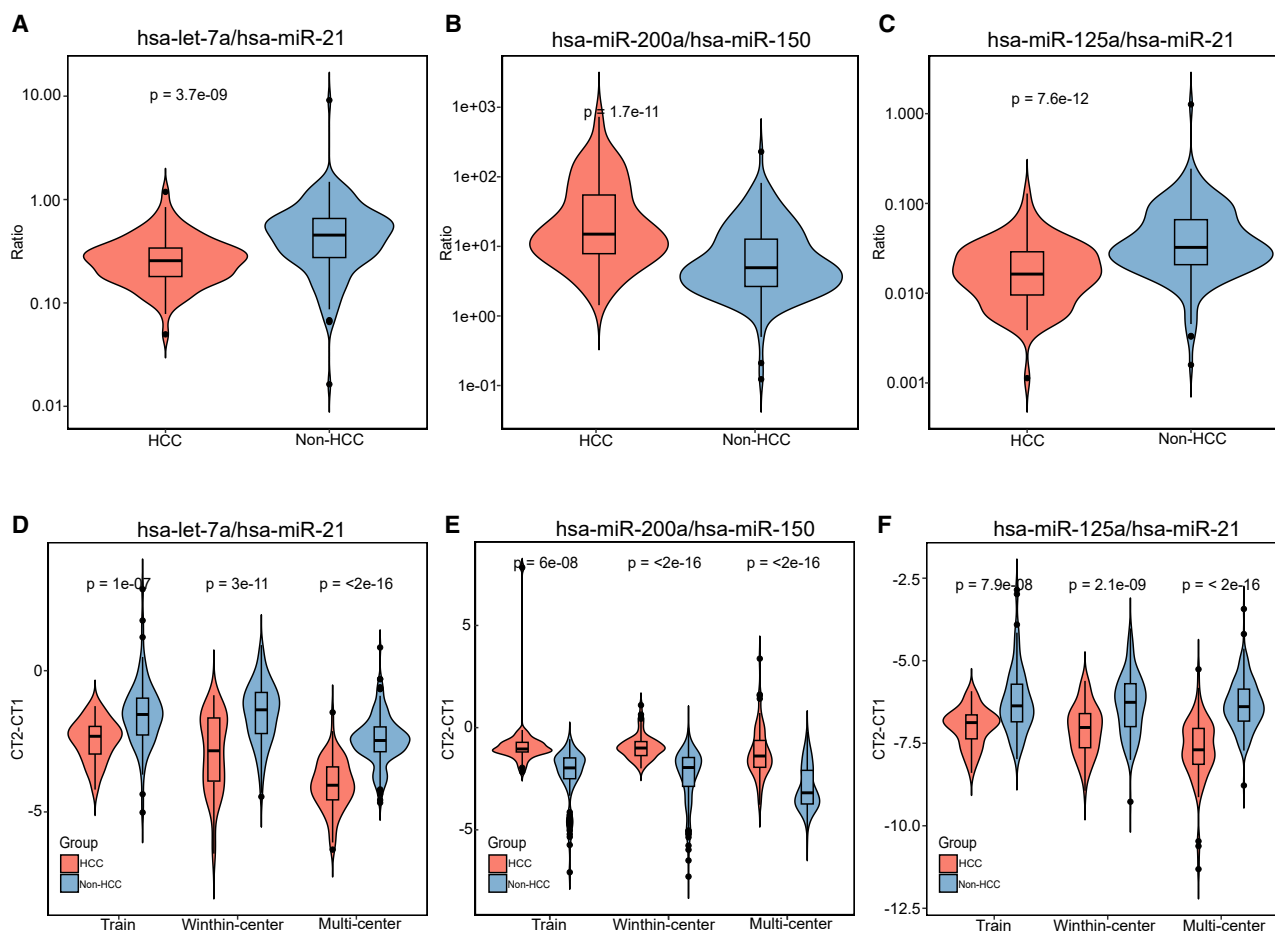
(DEMs) across the five groups. These five groups could be effectively segregated beginning from a healthy state, progressing to hepatitis and cirrhosis, and culminating in HCC (Figure 2B). Using the same approach, we conducted differential expression analysis for multiple comparisons, revealing a common set of DEMs common across the four non-healthy groups and the healthy group (Figure 2C). Interestingly, all 28 miRNAs in this common intersection were part of the 122 DEMs between the HCC and non-HCC groups. However, no overlap was detected in the DEMs between the HCC group and the three non-cancerous liver disease groups (Figure 2D).

We narrowed down the aforementioned 112 miRNAs to 35 miRNAs utilizing recursive feature elimination. From these, we identified the ten miRNAs based on their significance in the random forest feature importance analysis. The expression comparison of the ten miRNAs in both the HCC and non-HCC control groups is illustrated in Figures S5A–S5J. Thereafter, we validated these ten miRNAs using quantitative reverse-transcription PCR (RT-qPCR) and refined our selection to five miRNAs exhib-

iting significant signaling disparities. Since pairs of miRNAs can mitigate experimental system errors and enhance the signal-to-noise ratio, these five DEMs were selected to establish biologically relevant target pairs. Using five specific miRNAs, including hsa-let7a, hsa-miR-21, hsa-miR-200a, hsa-miR-150, and hsa-miR-125a, we identified three pairs of miRNAs. The resulting miRNA ratio signatures, acquired via both NGS (Figures 3A–3C) and RT-qPCR (Figures 3D–3F), consistently demonstrated significant differences between HCC and non-HCC control groups.

### The HCC prediction model using three miRNA ratios is successfully constructed and validated throughout multiple cohorts

To assess the efficacy of these candidate miRNA ratios as biomarker panels, logistic regression models were constructed using NGS data from a discovery cohort of 267 individuals and RT-qPCR data from training and validation cohorts totaling 779 individuals. Receiver operating characteristic (ROC) curve



**Figure 3. Comparison of the miRNA ratio signatures distinguishing HCC from controls using NGS and RT-qPCR**

Control groups, designated as non-HCC, include individuals with cirrhosis, hepatitis, benign hepatic tumors, and healthy controls.

(A–C) Expression ratio signatures measured by NGS: (A) hsa-let-7a/hsa-miR-21, (B) hsa-miR-200a/hsa-miR-150, and (C) hsa-miR-125a/hsa-miR-21.

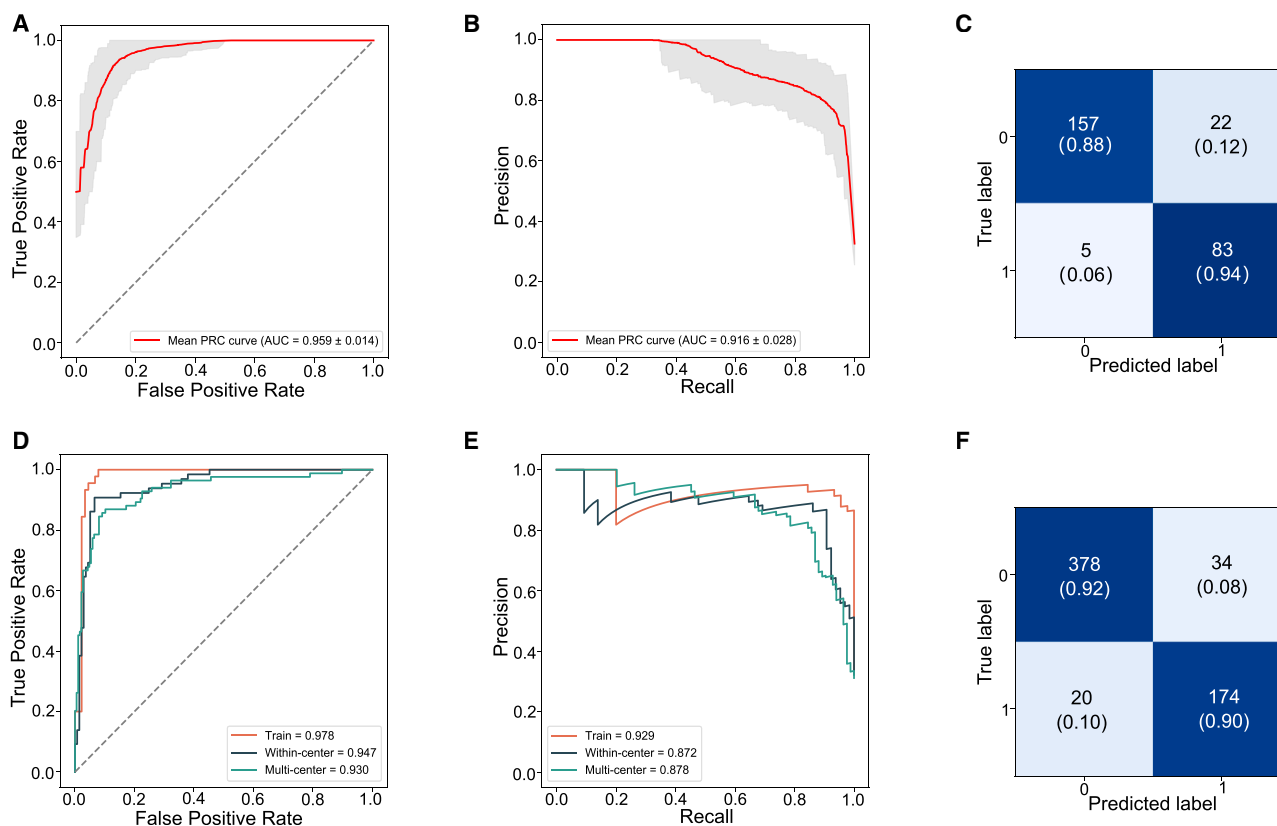
(D–F) Expression ratio signatures measured by RT-qPCR: (D) hsa-let-7a/hsa-miR-21, (E) hsa-miR-200a/hsa-miR-150, and (F) hsa-miR-125a/hsa-miR-21. Each experiment was conducted with biological replicates.

The *p* values represent those from the *t* test.

analysis demonstrated an area under the curve (AUC) of 0.959 (Figure 4A), while precision-recall curve (PRC) analysis had a precision-recall area under the curve (PRAUC) of 0.916 (Figure 4B) for the NGS data. The confusion matrix (Figure 4C) from the discovery cohort exhibited a sensitivity of 94.32% and a specificity of 87.71%, highlighting the model's robust diagnostic performance. Utilizing the RT-qPCR data for the same three candidate miRNA ratios, the model was trained using a training cohort and validated on two distinct cohorts: one from the same center (within-center) and another including multiple centers (multi-center). Across these cohorts, the model achieved AUCs of 0.978, 0.947, and 0.930 in the training, within-center, and multi-center cohorts, respectively (Figure 4D), whereas the corresponding PRAUCs were 0.929, 0.872, and 0.878 (Figure 4E). Sensitivity and specificity values were computed according to all RT-qPCR data, with values of 92% and 90%, respectively. Moreover, the model's performance metrics were calculated across the three cohorts and outlined in Table 1.

Notably, within the non-HCC group, specificity exceeded 97% in healthy control samples and surpassed 80% in cases of benign hepatic diseases. Furthermore, during external validation over multiple centers, the model demonstrated robust performance, attaining a sensitivity of 86.90% and a specificity of 89.25% at the established optimal cutoff derived from the training set.

To identify the specificity of the three-miRNA ratio signature for HCC, we obtained data from 173 cases of various malignancies, including gastric, lung, breast, and colorectal cancers (Table S1). Analysis of the three-miRNA ratio signature model prediction scores uncovered that most individuals in this cohort exhibited positive scores below 0.5 (Figure S6A). Notably, the highest specificity of 0.9 was observed for breast cancer, with the remaining three cancers also exhibiting specificity over 0.8 (Figure S6B). In summary, our study has established a model capable of highly accurate clinical diagnosis for HCC.



**Figure 4. Evaluation of the EVs-miRNA signature performance in clinical cohorts**

(A–C) Receiver operating characteristic (ROC) analysis, precision-recall curve (PRC) analysis, and confusion matrix analysis of prediction models using a three-miRNA ratio signature measured by NGS with 1,000 bootstrap sampling in 267 samples. (D–F) ROC analysis, PRC analysis, and confusion matrix for prediction models utilizing a three-miRNA ratio signature measured by RT-qPCR in training, within-center, and multi-center cohorts ( $n = 606$ ).

### Five miRNAs and their target genes enriched in tumorigenic pathways are altered in HCC-EV-receiving cells

To further confirm the alternation of the five miRNAs in Fu-EVs-receiving cells, human colon tumor cell line 116 (HCT-116) were treated with Fu-EVs isolated from HCC patients (HCC-EVs) or healthy controls (Ctr-EVs). Compared to cells treated with Ctr-EVs, the expression of the five miRNAs was significantly impacted in cells treated with HCC-EVs, determined by RT-qPCR (Figure S7). NGS analysis was employed to identify genes with altered expression related to the five miRNAs. Significant transcriptional changes were identified in cells treated with HCC-EVs (Table S2). We compared these changes with 234 predicted or reported target genes from the starBase database. The hypergeometric distribution test revealed a significant enrichment of differentially expressed genes within the 234 genes (Table S2), demonstrating a notable overrepresentation of targets specifically related to these miRNAs. The regulatory network between the five miRNAs and their 82 targeted genes is illustrated in Figure S8A. Pathway enrichment analysis of the identified 82 genes unveiled significant enrichment in tumorigenic pathways, especially the PI3K-Akt signaling pathway, cell-cycle regulation, and HCC pathway (Figure S8B).

### EV-miRNAs combined with serum protein biomarkers enhance the diagnostic accuracy of HCC

In clinical practice, AFP and des-gamma-carboxy prothrombin (DCP) serve as biomarkers for diagnosing HCC patients. However, their sensitivity and specificity for early detection in the general population remain inadequate. We hypothesized that integrating miRNAs with AFP and/or DCP could enhance clinical diagnostic efficacy for HCC. Consequently, we assessed the diagnostic performance of AFP and DCP alone, in combination with our miRNA ratio signatures. First, we conducted an analysis of model development and evaluation within the combined cohort of 706 samples from the three groups. As anticipated, developing the model using AFP or DCP alone resulted in AUC values of 0.79 and 0.81, respectively, significantly lower than the model based on the three-miRNA ratio signature, achieving an AUC of 0.95. Furthermore, combining AFP with miRNA signature enhanced overall diagnostic performance, yielding AUC values of 0.96 (Figure 5A). The PRC supports this conclusion, with PRAUC increasing from 0.89 for miRNA to 0.91 with the addition of AFP (Figure 5B). Additionally, we examined the detection outcomes of AFP, DCP, miRNA, and their combination (AFP+miRNAs, DCP+miRNAs, AFP+DCP+miRNAs) across all patients diagnosed with HCC (Figure 5C). The miRNA-based

**Table 1. Summary of diagnostic performance of classifier mode in the training and validation cohorts**

|                   |              | Train  |             |         | In-center validation |             |         | Multi-center validation |             |         |
|-------------------|--------------|--------|-------------|---------|----------------------|-------------|---------|-------------------------|-------------|---------|
|                   |              | HCC    | Non-HCC     |         | HCC                  | Non-HCC     |         | HCC                     | Non-HCC     |         |
|                   |              |        | Non-healthy | Healthy |                      | Non-healthy | Healthy |                         | Non-healthy | Healthy |
| miRNA model       | Pre-HCC      | 42     | 3           | 0       | 59                   | 10          | 1       | 73                      | 18          | 2       |
|                   | Pre-non-HCC  | 3      | 42          | 44      | 6                    | 62          | 64      | 11                      | 81          | 85      |
|                   | Sensitivity/ | 93.33% | 93.33%      | 100%    | 90.77%               | 86.11%      | 98.46%  | 86.90%                  | 81.82%      | 97.70%  |
|                   | specificity  |        | 96.63%      |         |                      | 91.97%      |         |                         | 89.25%      |         |
| miRNA & AFP model | Pre-HCC      | 44     | 3           | 0       | 60                   | 9           | 1       | 75                      | 16          | 0       |
|                   | Pre-non-HCC  | 1      | 42          | 44      | 5                    | 63          | 64      | 9                       | 83          | 87      |
|                   | Sensitivity/ | 97.78% | 93.33%      | 100%    | 92.31%               | 87.50%      | 98.46%  | 89.29%                  | 83.84%      | 100%    |
|                   | specificity  |        | 96.63%      |         |                      | 92.70%      |         |                         | 91.40%      |         |

classification model achieved an 88.06% detection rate in 67 AFP-negative patients, while the combined miRNA and AFP model achieved an overall recall rate of 90.70% over 194 HCC patients. This finding highlights the complementary nature of miRNA signatures and AFP, emphasizing the improved diagnostic performance achieved through combined detection. Further analyses specific to each cohort are shown in Figures 5D–5F, 5G–5I, and 5J–5L, illustrating ROC, PRC, and HCC recall for the Fifth Medical Center of Chinese PLA General Hospital (cohort 1), the First Medical Center of Chinese PLA General Hospital (cohort 2), and Capital Medical University Affiliated Beijing Ditan Hospital (cohort 3), respectively. Cohort 2 exhibited superior metrics, while cohort 1 and cohort 3 showed slightly lower performance. Nevertheless, the combined analysis across all cohorts provides a larger sample size, enhancing the model's capability for generalization. In summary, DCP exhibits superior specificity compared to AFP and miRNA features. In contrast, miRNA substantially enhances the sensitivity of HCC diagnosis compared to these protein biomarkers alone. Moreover, combining miRNA with AFP further enhances overall diagnostic accuracy (Tables S3–S6).

### The miRNA model exhibits superior diagnostic performance across various clinical indicators

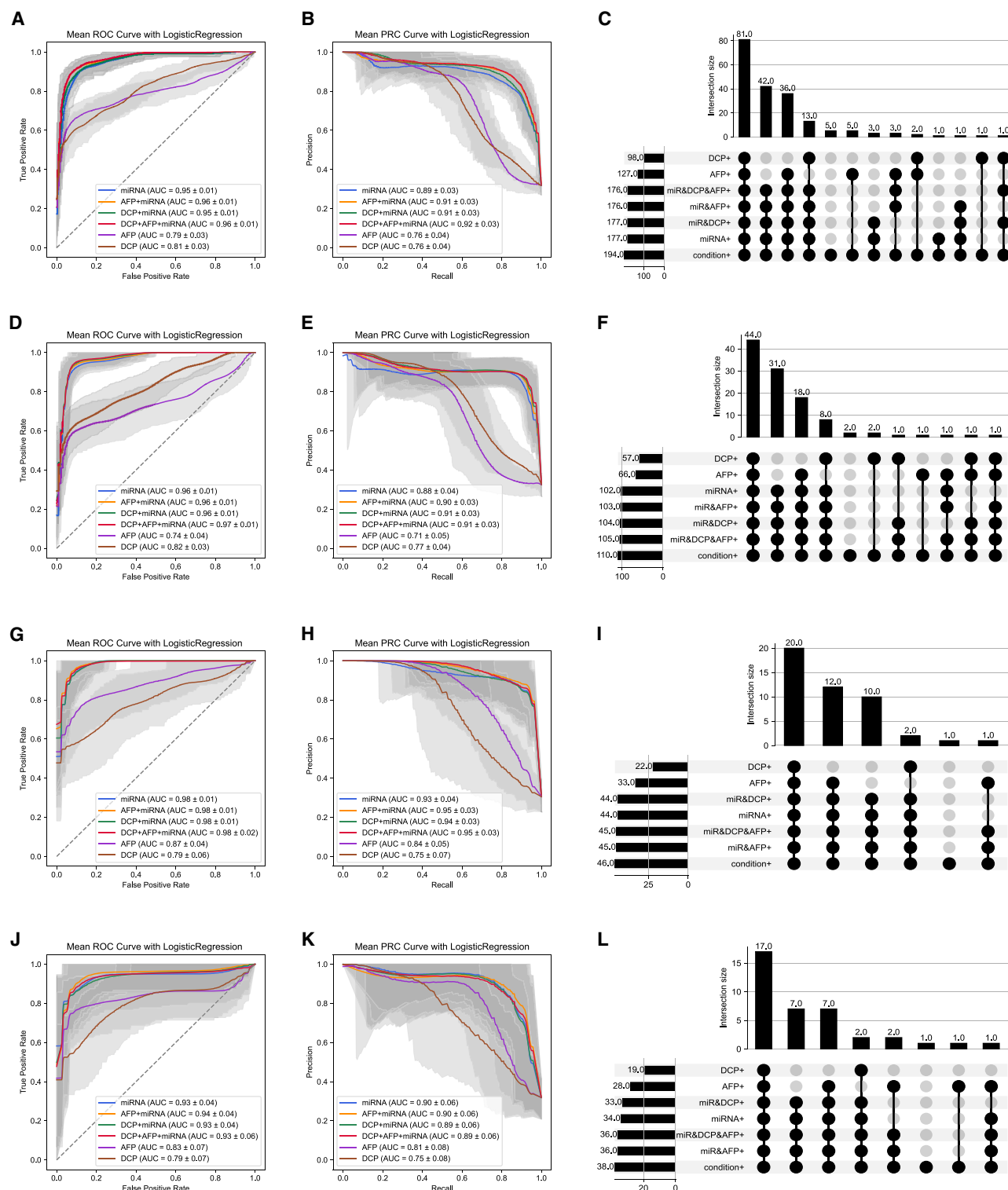
While the model's performance was previously evaluated across three distinct cohorts, we have conducted a more extensive investigation to better comprehend its clinical applicability. This involved considering additional clinical parameters and performing statistical analyses to evaluate the model's effectiveness across various indicators. In all RT-qPCR cohorts, comprising healthy controls, patients with benign hepatic diseases, and those with HCC, we systematically ranked the model's prediction probability scores in ascending order, as illustrated in Figure 6A. Notably, HCC samples with prediction scores below the threshold of 0.5 often exhibited DCP negativity and small tumor sizes, demonstrating low tumor burdens. The miRNA model accurately predicted all cases of benign hepatic tumors, 99.5% of healthy individuals, 85.2% of cirrhosis cases, and 86.3% of hepatitis cases (Figure 6B). For a more detailed analysis of diagnostic accuracy in cirrhosis, we examined the recall rates across different Child-Pugh stages. The model accurately predicted all patients classified under Child-Pugh stage C,

89.7% of those in stage A, and 73.3% of those in stage B (Figure 6C). This demonstrates the model's high accuracy in predicting cirrhosis in various stages. We also assessed the diagnostic accuracy of HCC patients, stratifying them according to AFP and DCP. The model showed a sensitivity of 88.1% among AFP-negative HCC patients (AFP <7) and 86.5% for DCP-negative patients (DCP <40) (Figures 6D and 6E). In comparison, the diagnostic performance of AFP and DCP alone in HCC patients across various disease stages, tumor sizes, and dimensions was consistently inferior to that of the miRNA classifier model (Figures 6F–6H). Notably, the miRNA model exhibited a positive detection rate of 85.7% in early-stage HCC patients classified under Barcelona Clinic Liver Cancer (BCLC) staging system, while both AFP and DCP showed positive detection rates of only 14.3%. These compelling results underscore the significant clinical value of the EV-miRNA ratio signature, demonstrating sustained efficacy even in early-stage HCC cases.

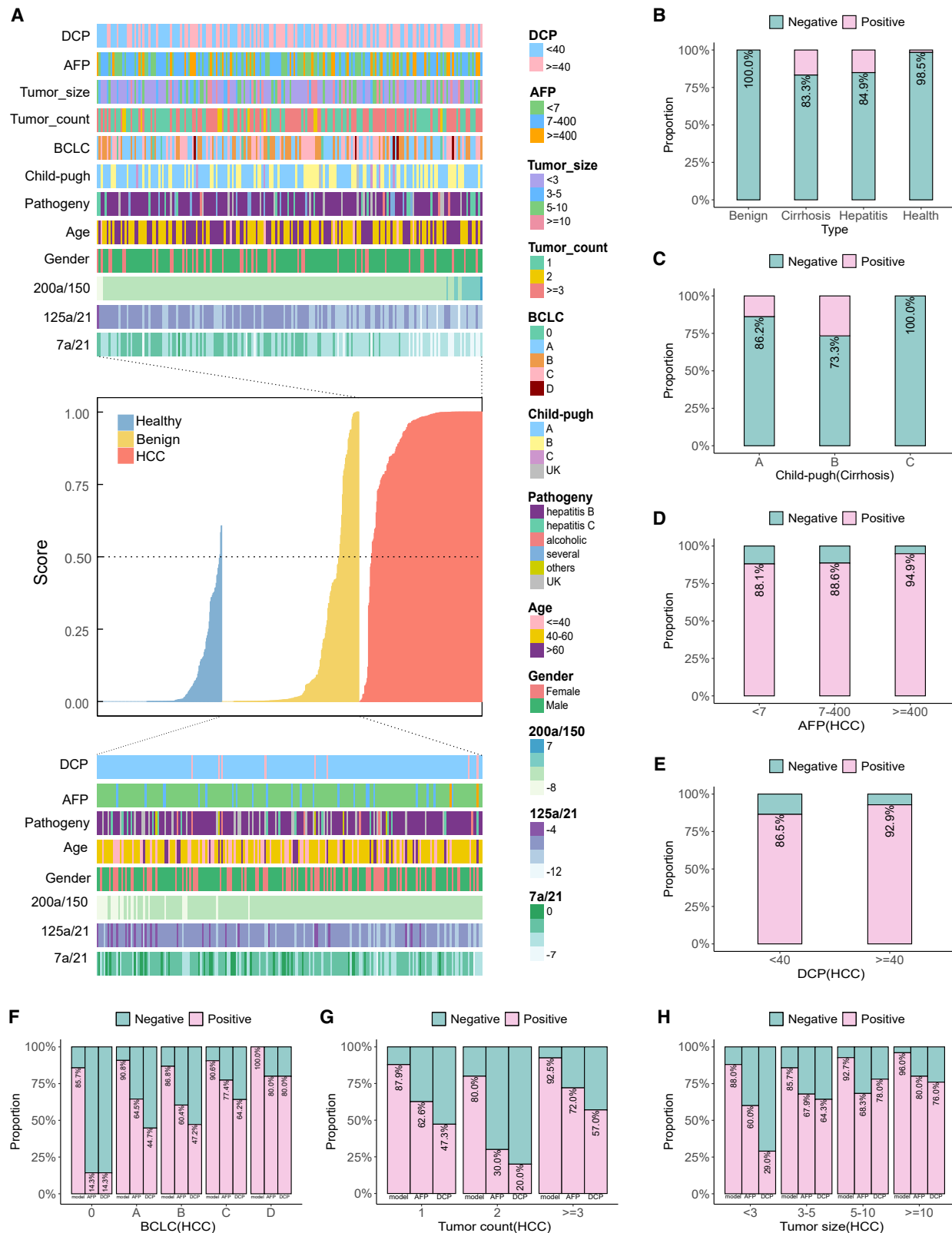
## DISCUSSION

In this study, we developed an approach for efficiently extracting and enriching Fu-EVs, characterized by high throughput, rapidity, and user-friendly operation, rendering it suitable for clinical biomarker detection. Using this extraction methodology, we isolated Fu-EVs from the serum of 276 individuals and analyzed their miRNA expression profiles using NGS. This extensive dataset enabled comparative analysis of Fu-EVs miRNA expression patterns across diverse groups, including patients diagnosed with HCC, cirrhosis, hepatitis, benign hepatic tumor, and healthy individuals. Moreover, we developed a highly accurate logistic regression model based on a three-miRNA ratio signature, enabling precise differentiation between HCC patients and non-HCC controls. The model underwent rigorous validation through RT-qPCR over multiple cohorts, comprising 779 individuals.

Despite the varied strategies developed for profiling EV composition,<sup>36</sup> the development of efficient and high-throughput isolation methods remains a persistent technical challenge. In clinical practice, altered glycosylation patterns within glycoproteins and glycoconjugates are valuable serological biomarkers for cancer diagnosis.<sup>37,38</sup> EVs secreted by cancer cells are enriched with diverse aberrant glyco-signatures, including



**Figure 5. Performance comparison of prediction models based on individual and combined features using logistic regression** (A, D, G, and J) Receiver operating characteristic (ROC) and (B, E, H, and K) precision-recall curve (PRC) analysis comparing diagnostic performance among AFP, DCP, miRNA, and their combinations of miRNA with these proteins. (C, F, I, and L) Upset plots showing positive detection of hepatocellular carcinoma (HCC) samples by models with different feature combinations. Numbers in the upper panel bars indicate sample counts identified by multiple models, while numbers in left panel bars denote detection by models with individual and combined features. Results are presented for combined cohorts and individual cohorts 1, 2, and 3.



(legend on next page)

mannose, complex type N-glycans, polylectosamine, and sialylated glycans,<sup>36</sup> making them promising targets for affinity purification. Therefore, we introduced a method for isolating Fu-EVs from serum samples using a fucose-specific lectin. Compared to UC, our GlyExo-Capture method preserves EV morphology, offers a broader size distribution, and enhances purity. Moreover, its rapid processing time and high throughput represent a significant advance in EV isolation, facilitating its potential streamlined clinical application. Our investigation reveals a heightened prevalence of glycosylated EVs in liver cancer cells relative to their healthy counterparts (Figure 1F). Furthermore, tumor cells exhibit a significantly increased uptake rate of glycosylated EVs relative to non-glycosylated EVs (Figures 1 and S2). These findings highlight the pronounced inclination of liver cancer cells toward both the secretion and uptake of glycosylated EVs. Prior research has uncovered the role of surface proteins on both EVs and recipient cells in facilitating EV uptake.<sup>39</sup> Therefore, we postulate that glycosylated EVs exert a more profound influence on the initiation and progression of liver cancer. This is supported by the promising diagnostic outcomes achieved using miRNA derived from Fu-EVs as biomarkers for HCC diagnosis in our study.

miRNAs play a pivotal role in maintaining biological processes essential for homeostasis, significantly contributing to the balance necessary for liver function.<sup>40</sup> Consequently, miRNAs, particularly those within EVs well-preserved by bilayer membranes, emerge as promising candidates for the early-stage diagnostic of various malignancies.<sup>12,41</sup> In this study, we developed classifier models integrating an Fu-EVs-derived three-miRNA ratio signature with AFP and/or DCP, employing logistic regression, random forest, support vector classification (SVC), and multilayer perceptron (MLP) algorithms (Figures 5 and S9–S11; Tables S3–S6). Among these models, logistic regression demonstrated the highest performance, followed by random forest, while MLP and SVC showed relatively poorer performance, particularly in the cohort from the First Medical Center of Chinese PLA General Hospital. The area under the precision-recall curve for SVC models was notably lower by approximately 10 percentage points compared to logistic regression and random forest. Our results showed superior performance of the miRNA-based model compared to models using AFP/DCP features alone (Figures 5 and S9; Tables S3–S6). While integrating miRNAs with AFP or DCP marginally improved performance, simultaneous integration did not produce additional enhancement. Notably, the miRNA model effectively identified a substantial proportion of AFP-negative HCC patients, complementing AFP's detection and reducing false negatives. Our comprehensive analysis, en-

compassing diverse clinical parameters, exhibited robust predictive capabilities across different cohorts, with particular efficacy observed in early-stage HCC diagnosis according to the BCLC staging criteria (Figure 6F). Lower prediction scores in HCC samples were related to DCP-negative statuses and smaller tumor burdens. Notably, the favorable predictive performance of the miRNA model in tumors outside of HCC suggests a robust specificity of the three-miRNA ratio signature for HCC. These findings underscore the clinical significance and robustness of our miRNA-based diagnostic model, especially for early-stage HCC detection.

In conclusion, our study presents Fu-EVs-derived miRNA signatures developed using our innovative EV capture method tailored for identifying HCC patients. The validated three-miRNA ratio signatures underwent rigorous validation across multiple independent cohorts, including HCC patients and non-HCC controls. When combined with AFP expression levels, our classifier model exhibited enhanced diagnostic performance compared to routine clinical assessment relying solely on AFP. Notably, our miRNA biomarkers demonstrated clinical relevance by accurately characterizing AFP-negative HCC patients, highlighting their potential as reliable and non-invasive biomarkers for early detection of HCC.

### Limitations of the study

Although our HCC diagnosis model has been validated in three distinct cohorts, prospective cohort studies are also crucial for confirming the utility of these biomarkers in early diagnosis within real-world clinical settings. Such studies will enhance the credibility and clinical feasibility of our findings. Additionally, a comparative analysis of the miRNA-based model with established clinical biomarkers and imaging techniques in prospective cohorts is essential. This approach will provide a comprehensive evaluation of the model's diagnostic accuracy and clinical relevance, facilitating its potential integration into routine clinical practice and improving patient management.

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Qi Gao ([gaoqi@youngen.net.cn](mailto:gaoqi@youngen.net.cn)).

#### Materials availability

Serum or RNA generated in this study are available from the lead contact with a completed Materials Transfer Agreement (some samples are depleted). Glycosylated Exosome Extraction Kits are available for purchase from Hotgen Biotech.

### Figure 6. Performance evaluation of the miRNA model under various clinical factors

- (A) Logistic regression model prediction probability scores for a cohort of 606 individuals, showing scores for various clinical factors and miRNA expression ratios.  
(B) Predictions for benign liver tumors, cirrhosis, hepatitis, and health populations.  
(C) Predictions for cirrhosis patients categorized by Child-Pugh class A, B, and C.  
(D) Predictions for HCC patients categorized by AFP levels (<7  $\mu\text{g/L}$ , 7–400  $\mu\text{g/L}$ , and >400  $\mu\text{g/L}$ ).  
(E) Predictions for HCC patients categorized by DCP levels (<40  $\mu\text{g/L}$ ,  $\geq 40$   $\mu\text{g/L}$ ).  
(F) Predictions for HCC patients categorized by BCLC stages (0, A, B, C, and D) using miRNA, AFP, and DCP model, respectively.  
(G) Predictions for HCC patients categorized by tumor count (1, 2, and  $\geq 3$ ) using miRNA, AFP, and DCP model, respectively.  
(H) Predictions for HCC patients categorized by tumor size (<3 cm, 3–5 cm, 5–10 cm, and  $\geq 10$  cm) using miRNA, AFP, and DCP model, respectively. Each model's performance is based on multiple biological replicates.

### Data and code availability

Messenger RNAseq data generated during the current study are available from the NCBI's Gene Expression Omnibus database (GEO: GSE266634). The scripts used for machine learning modeling are available on Zendo (<https://doi.org/10.5281/zenodo.13182532>). Any additional information required to re-analyze the data reported in this work is available from the [lead contact](#) upon request.

### ACKNOWLEDGMENTS

We thank Yufei Wang and Wade Wang (Quantum Design China and NanoView Biosciences) for supporting the characterization of extracellular vesicles. We also thank Professor Hui Zhong (Beijing Youngen Technology Co., Ltd., China) for her invaluable assistance in revising the manuscript. This work was funded by the Capital's Funds for Health Improvement and Research (SF2024-1-2171) and National Key Research and Development Program of China (2023YFC2507402). K.H. is partially supported by the Beijing Postdoctoral Research (no. 2021-ZZ-044).

### AUTHOR CONTRIBUTIONS

K.H., C.M., W.D., L.S., and X.H. conducted the experiments. B.L., M.L., A.W., and H.T. were responsible for manuscript drafting. H.T. and A.W. performed the bioinformatic analysis. B.L., M.L., T.J., A.L., and R.J. were responsible for sample collection. Q.G. and L.X. provided technical support. R.Y., R.J., Q.G., and Z.Z. acquired funding to financially support this study and helped with the conception for this study. A.W. and L.X. revised the manuscript.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
  - Human subjects and sample preparation
- **METHOD DETAILS**
  - Sera sample collection
  - EV isolation and characterization
  - Cellular uptake of EVs
  - EV-derived small RNA sequencing and analysis
  - qRT-PCR
  - Measurement of AFP and DCP levels
  - Construction and evaluation of the HCC prediction model
  - Messenger RNA NGS sequencing and data analysis
  - miRNA-mRNA regulatory network analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xcrm.2024.101716>.

Received: December 28, 2023

Revised: May 25, 2024

Accepted: August 13, 2024

Published: September 5, 2024

### REFERENCES

1. Llovet, J.M., Kelley, R.K., Villanueva, A., Singal, A.G., Pikarsky, E., Roayaie, S., Lencioni, R., Koike, K., Zucman-Rossi, J., and Finn, R.S. (2021). Hepatocellular carcinoma. *Nat. Rev. Dis. Primers* 7, 6–28.
2. Villanueva, A. (2019). Hepatocellular Carcinoma. *N. Engl. J. Med.* 380, 1450–1462.
3. Jemal, A., Ward, E.M., Johnson, C.J., Cronin, K.A., Ma, J., Ryerson, B., Mariotto, A., Lake, A.J., Wilson, R., Sherman, R.L., et al. (2017). Annual report to the nation on the status of cancer, 1975–2014, featuring survival. *J. Natl. Cancer Inst.* 109, dx030.
4. Parikh, N.D., Tayob, N., and Singal, A.G. (2023). Blood-based biomarkers for hepatocellular carcinoma screening: Approaching the end of the ultrasound era? *J. Hepatol.* 78, 207–216.
5. Nault, J.C., and Villanueva, A. (2021). Biomarkers for hepatobiliary cancers. *Hepatology* 73, 115–127.
6. Tzartzeva, K., Obi, J., Rich, N.E., Parikh, N.D., Marrero, J.A., Yopp, A., Waljee, A.K., and Singal, A.G. (2018). Surveillance imaging and alpha fetoprotein for early detection of hepatocellular carcinoma in patients with cirrhosis: a meta-analysis. *Gastroenterology* 154, 1706–1718.e1.
7. Labgaa, I., and Villanueva, A. (2015). Liquid biopsy in liver cancer. *Discov. Med.* 19, 263–273.
8. Di Leva, G., Calin, G.A., and Croce, C.M. (2006). MicroRNAs: fundamental facts and involvement in human diseases. *Birth Defects Res. C Embryo Today* 78, 180–189.
9. Kasinski, A.L., and Slack, F.J. (2011). MicroRNAs en route to the clinic: progress in validating and targeting microRNAs for cancer therapy. *Nat. Rev. Cancer* 11, 849–864.
10. Lin, S., and Gregory, R.I. (2015). MicroRNA biogenesis pathways in cancer. *Nat. Rev. Cancer* 15, 321–333.
11. Bottani, M., Banfi, G., and Lombardi, G. (2019). Circulating miRNAs as diagnostic and prognostic biomarkers in common solid tumors: focus on lung, breast, prostate cancers, and osteosarcoma. *J. Clin. Med.* 8, 1661.
12. Wang, X., He, Y., Mackowiak, B., and Gao, B. (2021). MicroRNAs as regulators, biomarkers and therapeutic targets in liver diseases. *Gut* 70, 784–795.
13. Shigeyasu, K., Toden, S., Zumwalt, T.J., Okugawa, Y., and Goel, A. (2017). Emerging role of microRNAs as liquid biopsy biomarkers in gastrointestinal cancers. *Clin. Cancer Res.* 23, 2391–2399.
14. Bahadori, M., Dabiri, S., Azizi, M.H., and Bahadori, N. (2021). Emerging Role of microRNAs as Liquid Biopsy Biomarkers in Lung Cancer: A Review. *Acta Med. Iran.*, 239–248.
15. Lin, X.-J., Chong, Y., Guo, Z.-W., Xie, C., Yang, X.-J., Zhang, Q., Li, S.-P., Xiong, Y., Yuan, Y., Min, J., et al. (2015). A serum microRNA classifier for early detection of hepatocellular carcinoma: a multicentre, retrospective, longitudinal biomarker identification study with a nested case-control study. *Lancet Oncol.* 16, 804–815.
16. Nishiwada, S., Cui, Y., Sho, M., Jun, E., Akahori, T., Nakamura, K., Sonohara, F., Yamada, S., Fujii, T., Han, I.W., et al. (2022). Transcriptomic profiling identifies an exosomal microRNA signature for predicting recurrence following surgery in patients with pancreatic ductal adenocarcinoma. *Ann. Surg.* 276, e876–e885.
17. Caby, M.-P., Lankar, D., Vincendeau-Scherrer, C., Raposo, G., and Bonnerot, C. (2005). Exosomal-like vesicles are present in human blood plasma. *Int. Immunol.* 17, 879–887.
18. Street, J.M., Barran, P.E., Mackay, C.L., Weidt, S., Balmforth, C., Walsh, T.S., Chalmers, R.T.A., Webb, D.J., and Dear, J.W. (2012). Identification and proteomic profiling of exosomes in human cerebrospinal fluid. *J. Transl. Med.* 10, 5–7.
19. Becker, A., Thakur, B.K., Weiss, J.M., Kim, H.S., Peinado, H., and Lyden, D. (2016). Extracellular vesicles in cancer: cell-to-cell mediators of metastasis. *Cancer Cell* 30, 836–848.
20. Hoshino, A., Costa-Silva, B., Shen, T.-L., Rodrigues, G., Hashimoto, A., Tesic Mark, M., Molina, H., Kohsaka, S., Di Giannatale, A., Ceder, S., et al. (2015). Tumour exosome integrins determine organotropic metastasis. *Nature* 527, 329–335.

21. Zhu, L., Sun, H.-T., Wang, S., Huang, S.-L., Zheng, Y., Wang, C.-Q., Hu, B.-Y., Qin, W., Zou, T.-T., Fu, Y., et al. (2020). Isolation and characterization of exosomes for cancer research. *J. Hematol. Oncol.* **13**, 152–224.
22. Nahand, J.S., Vandchali, N.R., Darabi, H., Doroudian, M., Banafshe, H.R., Moghoofei, M., Babaei, F., Salmanejad, A., and Mirzaei, H. (2020). Exosomal microRNAs: novel players in cervical cancer. *Epigenomics* **12**, 1651–1660.
23. Yu, W., Hurley, J., Roberts, D., Chakraborty, S.K., Enderle, D., Noerholm, M., Breakefield, X.O., and Skog, J.K. (2021). Exosome-based liquid biopsies in cancer: opportunities and challenges. *Ann. Oncol.* **32**, 466–477.
24. Yu, D., Li, Y., Wang, M., Gu, J., Xu, W., Cai, H., Fang, X., and Zhang, X. (2022). Exosomes as a new frontier of cancer liquid biopsy. *Mol. Cancer* **21**, 56.
25. Ye, Q., Ling, S., Zheng, S., and Xu, X. (2019). Liquid biopsy in hepatocellular carcinoma: circulating tumor cells and circulating tumor DNA. *Mol. Cancer* **18**, 114–213.
26. Vaidyanathan, R., Soon, R.H., Zhang, P., Jiang, K., and Lim, C.T. (2019). Cancer diagnosis: from tumor to liquid biopsy and beyond. *Lab Chip* **19**, 11–34.
27. Li, P., Kaslan, M., Lee, S.H., Yao, J., and Gao, Z. (2017). Progress in exosome isolation techniques. *Theranostics* **7**, 789–804.
28. He, C., Zheng, S., Luo, Y., and Wang, B. (2018). Exosome theranostics: biology and translational medicine. *Theranostics* **8**, 237–255.
29. Wang, W., Luo, J., and Wang, S. (2018). Recent progress in isolation and detection of extracellular vesicles for cancer diagnostics. *Adv. Healthcare Mater.* **7**, 1800484.
30. Patel, G.K., Khan, M.A., Zubair, H., Srivastava, S.K., Khushman, M., Singh, S., and Singh, A.P. (2019). Comparative analysis of exosome isolation methods using culture supernatant for optimum yield, purity and downstream applications. *Sci. Rep.* **9**, 5335.
31. Dobhal, G., Datta, A., Ayupova, D., Teesdale-Spittle, P., and Goreham, R.V. (2020). Isolation, characterisation and detection of breath-derived extracellular vesicles. *Sci. Rep.* **10**, 17381.
32. Shao, H., Chung, J., and Issadore, D. (2016). Diagnostic technologies for circulating tumour cells and exosomes. *Biosci. Rep.* **36**, e00292.
33. Chen, J., Li, P., Zhang, T., Xu, Z., Huang, X., Wang, R., and Du, L. (2021). Review on strategies and technologies for exosome isolation and purification. *Front. Bioeng. Biotechnol.* **9**, 811971.
34. Pinho, S.S., and Reis, C.A. (2015). Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* **15**, 540–555.
35. Noda, K., Miyoshi, E., Uozumi, N., Yanagidani, S., Ikeda, Y., Gao, C., Suzuki, K., Yoshihara, H., Yoshikawa, K., Kawano, K., et al. (1998). Gene expression of  $\alpha$ 1-6 fucosyltransferase in human hepatoma tissues: A possible implication for increased fucosylation of  $\alpha$ -fetoprotein. *Hepatology* **28**, 944–952.
36. Williams, C., Royo, F., Aizpurua-Olaizola, O., Pazos, R., Boons, G.-J., Reichardt, N.-C., and Falcon-Perez, J.M. (2018). Glycosylation of extracellular vesicles: current knowledge, tools and clinical perspectives. *J. Extracell. Vesicles* **7**, 1442985.
37. Descotes, J.-L. (2019). Diagnosis of prostate cancer. *Asian J. Urol.* **6**, 129–136.
38. Li, D., Mallory, T., and Satomura, S. (2001). AFP-L3: a new generation of tumor marker for hepatocellular carcinoma. *Clin. Chim. Acta* **313**, 15–19.
39. Escrivente, C., Keller, S., Altevogt, P., and Costa, J. (2011). Interaction and uptake of exosomes by ovarian cancer cells. *BMC Cancer* **11**, 108–110.
40. Szabo, G., and Bala, S. (2013). MicroRNAs in liver disease. *Nat. Rev. Gastroenterol. Hepatol.* **10**, 542–552.
41. Hayes, J., Peruzzi, P.P., and Lawler, S. (2014). MicroRNAs in cancer: biomarkers, functions and therapy. *Trends Mol. Med.* **20**, 460–469.
42. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. j.* **17**, 10–12.
43. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359.
44. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550–621.
45. Kolde, R., and Kolde, M.R. (2015). Package ‘pheatmap’. *R package* **1**, 790.
46. Wickham, H. (2011). ggplot2. *WIREs Computational Stats.* **3**, 180–185.
47. Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D., and Ripley, M.B. (2013). Package ‘mass’. *Cran r* **538**, 113–120.
48. Pedersen, T.L., Pedersen, M., LazyData, T., Rcpp, I., and Rcpp, L. (2017). Package ‘ggraph’. Retrieved January 1, 2018.
49. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS A J. Integr. Biol.* **16**, 284–287.
50. Gomez, C., Goonenko, A.V., and Soulikova, J.N. (2023). Constructing UpSet plot for survey data with weights using SAS and R software. *Commun. Stat. Simulat. Comput.* **52**, 2320–2326.
51. Li, J.-H., Liu, S., Zhou, H., Qu, L.-H., and Yang, J.-H. (2014). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **42**, D92–D97.
52. van Rossum, G. Guido van Rossum. Python (programming language) 1 CPython 13 Python Software Foundation **15**, 16..
53. Kramer, O., and Kramer, O. (2016). Scikit-learn. *Machine Learning for Evolution Strategies*, pp. 45–53.
54. Tosi, S. (2009). Matplotlib for Python Developers (Packt Publishing Ltd)).
55. Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., and Wingett, S. (2020). FastQC: A Quality Control Tool for High Throughput Sequence Data. [Online]. 2010.
56. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE                                                      | SOURCE                                                                                                                                                                            | IDENTIFIER                      |
|--------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------|
| <b>Antibodies</b>                                                        |                                                                                                                                                                                   |                                 |
| Mouse monoclonal Anti-CD81 antibody (M38)                                | Abcam                                                                                                                                                                             | Cat# ab79559; RRID:AB_1603682   |
| Mouse monoclonal anti-PDC6I antibody (3A9)                               | Abcam                                                                                                                                                                             | Cat# ab117600; RRID:AB_10899268 |
| Mouse monoclonal Anti-TSG101 antibody (4A10)                             | Abcam                                                                                                                                                                             | Cat# ab83; RRID:AB_306450       |
| Rabbit recombinant Anti-calnexin antibody (EPR3633)                      | Abcam                                                                                                                                                                             | Cat# ab133615; RRID:AB_2864299  |
| Mouse monoclonal anti-CD81 antibody (TS81)                               | Abcam                                                                                                                                                                             | Cat# ab59477; RRID:AB_943630    |
| Mouse monoclonal anti-CD63 antibody (H5C6)                               | Thermo Fisher Scientific                                                                                                                                                          | Cat#12-0639-42; RRID:AB_2572565 |
| Mouse monoclonal anti-CD9 antibody (HI9a)                                | BioLegend                                                                                                                                                                         | Cat# 312104; RRID:AB_2075894    |
| Mouse monoclonal anti-IgG1 antibody (MOPC 21)                            | Abcam                                                                                                                                                                             | Cat#ab18443; RRID:AB_2736846    |
| <b>Biological samples</b>                                                |                                                                                                                                                                                   |                                 |
| Human serum                                                              | the First Medical Center of Chinese PLA General Hospital; the Fifth Medical Center of Chinese PLA General Hospital; Capital Medical University Affiliated Beijing Ditan Hospital. | N/A                             |
| <b>Chemicals, peptides, and recombinant proteins</b>                     |                                                                                                                                                                                   |                                 |
| Penicillin/Streptomycin/Amphotericin B Solution                          | Sangon Biotech                                                                                                                                                                    | Cat #B540733-0010               |
| DMEM media                                                               | Thermo Fisher Scientific                                                                                                                                                          | Cat #C11995500BT                |
| Exosome-depleted Fetal Bovine Serum                                      | Shanghai VivaCell Biosciences                                                                                                                                                     | Cat #C3801-0050                 |
| PNGase F                                                                 | Yeasen Biotechnology                                                                                                                                                              | Cat# 20411ES01                  |
| Acryl/Bis 30% Solution (19:1)                                            | Sangon Biotech                                                                                                                                                                    | Cat# B546016-0500               |
| Blue Plus® IV Protein Marker (10–180 kDa)                                | TransGen Biotech                                                                                                                                                                  | Cat# DM131-01                   |
| PBS (pH7.4)                                                              | Thermo Fisher                                                                                                                                                                     | Cat# 10010023                   |
| DiD Perchlorate (DiIC18(5)) cell membrane red fluorescent                | Shanghai Xuanke Biological Technology                                                                                                                                             | Cat# 127274-91-3                |
| <b>Critical commercial assays</b>                                        |                                                                                                                                                                                   |                                 |
| Extracellular vesicles isolation kit (GlyExo-Capture method)             | Beijing Hotgen Biotech                                                                                                                                                            | Cat# EVs20200092                |
| $\alpha$ -fetoprotein test (Chemiluminescence Immunoassay)               | Beijing Hotgen Biotech                                                                                                                                                            | NA                              |
| Des- $\gamma$ -carboxy prothrombin test (Chemiluminescence Immunoassay)  | Beijing Hotgen Biotech                                                                                                                                                            | NA                              |
| NEBNext1 Small RNA Library Prep Set for Illumina1 (Multiplex Compatible) | NEB                                                                                                                                                                               | Cat# E7330                      |
| NEBNext Ultra II RNA Library reagent kit                                 | NEB                                                                                                                                                                               | Cat# E7765                      |
| miRcute plus miRNA cDNA first strand synthesis kit                       | Tiagen Biotech                                                                                                                                                                    | Cat# KR211                      |
| miRcute plus miRNA qPCR kit (SYBR Green)                                 | Tiagen Biotech                                                                                                                                                                    | Cat# FP411-02                   |
| miRNeasy Mini Kit                                                        | Qiagen                                                                                                                                                                            | Cat# 217004                     |
| QIAzol Lysis Reagent                                                     | Qiagen                                                                                                                                                                            | Cat# 79306                      |
| Qubit protein assay kit                                                  | Thermo Fisher Scientific                                                                                                                                                          | Cat# Q33211                     |
| BeyoECL Plus (ECL like Western reagent)                                  | Beyotime biotechnology                                                                                                                                                            | Cat# P0018S                     |
| <b>Deposited data</b>                                                    |                                                                                                                                                                                   |                                 |
| Raw RNA-seq data                                                         | This paper                                                                                                                                                                        | GEO: GSE266634                  |
| <b>Experimental models: Cell lines</b>                                   |                                                                                                                                                                                   |                                 |
| HepG2                                                                    | China Center for Type Culture Collection                                                                                                                                          | Cat# CRL10741                   |
| HCT116                                                                   | China Center for Type Culture Collection                                                                                                                                          | Cat# CCL-247                    |

(Continued on next page)

### Continued

| REAGENT or RESOURCE                           | SOURCE                                   | IDENTIFIER                                                                                                                                      |
|-----------------------------------------------|------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| Hela                                          | China Center for Type Culture Collection | Cat# CCL-2                                                                                                                                      |
| MIHA                                          | Beijing Crisprbio Biotechnology          | Cat# CE18544                                                                                                                                    |
| <b>Oligonucleotides</b>                       |                                          |                                                                                                                                                 |
| hsa-let-7a_fwd: GCGCGTGAGGTAGTAGTTGTATAGT     | This paper                               | N/A                                                                                                                                             |
| hsa-miR-21_fwdGCGCTAGCTTATCAGACTGATGTT        | This paper                               | N/A                                                                                                                                             |
| hsa-miR-125a_fwdGCGCTCCCTGAGACCCTTT           | This paper                               | N/A                                                                                                                                             |
| hsa-miR-200a_fwd<br>AACCACTTAACACTGTCTGGTAACG | This paper                               | N/A                                                                                                                                             |
| hsa-miR-150_fwdTCTCCCAACCCTTGACCACTG          | This paper                               | N/A                                                                                                                                             |
| <b>Software and algorithms</b>                |                                          |                                                                                                                                                 |
| Cutadapt 2.7                                  | Martin et al. <sup>42</sup>              | <a href="https://cutadapt.readthedocs.io/en/stable/">https://cutadapt.readthedocs.io/en/stable/</a>                                             |
| Bowtie2 2.4.5                                 | Langmead et al. <sup>43</sup>            | <a href="https://sourceforge.net/projects/bowtie-bio/files/bowtie2/2.4.5/">https://sourceforge.net/projects/bowtie-bio/files/bowtie2/2.4.5/</a> |
| R 4.2.3                                       | The R core                               | <a href="https://www.r-project.org/">https://www.r-project.org/</a>                                                                             |
| DESeq2 1.22.2                                 | Love et al. <sup>44</sup>                | <a href="https://bioconductor.org/packages/release/bioc/html/DESeq2.html">https://bioconductor.org/packages/release/bioc/html/DESeq2.html</a>   |
| R package Pheatmap 1.0.12                     | Kolde et al. <sup>45</sup>               | CRAN                                                                                                                                            |
| R package ggplot2 3.4.4                       | Wickham et al. <sup>46</sup>             | CRAN                                                                                                                                            |
| R package MASS 7.3–60                         | Ripley et al. <sup>47</sup>              | CRAN                                                                                                                                            |
| R package ggraph 2.1.0                        | Pederson et al. <sup>48</sup>            | CRAN                                                                                                                                            |
| R package clusterProfiler 4.6.2               | Yu et al. <sup>49</sup>                  | CRAN                                                                                                                                            |
| R package upsetplot 0.9.0                     | Gomez et al. <sup>50</sup>               | CRAN                                                                                                                                            |
| starBase 2.0                                  | Li jun-hao et al. <sup>51</sup>          | <a href="http://starbase.sysu.edu.cn/">http://starbase.sysu.edu.cn/</a>                                                                         |
| Python 3.11.3                                 | Guido van Rossum <sup>52</sup>           | <a href="https://www.python.org">https://www.python.org</a>                                                                                     |
| Python package sklearn 1.2.2                  | Kramer et al. <sup>53</sup>              | <a href="https://scikit-learn.org/">https://scikit-learn.org/</a>                                                                               |
| Python package matplotlib                     | Tosi et al. <sup>54</sup>                | <a href="https://pypi.org/project/matplotlib/">https://pypi.org/project/matplotlib/</a>                                                         |
| Model code                                    | This paper                               | <a href="https://doi.org/10.5281/zenodo.13182532">https://doi.org/10.5281/zenodo.13182532</a>                                                   |

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

### Human subjects and sample preparation

This multicenter study was conducted across the Fifth Medical Center of Chinese PLA General Hospital (cohort 1), the First Medical Center of Chinese PLA General Hospital (cohort 2), and Capital Medical University Affiliated Beijing Ditan Hospital (cohort 3). These cohorts were used for discovery, training, and validation purposes. The study establishes rigorous inclusion and exclusion criteria for participant selection. Participants must meet one of the following inclusion conditions: newly diagnosed with primary hepatocellular carcinoma (HCC) at various stages (BCLC stage 0, A, B, C, D); diagnosed with other gastrointestinal or non-gastrointestinal tumors such as gastric cancer, colorectal cancer, lung cancer, or breast cancer; diagnosed with benign liver diseases such as viral hepatitis or cirrhosis; or healthy individuals without a history of tumors, and with no abnormalities found on blood biochemistry and imaging examinations; Additionally, participants must provide comprehensive case information, including outpatient or inpatient number, sample number, age, gender, and clinical diagnosis. They must be at least 18 years old, with no gender restrictions, and voluntarily participate by providing peripheral blood samples. Exclusion criteria include participants who have received immunotherapy or targeted therapy within the past 6 months, those with incomplete case information, and pregnant or lactating women. The study allows for termination if participants withdraw voluntarily at any stage. Cases will be excluded if participants are found not meeting the inclusion criteria, meeting the exclusion criteria, lacking essential clinical or imaging information, or if there are failures in extracellular vesicle extraction or RNA quality control, resulting in an insufficient sample size to complete the experiment. A total of 1046 samples were successfully collected, including 182 cases of HCC, 173 cases of non-hepatic malignancies, 245 from healthy controls, and 346 cases from individuals with benign liver disease including benign liver tumors, cirrhosis, and hepatitis (Figure S12). These samples were separated into three cohorts, shown in Table S1. The discovery cohort used for the initial biomarker screening, consisting of 88 HCC patients, 130 individuals with benign liver disease, and 49 healthy controls, was enrolled at the Fifth Medical Center of Chinese PLA General Hospital between November 2020 and May 2022. Next-generation deep sequencing (NGS) was performed to reveal

EVs-miRNA profiles from serum samples. Subsequently, the expression levels of candidate EVs-miRNA biomarkers were assessed using the qRT-PCR assay in an additional 779 samples for clinical biomarker identification and model performance confirmation. These samples were enrolled at the Fifth Medical Center of Chinese PLA General Hospital, the First Medical Center of Chinese PLA General Hospital, and Capital Medical University Affiliated Beijing Ditan Hospital. The training cohort, composed of 134 samples collected at the Fifth Medical Center of the Chinese PLA General Hospital between August 2021 and February 2022, was utilized to establish the model. Three validation cohorts, encompassing 645 samples, were used to assess the model's performance (Figure S12). The in-center validation cohort enrolled 202 participants from November 2021, to May 2022. The multi-center validation cohort was composed of 270 participants from the First Medical Center of Chinese PLA General Hospital ( $n = 151$ ) enrolled from April 2022 to May 2022, and Capital Medical University Affiliated Beijing Ditan Hospital ( $n = 119$ ) enrolled from March 2022 to September 2022. The cohort for other tumor species included 173 participants from Capital Medical University Affiliated Beijing Ditan Hospital, enrolled from March 2022 to May 2024. Demographic and clinical data (i.e., age, gender, BCLC stage, Etiology, AFP, and DCP) were obtained at the time of enrollment (Table S1).

This study was approved by the Ethics Committee of the Fifth Medical Center of Chinese PLA General Hospital (2016-004-D) and received a waiver of informed consent due to its minimal risk to participants and retrospective nature. The authors are responsible for all aspects of the work and ensure that any concerns regarding the integrity and accuracy of the research are thoroughly addressed. The study adhered to ethical standards and regulations to maintain participant confidentiality and privacy.

## METHOD DETAILS

### Sera sample collection

Blood was collected via venipuncture using a 21-gauge needle and immediately transferred into 5 mL vacuum tubes containing a separator gel (Improve Medical Instruments Co., Ltd). The harvested samples were then centrifuged at  $1800\times g$  for 10 min at room temperature within 1 h of collection. Subsequently, the resulting serum underwent additional centrifugation at  $3000\times g$  for 10 min at  $4^{\circ}\text{C}$  to remove any cellular debris. The obtained sera were divided into aliquots and stored at  $-80^{\circ}\text{C}$  for further analysis.

### EV isolation and characterization

EVs were isolated from the serum samples using the Extracellular vesicles isolation kit (GlyExo-Capture method) (Beijing Hotgen Biotech Co., Ltd., Beijing, China) following the standardized protocol. This approach utilized lectin-immobilized magnetic beads to capture EVs, leveraging the strong affinity between lectin and fucosylated proteins on EV surface. This approach enhanced the enrichment of EVs secreted by tumor cells, particularly abundant in membrane proteins glycosylated with fucose. Characterization and identification of EVs extracted using GlyExo-Capture were performed and contrasted with EVs obtained via ultra-centrifugation (UC). Transmission Electron Microscopy (TEM) was performed using an FEI Tecnai Spirit microscope (FEI, Eindhoven, The Netherlands) at 120 kV, with electron micrographs captured using a Gatan UltraScan 1000 charge-coupled device camera (Pleasanton, USA). Nanoparticle tracking analysis (NTA) was performed using a Malvern NanoSight NS300 instrument (Malvern Instruments, Ltd.) to characterize particle concentration and size distribution, following the manufacturer's instructions. The protein concentration of EVs was analyzed using a Qubit Protein Assay (Thermo Fisher Scientific, Inc., Waltham, USA) and a Qubit 3.0 fluorometer (Thermo Fisher Scientific, Inc., Waltham, USA). Western blot analysis was performed on 25  $\mu\text{g}$  of EVs using antibodies M38 (1:400 dilution, Abcam, Cambridge, USA), 3A9 (1:400 dilution, Abcam, Cambridge, USA), 4A10 (1:400 dilution, Abcam, Cambridge, USA), and EPR3633 (1:2000 dilution, Abcam, Cambridge, USA) to detect CD81, Alix, TSG101, and calnexin expression, respectively. Additionally, the Exoview R100 instrument (NanoView Biosciences) was employed to evaluate CD81, CD63, CD9, and IgG1 expression on EVs, employing TS81, H5C6, HI9a, and MOPC 21 antibodies (Abcam, Cambridge, USA), respectively. Polyacrylamide gel electrophoresis was performed by using Acryl/Bis 30% Solution (19:1) (Sangon Biotech Co., Ltd., Shanghai, China) and Blue Plus IV Protein Marker (10–180 kDa) (TransGen Biotech Co., Ltd., Beijing, China). BeyoECL Plus (ECL like Western reagent) (Beyotime biotechnology Co., Ltd., Shanghai, China) was used in Western blot analysis. Additionally, the Exoview R100 instrument (NanoView Biosciences) was employed to evaluate CD81, CD63, CD9, and IgG1 expression on EVs, employing TS81 (Abcam, Cambridge, USA), H5C6 (Thermo Fisher Scientific, Inc., Waltham, USA), HI9a (BioLegend, Beijing, China), and MOPC 21 (Abcam, Cambridge, USA) antibodies, respectively.

### Cellular uptake of EVs

Quantification of fucosylated and non-fucosylated EVs was performed using NTA. EVs treated with PNGase F (Yeast Biotechnology Co., Ltd., Shanghai, China) or without PNGase F were adjusted to the same concentration and labeled using a final concentration of 5  $\mu\text{M}$  DiD membrane dye (XK4019, Excitation/Emission (MeOH) = 644/663 nm, Shanghai Xuanke Biological Technology Co., Ltd.). After thorough mixing, the combination was incubated at  $37^{\circ}\text{C}$  for 40 min and filtered through a 0.2  $\mu\text{M}$  disposable filter membrane to remove excess DiD dye. Concurrently, EVs were incubated with PBS (Thermo Fisher Scientific, Inc., Waltham, USA), and the dye served as a negative control. HepG2, HCT-116, and HeLa from China Center for Type Culture and MIHA (Beijing Crisprbio Biotechnology Co., Ltd., Beijing, China) were assessed for mycoplasma contamination and cultured in DMEM media (Thermo Fisher Scientific, Inc., Waltham, USA) supplemented with 10% EVs-depleted fetal bovine serum (Shanghai VivaCell Biosciences Ltd., Shanghai, China) and 1% penicillin-streptomycin solution (Sangon Biotech Co., Ltd., Shanghai, China) at  $37^{\circ}\text{C}$  with 5%  $\text{CO}_2$ . The cells were

seeded into 24-well plates at a density of 100,000 cells per well in triplicate. Fucosylated and non-fucosylated EVs derived from HepG2 cells were included in each well, and the cells were incubated at 37°C for 3 h. After incubation, the cells were harvested, washed with PBS, and subjected to fluorescence detection using a BD Accuri C6 flow cytometer. A total of 5000 cells were obtained for each sample, with detection performed using the APC-A channel.

## EV-derived small RNA sequencing and analysis

### RNA isolation and sequencing

Total RNA was extracted from EVs using the miRNeasy Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's directions. RNA quality was assessed using a high-resolution cartridge of a Qsep100 fully automated nucleic acid analysis system (Bioptronic, Inc., LA, USA). The extracted miRNAs were reverse-transcribed into single-stranded cDNA libraries using the NEBNext Small RNA Library Prep Set kit (Multiplex Compatible) (NEB, Ipswich, USA) according to the manufacturer's instructions. Library size selection was conducted utilizing the E-Gel In Review SizeSelect II gel (Thermo Fisher Scientific, Inc., Waltham, USA) on an E-Gel Power Snap electrophoresis system (Thermo Fisher Scientific, Inc., Waltham, USA). Ultimately, groups of samples with equivalent concentrations were pooled together before being subjected to 75-bp single-read sequencing on an Illumina NextSeq 550 Sequencing System (Illumina, San Diego, CA, USA).

### NGS data analysis

Sequence quality was evaluated using FastQC version 0.11.9,<sup>55</sup> and primer adapters were removed using Cutadapt version 3.7.<sup>42</sup> Sequences were aligned to the human reference genome (HG38) using Bowtie 2.<sup>43</sup> Aligned data were annotated using HT-seq with the miRBase version 23 gff3 file, and miRNA reads were normalized to transcripts per million (TPM). Differential analysis was performed using the DESeq2 package in R version 4.2.3, with statistical significance identified using false discovery rate (FDR) adjustment. Differentially expressed miRNAs (DEMs) were found based on the fold changes exceeding 2 and adjusted *p*-values below 0.05. Clustering was performed using the pheatmap package (version 1.0.12) using the ward.D2 method. Linear Discriminant Analysis (LDA) was utilized on the TPM expression matrix with the MASS package (version 7.3) to visualize the differences between HCC and non-HCC groups.

### qRT-PCR

cDNA synthesis was performed using the miRcute Plus miRNA cDNA first-strand synthesis kit (Cat# KR211, Tiangen Biotech Co., Ltd., Beijing, China) following a standard protocol. Subsequently, quantitative real-time PCR (qRT-PCR) was performed using the miRcute Plus miRNA qPCR kit (SYBR Green) (Tiangen Biotech Co., Ltd., Beijing, China) on an ABI 7500 Real-Time PCR System (Applied Biosystems, Foster City, USA). The specific primers utilized for qRT-PCR were synthesized by Sangon Biotech (Shanghai, China).

### Measurement of AFP and DCP levels

Serum AFP and DCP was evaluated using a chemiluminescence immunoassay (CLIA) with magnetic microparticles. The analysis was carried out with the  $\alpha$ -fetoprotein test (Chemiluminescence Immunoassay) and Des- $\gamma$ -carboxy prothrombin test (Chemiluminescence Immunoassay) kits (Beijing Hotgen Biotech Co., Ltd, Beijing, China) by MQ60 AUTO (Beijing Hotgen Biotech Co., Ltd, Beijing, China), as per the manufacturer's directions.

## Construction and evaluation of the HCC prediction model

### Feature selection

We conducted recursive feature elimination on TPM data of differentially expressed miRNAs (DEMs) using the RFE module in the sklearn.feature\_selection package. The combination of miRNAs with the highest average correctness at five cross-validation was selected. The importance of the features was further evaluated using the RandomForest model to obtain the top ten miRNAs. Finally, five miRNAs were obtained using qRT-PCR validation and combined into three miRNA pairs for subsequent analysis.

### Bootstrapping approach for model assessment

To evaluate the robustness and performance of our models, we utilized a bootstrapping approach. The dataset was randomly partitioned into training and validation sets, and this process was repeated 1000 times. In each iteration, the optimal cutoff value was determined using the Youden Index from the training dataset. Predictions for each sample in validation dataset were recorded for every model generated. The final predicted class for each sample was determined by majority voting, where the class predicted most frequently across all models was assigned as the final prediction. This method allowed us to estimate performance metrics, including mean ROC and PR curves and their 95% confidence intervals, by averaging results from multiple iterations.

### NGS data-based model construction

We applied the LogisticRegression module from sklearn (parameters: C = 10, max\_iter = 1000, class\_weight = 'balanced') to assess the robustness of the 3 miRNA ratio signature. We then generated Receiver Operating Characteristic (ROC) and Precision-Recall curve (PRC), along with 95% confidence intervals, using matplotlib, and calculated the sensitivity and specificity.

### qRT-PCR data-based model construction

Logistic regression models (parameters: C = 10, max\_iter = 1000, class\_weight = 'balanced') were produced on the training dataset and assessed on the validation dataset. Evaluation techniques included ROC and PRC, confusion matrices, bar charts delineating

performance across distinct data subsets, and plots illustrating model score distributions. Additionally, box-and-whisker plots were generated for samples of non-hepatic malignancies, including gastric, lung, breast, and colorectal cancers.

### **Comparison of model performance**

We combined the training and validation cohorts, and employed Logistic Regression (parameters:  $C = 10$ ,  $\text{max\_iter} = 1000$ ,  $\text{class\_weight} = \text{'balanced'}$ ), RandomForestClassifier ( $\text{max\_depth} = 2$ ,  $\text{n\_estimators} = 100$ ,  $\text{class\_weight} = \text{'balanced'}$ ), SVC ( $\text{probability} = \text{True}$ ,  $\text{class\_weight} = \text{'balanced'}$ ,  $\text{kernel} = \text{'rbf'}$ ,  $C = 0.4$ ,  $\text{gamma} = 0.15$ ), and MLP ( $\text{hidden\_layer\_sizes} = (8,6)$ ,  $\text{max\_iter} = 1000$ ) Classifier to model the entire dataset. Subsequently, models integrating miRNAs, miRNA&AFP, miRNA&DCP, and miRNA&AFP&DCP were produced through random sampling and 1000 iterations. ROC and PRC, along with 95% confidence intervals, were plotted to contrast their performance with separate protein metrics AFP and DCP. Model cutoffs were established at optimal thresholds from the training set, while clinical thresholds were used for protein features. This analysis was also independently performed in each of the three cohorts.

### **Messenger RNA NGS sequencing and data analysis**

#### **RNA sample preparation**

RNA samples underwent initial quality assessment for purity and concentration using a Nanodrop2000 spectrophotometer (Thermo Fisher Scientific, Inc., Waltham, USA). RNA integrity was assessed using the Labchip GX Touch microfluidic capillary system (PerkinElmer, Waltham, MA, USA).

#### **mRNA library Construction**

Total RNA samples were employed to construct mRNA libraries using the NEBNext Ultra II RNA Library reagent kit (NEB, Ipswich, USA). Eukaryotic mRNA enrichment was attained with Oligo (dT) magnetic beads, followed by mRNA fragmentation. First-strand cDNA synthesis employed a six-base random primer, followed by second-strand cDNA synthesis including buffer, dNTPs, DNA Polymerase I, and RNase H. Purified double-stranded cDNA underwent end repair, A-tailing, and adapter ligation. Fragments of approximately 350 bp were size-selected, and the cDNA library was enriched via PCR amplification.

#### **Library quantification and pooling**

Library concentration was evaluated using a Qubit 3.0 fluorometer, and fragment size was assessed with the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). Precise quantification was performed using the Bio-RAD CFX 96 Real-Time PCR instrument (Bio-Rad, Hercules, CA, USA). Qualified libraries were combined based on effective concentration and desired downstream data volume for sequencing, generating 150 bp paired-end reads.

#### **Sequencing**

Sequencing was conducted using the Illumina HiSeq X Sequencing System (Illumina, San Diego, CA, USA). The flow cell was loaded with fluorescently labeled dNTPs, DNA polymerase, and primers for amplification. Sequencing cycles integrated fluorescently labeled dNTPs, emitting corresponding signals captured by the sequencing instrument. Computer software converted optical signals into sequencing peaks to identify the sequence information of the target fragment.

#### **Data preprocessing and analysis**

Raw fastq files underwent preprocessing to eliminate adapter sequences and low-quality reads using cutadapt. Processed fastq files underwent quality assessment using FastQC to assess Q30 scores and GC content. High-quality sequences were aligned to the reference genome utilizing HISAT2, and annotation-guided quantification was performed using featureCounts to produce the final counts matrix. FPKM values were derived for subsequent differential analysis using the DESeq2 package in R, with statistical significance characterized through false discovery rate (FDR) adjustment.

### **miRNA-mRNA regulatory network analysis**

To visualize the miRNA-mRNA regulatory network, we utilized the 'ggraph' package in R. Furthermore, using 'clusterProfiler' package in R, we conducted KEGG pathway enrichment analysis on the miRNA-targeted mRNAs within the network to elucidate the functional roles of the DEM.<sup>56</sup>

### **QUANTIFICATION AND STATISTICAL ANALYSIS**

We employed R version 4.2.3 to conduct both statistical computations and data visualization. To assess the statistical differences in efficiency between EV capture approaches using the BSA-beads and lectin-beads-based methods, Student's t-tests were employed. For the differential analysis of qRT-PCR data, the Wilcoxon signed-rank test was applied. All statistical tests adhered to a two-tailed approach unless specifically indicated otherwise specified. The enrichment analysis for the predicted targets of five miRNAs among differentially expressed genes, derived from messenger RNA sequencing (mRNA-Seq) data, was conducted using the hypergeometric distribution test implemented within the hyper function in R.