



OPEN

# Predictive three-biomarker panel in peripheral blood mononuclear cells for detecting hepatocellular carcinoma

Sara Fayazzadeh<sup>1</sup>, Mahsa Ghorbaninejad<sup>2</sup>, Amirhassan Rabbani<sup>3</sup>, Javad Zahiri<sup>1,4</sup> & Anna Meyfour<sup>2</sup>✉

Hepatocellular carcinoma (HCC) ranks among the most prevalent cancers and accounts for a significant proportion of cancer-associated deaths worldwide. This disease, marked by multifaceted etiology, often poses diagnostic challenges. Finding a reliable and non-invasive diagnostic method seems to be necessary. In this study, we analyzed the gene expression profiles of 20 HCC patients, 12 individuals with chronic hepatitis, and 15 healthy controls. Enrichment analysis revealed that platelet aggregation, secretory granule lumen, and G-protein-coupled purinergic nucleotide receptor activity were common biological processes, cellular components, and molecular function in HCC and chronic hepatitis B (CHB) compared to healthy controls, respectively. Furthermore, pathway analysis demonstrated that “estrogen response” was involved in the pathogenesis of HCC and CHB conditions, while, “apoptosis” and “coagulation” pathways were specific for HCC. Employing computational feature selection and logistic regression classification, we identified candidate genes pivotal for diagnostic panel development and evaluated the performance of these panels. Subsequent machine learning evaluations assessed these panels’ performance in an independent cohort. Remarkably, a 3-marker panel, comprising RANSE2, TNF- $\alpha$ , and MAP3K7, demonstrated the best performance in qRT-PCR-validated experimental data, achieving 98.4% accuracy and an area under the curve of 1. Our findings highlight this panel’s promising potential as a non-invasive approach not only for detecting HCC but also for distinguishing HCC from CHB patients.

Primary liver cancer, approximately 75% histology of which is hepatocellular carcinoma (HCC)<sup>1</sup>, is the sixth most frequently occurring cancer and the third most common cause of cancer mortality in both sexes worldwide<sup>2</sup>. Most HCC patients are diagnosed at late or advanced stages when the treatment options are limited and it is hard to cure if not untreatable. HCC is a complex multifactorial disease. The most important HCC risk factors mainly include cirrhosis, chronic infection with hepatitis B virus (HBV), hepatitis C virus (HCV), alcoholic liver disease, and nonalcoholic steatohepatitis (NASH). Additional risk factors that are also known to increase the chance of developing HCC include tobacco smoking, aflatoxin-contaminated food intake, diabetes, obesity, genetic factors, and heredity<sup>3,4</sup>.

HCC is associated with poor prognosis and low survival rates. Therefore, early detection of HCC is beneficial for prolonging patient survival<sup>5</sup>. To date, the gold standard for HCC diagnosis is liver biopsy, which is an invasive procedure as well as being time-consuming and expensive. This procedure also brings difficulties and complications to the patients such as tumor seeding, intra-abdominal bleeding, pain, discomfort, bacteremia, or even death<sup>6–8</sup>. Therefore, non-invasive diagnostic methods are preferred nowadays. Even though detection of liver cancer in high-risk individuals has been improved by some imaging-based methods, including ultrasonography (US), computed tomography (CT), and/or high-cost magnetic resonance imaging (MRI), these radiological imaging techniques have shown some limitations like high false-positive rates leading to unnecessary patient

<sup>1</sup>Bioinformatics and Computational Omics Lab (BioCOOL), Department of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran. <sup>2</sup>Basic and Molecular Epidemiology of Gastrointestinal Disorders Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran. <sup>3</sup>Department of Transplant and Hepatobiliary Surgery, Taleghani Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran. <sup>4</sup>Department of Neurosciences, University of California, San Diego, CA, USA. ✉email: a.meyfour@sbmu.ac.ir

anxiety and may need to be followed by biopsy as an invasive procedure<sup>9</sup>. Conventional serum markers, such as  $\alpha$ -feto-protein (AFP), which is widely used in the clinical setting, lack specificity and show limited ability for the detection of HCC<sup>10</sup>. Identifying new, reliable, non or minimally invasive biomarkers is important to improve the detection of HCC.

Peripheral mononuclear cells (PBMCs) including lymphocytes and monocytes are critical components in the host immune system. Previous studies demonstrated that alteration of PBMC genes was observed in various malignancies and cancers such as non-small cell lung cancer (NSCLC), pancreatic and breast cancers<sup>11–13</sup>. Therefore, the detection of these changes could serve as a potential non-invasive diagnostic method. The suitable biomarker should be characterized as widely used for screening purposes and diagnosing asymptomatic patients, having high sensitivity and specificity, low cost, and easy detection.

In the present study, we integrated datasets retrieved from gene expression omnibus (GEO) and analyzed the data to identify differentially expressed genes (DEGs) between HCC, hepatitis B patients, and healthy individuals. Combining multiple datasets and integrated analysis can help in overcoming the individual dataset's limitations, including batch effects and dataset specific biases, increasing the number of samples, generalizability, statistical power and reliability<sup>14</sup>.

Furthermore, enrichment and functional analyses as well as network construction were performed to find important signaling pathways and key players in HCC. Feature selection using multinomial LASSO was done to find important genes in the diagnosis of each condition. Then, multinomial logistic regression as a machine learning algorithm was used to investigate the ability of selected features to distinguish cases from controls. Gene panels with the best performance were introduced as HCC diagnostic biomarker panels and verified in a real-life patient cohort.

## Results

### Basic clinical characteristics and patients' information

In this study, the gene expression profiles of HCC patients, chronic hepatitis B (CHB) patients as positive controls, and healthy individuals of the GSE49515 and GSE58208 data series were integrated to form the discovery cohort (Supplementary Fig. S1 and Supplementary Table 1). This dataset was utilized to identify significantly altered molecular mechanisms in PBMCs of HCC patients, discover potential biomarker panels and develop a machine-learning approach that can predict the diagnosis of the disease. To examine the predictive power and reproducibility, we conducted an experimental evaluation of the panel using gene expression data from 39 HCC patients, 15 CHB patients and 24 healthy controls. The study workflow is depicted in Fig. 1. For participant characteristics in the validation cohort refer to Supplementary Table 2. HCC patients comprised 29 (74.4%) men and 10 (25.6%) women. Approximately 38% of these patients reported no history of smoking, while 31% were current smoker and another 31% were former smokers. Half of the patients were overweight, and 71.8% of patients reported no family history of liver disease. Additionally, 36% of all patients reported alcohol consumption. The average total bilirubin of HCC patients was 2.51 mg/dL, while the serum albumin level averaged at 3.63 g/dL. The mean values of aspartate transaminase, alanine transaminase, and alkaline phosphatase were 58.29, 57.29, and 320.57 IU/L, respectively. Additionally, the average alpha-fetoprotein level among HCC patients was 163.7 ng/mL.

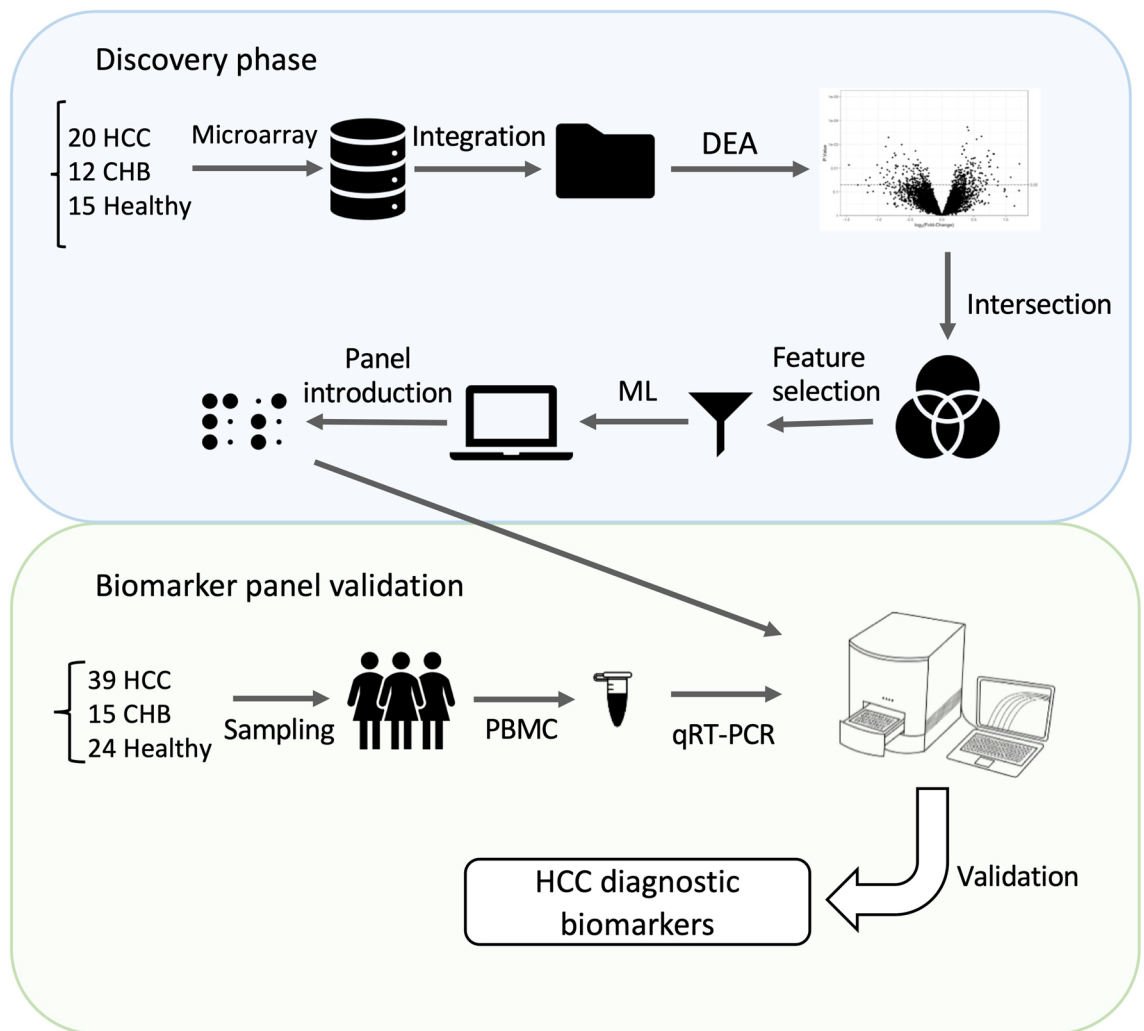
### PBMC transcriptome analysis of patients with HCC and CHB

In the present study first, we integrated samples of two datasets GSE58208 and GSE49515 by combat function of the *sva* package to form a unique dataset of 20 HCC, 12 CHB patients, and 15 normal controls called discovery cohort. Samples' bar plot and PCA were plotted before and after performing combat to assure the batch effect removal (Supplementary Fig. S1).

A total of 107 (99 upregulated and 8 downregulated) and 54 (46 upregulated and 8 downregulated) genes were identified under the threshold of absolute  $\log_2$ (fold change)  $\geq 1$  and adjusted  $p$  value  $< 0.05$  for HCC and CHB compared to normal conditions, respectively (Fig. 2a,b). Correlation heatmap was plotted to demonstrate the relationship between groups (Fig. 2c,d), it clearly illustrated the similarities within groups and distinction between different groups in both HCC vs healthy controls (Fig. 2c) and CHB vs healthy controls (Fig. 2d) and the conditions were greatly separated. Furthermore, we conducted hierarchical clustering and visualized the results through the heatmap. The expression patterns of 107 and 54 DEGs of HCC and CHB versus control were illustrated in Fig. 3a,b, respectively. These heatmaps demonstrated the upregulated and downregulated patterns of DEGs in each condition.

### Common and distinct gene signatures associated with HCC in PBMC

We next investigated the gene ontologies (GOs) of DEGs in HCC (Fig. 4a) and CHB (Fig. 4b) patients compared to healthy controls using Enrichr. Terms and processes with  $p$  value  $< 0.05$  were considered significant and the top 10 biological processes (BP), molecular functions (MF), and cellular components (CC) were visualized for each condition in comparison to normal controls. Platelet aggregation was the most enriched BP when comparing both HCC and CHB groups to healthy subjects. The secretory granule lumen was among enriched CC in both comparisons. Furthermore, G-protein-coupled purinergic nucleotide receptor activity and hemoglobin alpha binding were enriched MF in HCC and CHB comparisons with the healthy group, respectively. Pathway analysis demonstrated that "estrogen response" was involved in both HCC and CHB pathogenesis (Fig. 4c,d). Furthermore, enrichment analysis showed that DEGs in the HCC group were significantly members of "apoptosis" and "coagulation" signaling pathways (Fig. 4c), while "Interferon alpha response" was a substantially enriched pathway in CHB patients compared to normal controls (Fig. 4d).



**Figure 1.** The schematic overview of the study. In discovery phase, two datasets were retrieved from GEO, integrated and analyzed to identify predictive biomarkers of HCC using feature selection and machine learning (ML) algorithm. Additionally, in the validation phase, the predictive panel was experimentally confirmed on a real-life cohort.

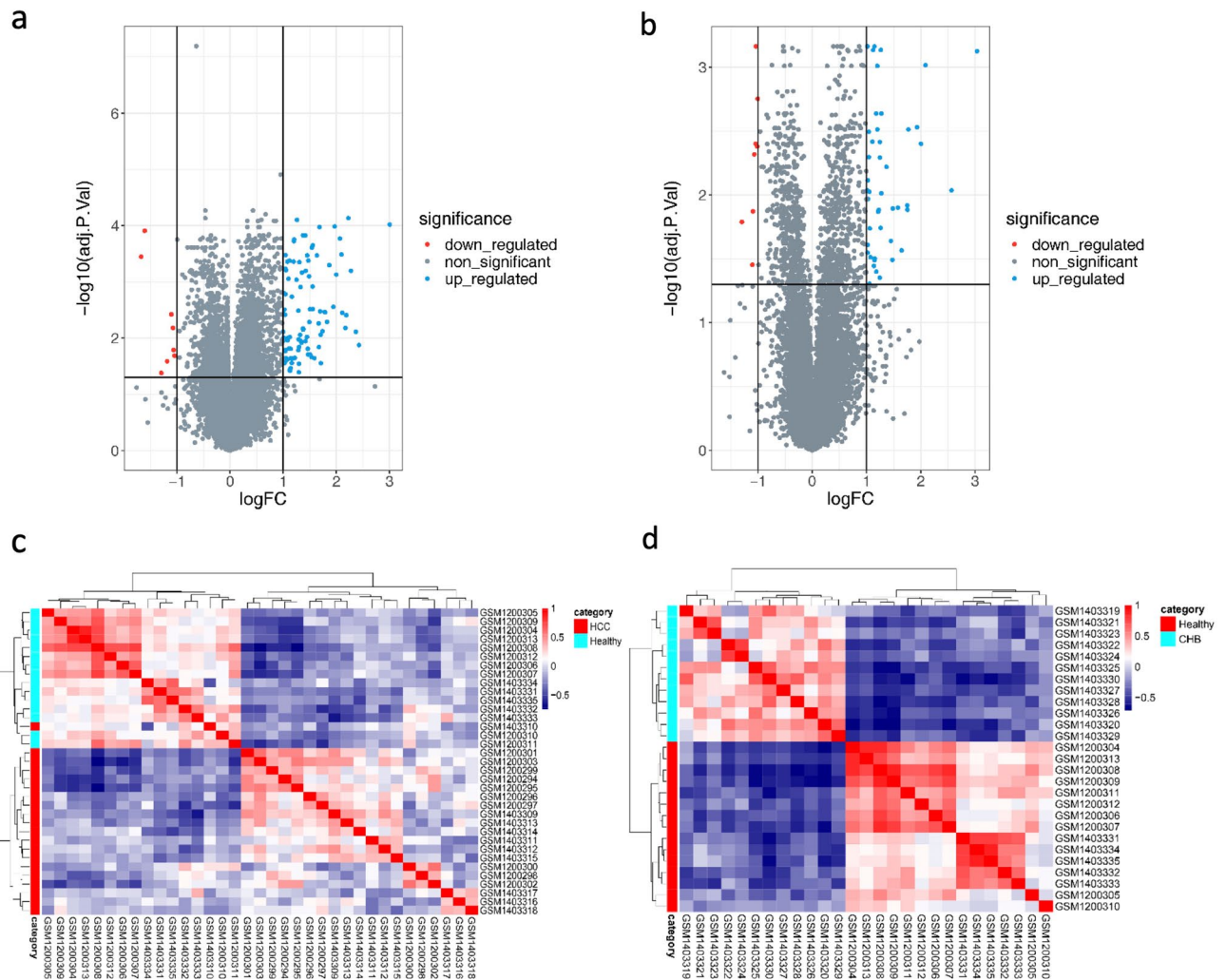
### The protein–protein interaction network of DEGs in HCC

Significant altered genes of HCC compared to the healthy group were applied to construct the PPI network using STRING. Visualization of the network was achieved using Cytoscape, highlighting functional interactions. The network comprised 96 nodes and 107 edges as shown in Fig. 4e. Centrality parameters, including betweenness, closeness, degree, eccentricity, radiality, and stress of the nodes, were computed with the CentiScape plugin. The top 10 nodes with the highest values for each centrality were identified. Subsequently, genes appearing in the top 10 across all centrality measures were recognized as hub genes. Key nodes, according to the CentiScape analysis, included tumor necrosis factor (TNF), interleukin 1 beta (IL-1 $\beta$ ), integrin subunit alpha 2b (ITGA2B), and purinergic receptor P2Y12 (P2RY12), as detailed in Supplementary Fig. S2.

### Discovery of candidate predictive biomarker panels in PBMCs of HCC patients

To investigate the intersection of DEGs, a Venn diagram was constructed using all significant up regulated and down regulated genes in HCC and CHB compared to healthy controls (Fig. 5a). Upregulated genes that did not have an overlap for each condition compared to healthy controls were then selected for the feature selection step. There were 73 unique genes that were only upregulated in HCC compared to healthy controls, while 22 genes were upregulated in CHB in comparison with healthy controls. Furthermore, there were two genes that not only upregulated in HCC compared to normal samples but also upregulated in HCC patients compared to CHB ones. These 97 genes were then fed to multinomial LASSO for feature selection in R. A list of 11, 7, and 5 genes with non-zero LASSO coefficient were reported in Supplementary Table 3, which that may have the potential to detect HCC, CHB, and healthy, respectively.

To assess the diagnostic performance of candidate genes in distinguishing the three specified groups, we utilized the top three genes with the highest absolute LASSO coefficient for each condition. These genes included



**Figure 2.** Distinct transcriptomes of HCC and CHB patients compared to healthy transcriptomes. Volcano plots show the number of statistically (adjusted  $p$  value  $< 0.05$ , absolute  $\log_2$ (fold change)  $\geq 1$ ) significant DEGs in PBMCs samples of (a) HCC and (b) CHB patients compared to Healthy controls. The volcano plots were created using R. Pearson correlation analysis along with the hierarchical clustering of (c) 107 DEGs in HCC vs. healthy and (d) 54 DEGs in CHB vs. healthy reveals within groups' cohesion and altered gene expression profiles of PBMCs in HCC development. Red represents the positive correlation and blue represents the negative correlation.

CCNA1, ANG, RNASE2, FADD, TNF- $\alpha$ , KCNJ2, ZNF324, DNAJB2, and MAP3K7, allowing for the creation of 3- to 7-biomarker panels. Then we used multinomial logistic regression, and built a model for each 3- to 7-marker panel that is evaluated by fivefold cross-validation. Sensitivity and specificity of the best panels for each condition were reported in Table 1. The best 3-marker panel, comprising RNASE2, TNF- $\alpha$ , and DNAJB2 genes, demonstrated an accuracy of 0.915. The accuracy increased with the addition of more markers: 0.936, 0.957, and 0.979 for the 4-, 5-, and 6-marker panels, respectively. It reached 1 for the 7-marker panel, which consists of RNASE2, TNF- $\alpha$ , KCNJ2, CCNA1, FADD, ZNF325, and MAP3K7 genes. The expression profiles of these genes, contributing to the construction of the best 3- to 7-marker panels, are illustrated in Fig. 5b–h. The results indicated significant upregulation of RNASE2, TNF- $\alpha$ , CCNA1, and MAP3K7 in HCC compared to both CHB and healthy controls. Additionally, DNAJB2 and KCNJ2 were found to be upregulated in CHB relative to both HCC and healthy controls. ZNF324, on the other hand, was upregulated in both HCC and CHB when compared to healthy controls.

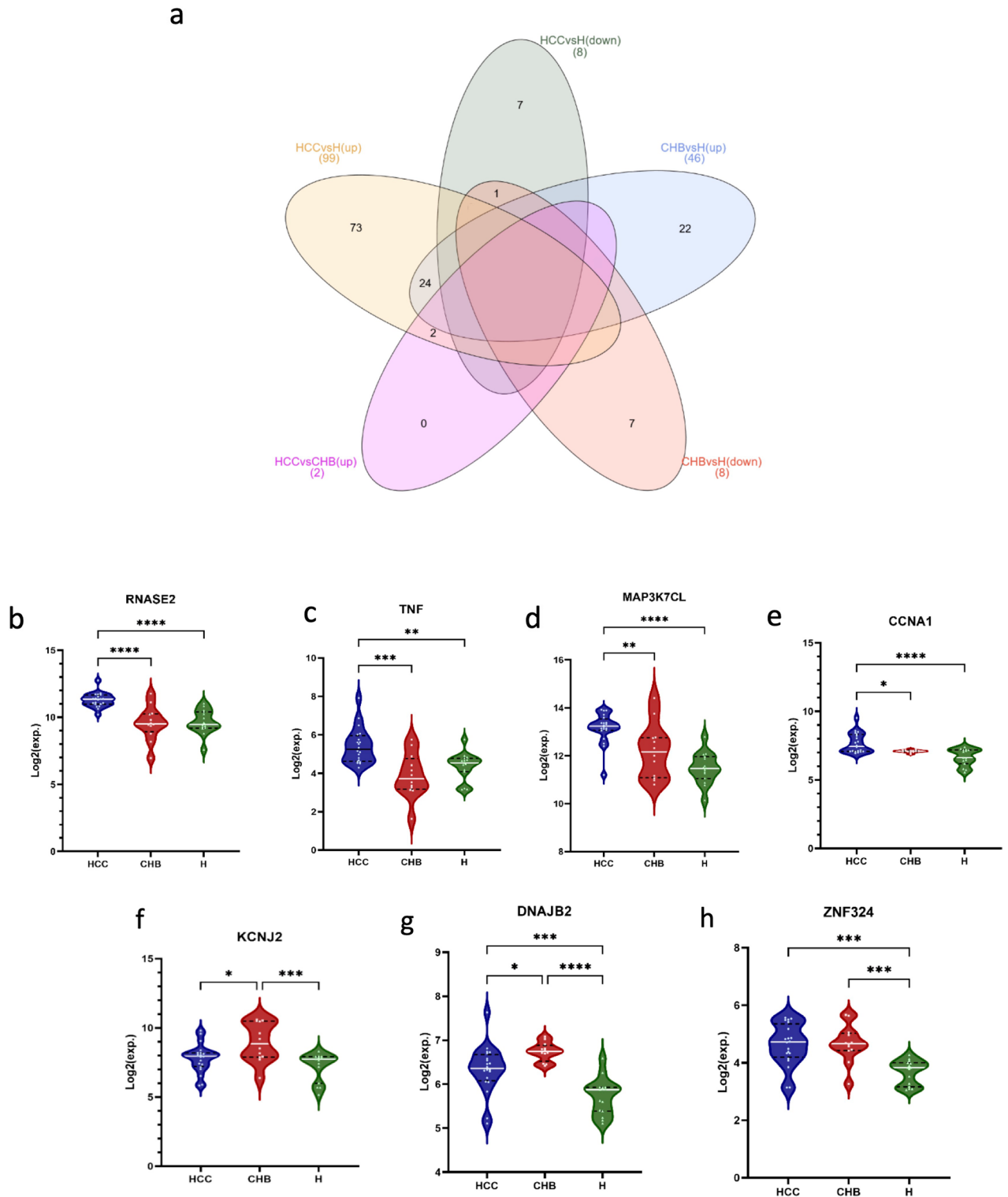
#### Experimental validation of the best *in-silico* predictive panels in the real-life HCC patient cohort

The expression of gene members in top models were then evaluated using qRT-PCR. Their expression levels were plotted using GraphPad Prism version 9. It was observed that RNASE2, TNF- $\alpha$ , MAP3K7, and CCNA1 were significantly differentially expressed in HCC patients compared to healthy controls. As expected, DNAJB2 and KCNJ2, showed no significant changes in HCC compared to healthy samples (Fig. 6a–f). Furthermore, RNASE2,









**Figure 5.** Identification of specific genes for HCC and CHB cases. (a) Venn diagram represents the intersection between upregulated and downregulated genes in HCC and CHB compared to healthy controls. It demonstrated that there are 73, and 22 unique upregulated genes for HCC and CHB compared to healthy controls, respectively. The violin plots represent the expression levels of (b) RNASE2, (c) TNF, (d) MAP3K7, (e) CCNA1, (f) KCNJ2, (g) DNAJB2, and (h) ZNF324. X-axis represents HCC (left), CHB (middle) and healthy control (right), Y-axis represents mRNA expression in log<sub>2</sub>.

Panel	No. features	HCC sen	CHB sen	Healthy sen	HCC spe	CHB spe	Healthy spe	Accuracy
RNASE2, TNF- $\alpha$ , DNAB2	3	0.95	0.833	0.933	0.963	0.943	0.969	0.915
RNASE2, TNF- $\alpha$ , MAP3K7, KCN2	4	0.95	0.917	0.933	0.926	1	0.969	0.936
RNASE2, TNF- $\alpha$ , MAP3K7, KCN2, DNAB2	5	0.95	0.917	1	0.963	0.971	1	0.957
RNASE2, TNF- $\alpha$ , KCN2, CCNA1, FADD, ZNF324	6	1	0.917	1	1	1	0.969	0.979
RNASE2, TNF- $\alpha$ , KCN2, CCNA1, FADD, ZNF325, MAP3K7	7	1	1	1	1	1	1	1

**Table 1.** The best 3- to 7-marker panels on microarray data (discovery cohort).

comprising RNASE2, TNF- $\alpha$ , and MAP3K7, which showed the best performance in detecting HCC from healthy controls, also exhibited acceptable performance in distinguishing HCC from CHB patients with an accuracy of 0.907 and AUC of 0.92 (Supplementary Table 5).

## Discussion

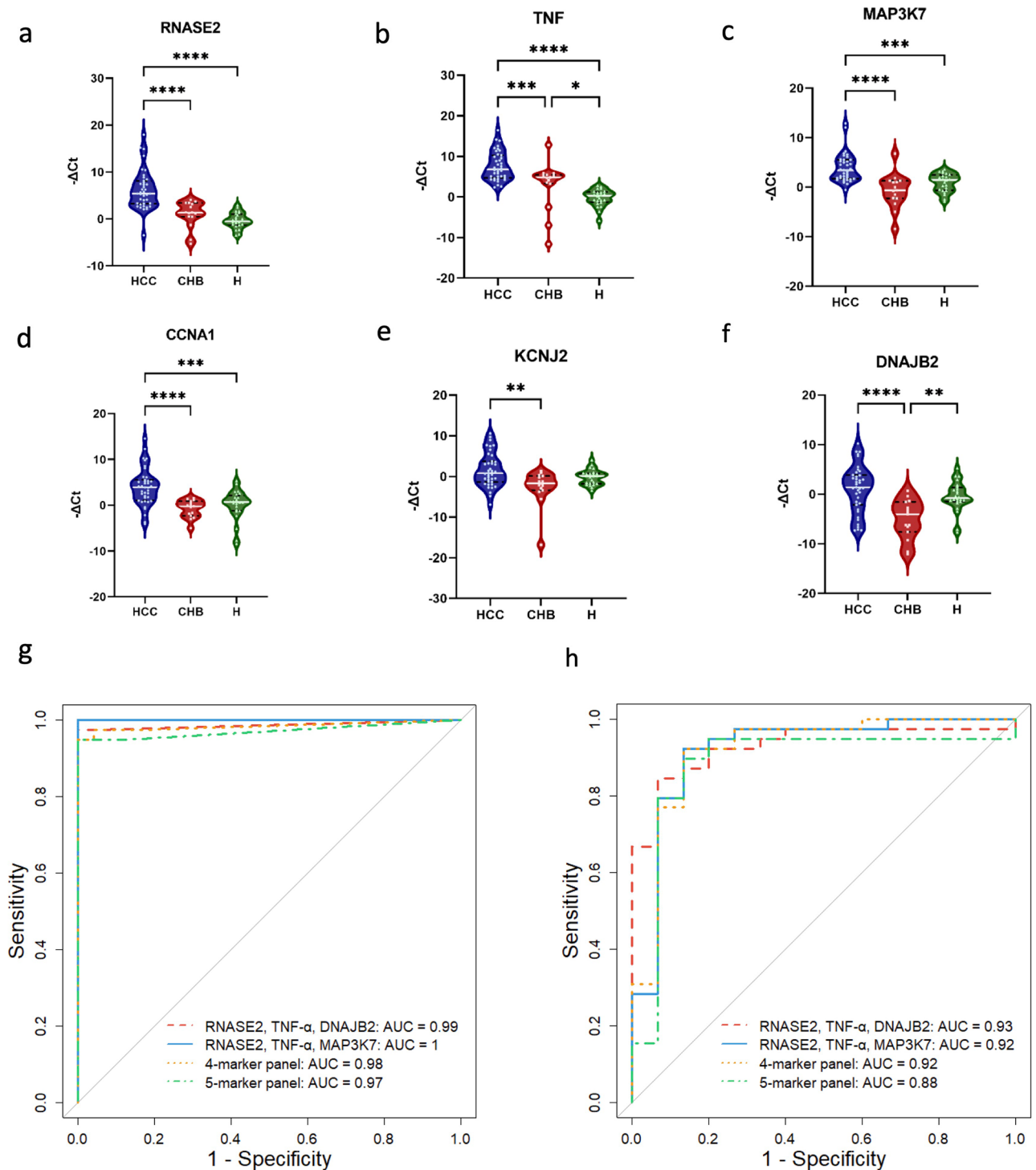
HCC is the most common type of primary liver cancer with a raising global burden of mortality. Being asymptomatic in the early stages delays timely diagnosis and leads to limited curative options in advanced stages<sup>15</sup>. PBMCs including monocytes, lymphocytes, and natural killer (NK) cells have a critical role in immune system functions. Studies showed that they have applications in the diagnosis, screening, and prognosis of diseases. Accumulating evidence has indicated that gene expression and methylation profiles of PBMCs alter in different types of disorders and malignancies<sup>16</sup>, such as HCC<sup>17</sup>. Alpha-fetoprotein (AFP) and imaging techniques are conventional methods for screening and early detection of HCC. However, the sensitivity of these methods even in combination is still suboptimal, ranging from 39 to 65% for AFP<sup>18</sup> and just 63% for the combination of ultrasound and AFP<sup>19</sup>, highlighting the need for more reliable and accurate methods to improve the prognosis and diagnosis of HCC patients. Furthermore, transcriptome analysis and identifying key players could help to understand the underlying mechanism of tumorigenesis of HCC<sup>20</sup>.

In this study, two different datasets, retrieved from the GEO database, were integrated to improve the sample size and reduce the study bias. The unified dataset which contained 20 HCC, 12 CHB patients, and 15 healthy controls was used to identify genes with significant alteration in their expression known as DEGs. Our findings showed that 107 and 54 genes were significantly expressed in HCC and CHB compared to healthy controls, respectively. Functional annotation analysis that was conducted using DEGs, revealed that platelet aggregation was the common enriched term among the BPs for both HCC and CHB conditions in comparison to healthy controls. Accumulating evidence suggests that platelets play a pivotal role in both inflammation and the progression of cancer<sup>21,22</sup>. Their contact with PBMCs modulated the immune response against viral infections<sup>23</sup>. Furthermore, it has been observed that platelets modulate T cell immunity against cancer, emphasizing their role in the immune response in the tumor microenvironment<sup>24</sup>. The interaction of platelet-leukocyte, as a predominant member of PBMCs, can lead to inflammatory immune reactions in different diseases<sup>25</sup>, including cancer<sup>26</sup>. The effector functions of innate leukocyte are modulated by platelets that are involved in immune responses but also promote thrombosis and metastasis<sup>26</sup>.

Platelets are small anucleate cellular fragments, that has functional organelles including mitochondria, Golgi apparatus, and endoplasmic reticulum (ER). Megakaryocytes generate platelets in bone marrow and release them into the blood stream<sup>27–29</sup>. In homeostasis and the initiation of coagulation, platelets play a crucial role by binding to injured vascular sites, forming aggregates, and subsequently arresting bleeding<sup>28,29</sup>. The role of platelets extends beyond clotting. Notably, they possess mRNA, which equips them with the capability to synthesize, express, and release proteins<sup>28,29</sup>. Upon activation, platelets initiate the release of various bioactive molecules, including multiple growth factors, serotonin, inflammatory cytokines, tumor necrosis factor alpha, chemokines, adrenaline, and more, via platelet granules. Bioactive molecules released from activated platelets can exhibit multifaceted effects, such as shielding cancer cells against immune surveillance, creating a stimulant environment for tumor growth, invasion and facilitating metastasis, or even exerting all of these actions concurrently<sup>22,29,30</sup>. While a reduction in platelet counts has been consistently reported in patients afflicted with chronic HBV infection and various other chronic liver diseases, it is noteworthy that an elevated platelet count has frequently been associated with HCC aggressiveness and tumor size, an increased propensity for tumor recurrence, and an elevated risk of metastatic dissemination<sup>30</sup>. Henceforth, conventional antiplatelet medications or platelet activation inhibitors may have the capacity to impede platelet aggregation, diminish hepatocyte damage arising from viral infection, and suppress the development of HCC<sup>30</sup>. Moreover, multiple BPs related to platelet function demonstrated significant enrichment in HCC compared to healthy controls. Key processes included platelet degranulation, which pertains to the platelet response to aggregating agents, and the regulation of megakaryocyte differentiation.

Other significant BPs in HCC against healthy controls were regulation of nitric oxide synthetic and metabolic processes. Nitric oxide (NO) serves as a transcellular signaling molecule, and its production is catalyzed by enzymes known as nitric oxide synthases (NOS). All isoforms of NOS have been identified in tumor cells, and they play a pivotal role in both promoting and suppressing cancer progression. NOS exhibits a dual role in cancer pathogenesis, correlating with tumor grade, proliferation rate, and the expression of cancer-development signaling molecules, such as the estrogen receptor. As a free radical gas molecule at room and body temperature,





**Figure 6.** Real-life cohort evaluation of predictive panels and ROC curves for the 3- to 5-marker panels. Experimental validation of the predictive panels in the real-life cohort. Relative expression of candidate markers (a) RNASE2, (b) TNF, (c) MAP3K7, (d) CCNA1, (e) KCNJ2, and (f) DNAJB2 in PBMC samples of patients with HCC, CHB and healthy controls. Green violin plots represent the expression levels of genes in normal patients ( $n = 24$ ), blue and red violin plots represent the expression levels of genes in patients with HCC ( $n = 39$ ) and CHB ( $n = 15$ ), respectively. Relative expression of genes represents as  $-\Delta Ct$  on y-axis. (g) The ROC curve depicts the performance of the three-marker panel comprising RNASE2, TNF- $\alpha$ , and DNAJB2 (red line), as well as the four-marker (yellow line) and five-marker (green line) panels, alongside the best introduced three-marker panel containing RNASE2, TNF- $\alpha$ , and MAP3K7 (blue line) for distinguishing HCC from healthy controls. (h) The ROC curve illustrates the ability to distinguish HCC from CHB, showing the performance of the three-marker panel containing RNASE2, TNF- $\alpha$ , and DNAJB2 (red line), the best three-marker panel introduced, RNASE2, TNF- $\alpha$ , and MAP3K7 (blue line), as well as the four-marker (yellow line) and five-marker (green line) panels. The AUC is displayed in the plot.

Panel	Sensitivity	Specificity	Accuracy	AUC
RNASE2, TNF- $\alpha$ , DNAJB2	0.974	1	0.984	0.99
RNASE2, TNF- $\alpha$ , MAP3K7, KCNJ2	0.949	1	0.968	0.98
RNASE2, TNF- $\alpha$ , MAP3K7, KCNJ3, DNAJB2	0.949	1	0.968	0.97

**Table 2.** Performance of machine learning on the real-life patient cohort data (validation cohort).

NO is highly diffusible and participates actively in various BPs due to its high reactivity. It generates metabolites that are central to genotoxic effects, including DNA damage. High concentrations of NO can hinder cancer growth and trigger cell apoptosis, whereas lower concentrations can stimulate tumor growth and proliferation<sup>31</sup>.

Furthermore, there were several remarkable pathways in HCC such as estrogen response, that known to be involved in the development and progression of cancers<sup>32,33</sup>, including HCC<sup>34</sup>. Coagulation was another significantly enriched pathway in HCC vs healthy controls. Various studies indicated the role of blood coagulation proteins in tumor progression<sup>35</sup>. Since the main site of coagulation factor production is liver, it has been suggested that these factors could be associated with the progression of HCC<sup>36</sup>.

In this study, we constructed a biological network based on the detected DEGs. The genes TNF- $\alpha$ , IL1B, ITGA2B, and P2RY12 were subsequently identified as hub genes in the PBMCs of HCC and can be considered therapeutic targets in HCC management. TNF- $\alpha$  is a crucial cytokine implicated in numerous signaling pathways, including inflammation, immunity, and even tumorigenesis. Predominantly generated by macrophages, its production is also observed in other cell types such as endothelial cells<sup>37</sup>. Its overexpression has been observed in monocytes isolated from PBMCs of colorectal cancer patients in comparison to normal controls<sup>38</sup>. However, in the liver, TNF- $\alpha$  possesses a dichotomous role: it can instigate liver tissue damage through hepatocyte apoptosis, while concurrently stimulating a regenerative effect and possibly promoting hepatocarcinogenesis via cell proliferation. Experimental findings suggested that complete inhibition or ablation of TNF- $\alpha$  attenuates liver cancer progression in laboratory mice. However, this modulation also correlates with reduced survival and the onset of hepatic failure.

Dysregulation of IL-1, another hub gene in the network, has been reported to be associated with various types of human malignancies. Accumulating evidence suggests that IL-1 plays an important role in tumorigenesis, cancer progression, metastasis, and the response to cancer treatment. IL-1 $\beta$  is one of the main agonists of the IL-1 family<sup>39</sup> which is produced by tumor and immune cells<sup>40</sup>. The serum level of IL-1B shows an elevation in the case of hepatitis, cirrhosis, liver fibrosis and HCC. The elevation of IL-1 $\beta$  serum levels can lead to overexpression of gankyrin which is an oncoprotein. Its overexpression induces the cell growth, invasiveness, and metastasis by regulating IL-1 $\beta$ /IRAK-1 signaling<sup>39</sup>. COX-2 expression that is also induced by IL-1 $\beta$ , inhibits antigen presenting cells from maturation and activation at the tumor microenvironment<sup>40</sup>.

ITGA2 constitutes one of the pivotal subunits, forming a heterodimer along with the  $\beta$ 1 subunit. The protein encoded by this gene functions as a transmembrane receptor integral to cellular adhesion processes, facilitating both cell–cell adhesion and the adherence of platelets and various cell types to the extracellular matrix. Moreover, it involves in cell surface-mediated signal transduction, modulating cellular growth and differentiation through the interaction with growth factors and chemokines<sup>41</sup>. In normal tissues and organs, the expression of ITGA2 remains very low; however, its overexpression has been identified in multiple cancer types, including HCC<sup>42</sup>. This is attributed to its role in critical cellular processes such as tumor cell proliferation, migration, invasion, angiogenesis, and metastasis<sup>41,43</sup>.

P2RY12 is an adenosine diphosphate (ADP)-responsive G protein-coupled receptor, whose expression is not exclusive to the surface of platelets but is also prominently observed in lymphocytes, monocytes, and megakaryocytes<sup>44</sup>, as well as cancer cells<sup>45</sup>. Research has elucidated that adenosine triphosphate (ATP) and ADP are secreted into the tumor site by both platelets and tumor cells, serving as activators for P2Y12, subsequently modulating the inflammatory response. Further investigations have shown that hepatic macrophages also express P2Y12 receptor. This receptor plays a crucial role in the induction of ER stress pathways. The induction of these pathways has a pivotal role in the pathogenesis of chronic liver diseases and HCC<sup>46</sup>.

In the next step, a computational feature selection was conducted on the selected DEGs. LASSO algorithm was used for multinomial computational feature selection. The three most significant features with the most absolute coefficient in each condition were utilized to construct 3 to 7-biomarker panels. Panels with the best performances included TNF- $\alpha$ , RNASE2, MAP3K7, CCNA1, DNAJB2, and KCNJ2.

Ribonuclease A family member 2 or RNASE2, also known as eosinophil-derived neurotoxin (EDN), belongs to the RNase superfamily, playing a pivotal role in the immune system and host defense mechanisms against pathogens. Upon activation by proinflammatory stimuli, eosinophils secrete a suite of proteins, with EDN being one of the four major proteins<sup>47</sup>. Additionally, monocyte-derived macrophages produce EDN upon stimulation, demonstrating capabilities in inducing cell proliferation, migration, and invasion<sup>48</sup>. Recent literature reports an overexpression of EDN in various tumors<sup>48–50</sup>. While existing studies have underscored the influence of immune-related genes on the tumor microenvironment, facilitating the progression of HCC<sup>51,52</sup>, this study marks the inaugural mention of RNASE2 in the context of HCC.

Mitogen-activated protein kinase kinase kinase 7 or MAP3K7, which is also known as transforming growth factor- $\beta$  activated kinase 1 (TAK1), is a pivotal member of the MAPKKK family. This kinase can be activated by a plethora of molecules, encompassing TGF- $\beta$ , cytokines (e.g., TNF- $\alpha$  and IL-1), Toll-like receptors, CD40, and B cell receptors. Additionally, several viruses, notably HBV and HCV, which are associated with a heightened risk of HCC, can induce MAP3K7 activation. Notably, there exists a positive correlation between MAP3K7 and

the expression of the mammalian target of rapamycin (mTOR), which is often linked to poor survival in HCC patients<sup>53</sup>. Furthermore, MAP3K7 can activate pathways such as NF- $\kappa$ B and MAPK, both of which are implicated in tumorigenesis<sup>54</sup>. Consequently, in agreement with our findings, overexpression of MAP3K7 can play a pivotal role in the inflammation and the progression of HCC<sup>53</sup>.

Among biomarker candidates, Cyclin A1 is a member of the A-type cyclin protein family that plays a pivotal role in regulating the G1/S phase transition of the cell division cycle. This regulation is achieved through its interaction with cyclin-dependent kinase 2 (CDK2) and cell division cycle 2 (CDC2) kinase. Additionally, Cyclin A1 has been identified to be associated with key cell cycle regulators such as retinoblastoma (Rb), p21 family proteins, and transcription factor E2F-1<sup>55</sup>. Dysregulation of Cyclin A1 has been implicated in various cancers including breast and thyroid cancers<sup>56</sup>. While the roles of cyclins CCNE1 and CCNB1 in HCC have been elucidated<sup>57,58</sup>, the involvement of CCNA1 in HCC remains unreported.

Potassium inwardly rectifying channel subfamily J member 2 or KCNJ2 belongs to the potassium inwardly rectifying channel family. Previous studies have highlighted the oncogenic roles of members of this family in various cancers. For instance, the interaction between KCNJ2 and the HIF1 $\alpha$  transcription factor has been shown to establish a positive-feedback loop, contributing to osteosarcoma metastasis<sup>59</sup>. Furthermore, inhibition of KCNJ2 has been associated with reduced proliferation, migration, and epithelial-mesenchymal transition (EMT) progression in papillary thyroid carcinoma cells<sup>60</sup>. In prostate cancer cells, KCNJ2 promotes proliferation by binding to the RELA protein in the nucleus, thereby activating the NF- $\kappa$ B signaling pathway<sup>61</sup>. Based on these findings, it is suggested that KCNJ2 might exert similar effects on HCC cells.

DnaJ heat shock protein family (HSP40) member B2, a member of HCC biomarker panels, belongs to HSP40 subclass DNAJB proteins. Emerging evidence indicates that DNAJB proteins play a crucial role in cancer invasion and metastasis by modulating diverse signaling pathways. However, certain members, such as DNAJB2, remain under-investigated. Even with limited data on these specific DNAJB proteins, they are believed to have a potential role in cancer development<sup>62</sup>.

While previous studies have been conducted to identify biomarkers for diagnosing HCC<sup>63</sup>, some have merely identified and introduced biomarkers based on bioinformatic analyses, lacking validation in independent real-life patient cohorts<sup>64,65</sup>. Using feature selection method, 3- to 7-biomarker panel was utilized as machine learning entry. Then, multinomial logistic regression algorithm was performed with fivefold cross validation to prevent the model from overfitting. Performing machine learning on experimental data obtained from an independent patient cohort and finding a 3-marker panel with the best performance confirmed the *in-silico* results.

Although some studies have introduced diagnostic biomarkers for HCC in PBMCs<sup>8,66</sup>, we utilized an integrated dataset comprising two distinct datasets. This approach was employed to eliminate batch effects from individual studies and to build a generalized predictive model using multi-centric data. Furthermore, this study represents a novel application of a predictive machine learning approach in PBMC samples for HCC, focusing on creating reliable models and accurately predicting this cancer from healthy controls and CHB that shares similar signaling pathways and biological processes with HCC. However, the performance of the introduced panels should be further investigated in a larger cohort to ensure their applicability in a clinical setting.

In conclusion, DEGs were identified from PBMC microarray data of HCC and CHB patients in the present study. Upregulated DEGs for each condition were then fed to feature selection method and a combination of features was determined to develop an input for machine learning. The best model of each 3-, 4-, and 5-biomarker panels were confirmed by qRT-PCR, and the 3-marker panel of TNF- $\alpha$ , RNASE2, and MAP3K7 found to have the best performance on experimental data, suggesting that it has the potential to be used as a non-invasive diagnostic panel with high accuracy for detecting HCC compared to healthy controls and CHB patients.

## Material and methods

### Data collection and sample preparation

The GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) was searched and two gene expression profile datasets (GSE49515 and GSE58208) that met our criteria were downloaded. Both of these datasets were based on the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array); therefore, probe annotation files were utilized for further analysis. One of the dataset profiles (GSE49515) consisted of 10 PBMC samples of HCC patients and 10 normal controls and the other one (GSE58208) contained the data of 10 HCC, 5 healthy controls, and 12 CHB samples as positive controls. These two datasets were then integrated and utilized as the discovery cohort.

Additionally, blood samples were obtained in EDTA tube from 39 patients with HCC and 15 CHB, who were diagnosed at Taleghani Hospital, Shahid Beheshti University of Medical Sciences (SBMU), and 24 healthy controls, between 2021 and 2023. The demographic characteristics of patients were also collected. These features included age, gender, family history of liver disease, overweight, drinking, smoking state, total bilirubin, serum albumin, aspartate transaminase, alanine transaminase, alkaline phosphatase, platelet count, and alpha fetoprotein. The blood samples of healthy controls were also collected from individuals without any evidence of liver diseases or HCC. Written informed consent was obtained from each participant and/or their legal guardian(s) and this research was approved by the ethical review board of Shahid Beheshti University of Medical Sciences, Tehran, Iran (IR.SBMU.RIGLD.REC.1402.0200). All methods were carried out in accordance with relevant guidelines and regulations. PBMC isolation was permed using Ficoll density gradient sedimentation (Lymphodex, Inno-Train, Germany).

### RNA isolation and quantitative reverse-transcription PCR

RNA isolation was performed on PBMCs using the TRIzol (Qiagen, USA) reagent based on the manufacturer's instruction. Then, cDNA was synthesized using cDNA synthesis kit (Parstous, Iran) in the presence of random hexamers. The synthesized cDNAs were stored at  $-20^{\circ}\text{C}$  for further experiments. Expression levels of the genes

were quantified by qRT-PCR. Briefly, quantitative real-time PCR (qRT-PCR) reactions were performed on a Rotor Gene Q System (QIAGEN) using SYBR Green master mix (AmpliQon, Denmark). The  $\Delta\text{Ct}$  was calculated for the obtained data. GAPDH (Glyceraldehyde-3-phosphate dehydrogenase) was used as an internal control gene. The detailed information about the primers is presented in Supplementary Table 6.

### Identification of DEGs

First, we utilized the combat function from the sva package in R to integrate datasets, improve the number of samples, and remove the batch effect. Then we performed differential analysis to statistically compare HCC, CHB, and normal samples. The genes with absolute  $\log_2(\text{fold change}) \geq 1$  and adjusted  $p$  value  $< 0.05$  in all comparisons including HCC vs. healthy, CHB vs. healthy, and HCC vs. CHB were considered as statistically differentially expressed. Limma package was used to determine DEGs in R. The correlation heatmap was plotted to illustrate the relationship between groups using Pearson correlation and the expression heatmap was used to explore the relationship of DEGs.

### Enrichment and functional analysis

We utilized Enrichr online tool (<https://maayanlab.cloud/Enrichr/>) to detect GO categories, including biological process (BP), molecular function (MF), and cellular components (CC) as well as signaling pathway analysis. Results with  $p < 0.05$  were considered significant. The results of top GOs and pathways were visualized using R.

### Biological network construction

DEGs were uploaded into the Search Tool for Retrieval of Interacting Genes/Proteins (STRING) (<https://string-db.org/>), which is an online tool designed for predicting protein–protein interactions (PPI), to find out the functional interactions between the proteins encoded by identified DEGs and construct a PPI network. For this purpose, all DEGs of HCC compared to healthy controls were uploaded to the STRING database. Various types of evidence, including text mining, experiments, databases, neighborhood, gene fusion, co-occurrence, and co-expression, were selected as sources of interaction. The minimum required interaction score was set at 0.4, meaning that only interactions above this threshold were included in the predicted network. The network was visualized using Cytoscape (version 3.9.1). The CentiScape plugin was employed to examine various parameters of the nodes, which included betweenness, closeness, degree, eccentricity, radiality, and stress. Furthermore, enrichment analysis was conducted to investigate BPs, MFs, and CCs associated with the network.

### Feature selection and machine learning

First, an online tool was used to obtain a Venn diagram of DEGs for each condition and healthy controls. Then, to identify the most important and effective genes among DEGs, unique upregulated DEGs selected from the Venn diagram were fed to the multinomial LASSO algorithm in R. Top 3 features in each condition (HCC, CHB, and healthy controls) were used to create a 3- to 7-marker panels of gene combinations. Multinomial logistic regression was used as a machine learning algorithm to perform classification and determine the diagnostic value of each model using glm in R. Since the sample size was small, logistic regression would be an appropriate choice for the classification. Five-fold cross-validation was used to evaluate each model and evaluation parameters such as sensitivity, specificity and accuracy was reported for each condition. pROC package was also utilized to plot the receiver operating characteristic (ROC) curves and estimate the area under the ROC curve (AUC).

To validate the potential predictive ability of the panels, a machine learning approach was employed using a real-life cohort consisting of 39 HCC, 15 CHB patients and 24 healthy individuals. For this step, binomial logistic regression was used. In addition to being suitable for a small number of samples, as mentioned previously, this algorithm is appropriate for building models for disease state (diseased/healthy) classification and decision-making problems and has been widely employed in health sciences studies. Therefore, this algorithm was used to build models based on the expression results obtained from qRT-PCR. Five-fold cross-validation was employed to improve the reliability of the model and to prevent overfitting.

### Statistical analysis

We performed statistical analysis using GraphPad Prism version 9. The comparison of selected genes between conditions in the discovery cohort was analyzed using one-way ANOVA, while t-test was employed in the validation cohort.  $p$ -value  $< 0.05$  was considered a statistically significant difference.

### Data availability

The data generated during the current study is available from the corresponding author on reasonable request.

Received: 13 January 2024; Accepted: 26 March 2024

Published online: 29 March 2024

### References

1. Petrick, J. L. *et al.* International trends in hepatocellular carcinoma incidence, 1978–2012. *Int. J. Cancer* **147**, 317–330 (2020).
2. Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
3. Suresh, D., Srinivas, A. N. & Kumar, D. P. Etiology of hepatocellular carcinoma: Special focus on fatty liver disease. *Front. Oncol.* **10**, 601710 (2020).
4. Yang, J. D. *et al.* A global view of hepatocellular carcinoma: Trends, risk, prevention and management. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 589–604 (2019).



5. Bharali, D. *et al.* Expression analysis of serum microRNA-34a and microRNA-183 in hepatocellular carcinoma. *Asian Pac. J. Cancer Prev. APJCP* **19**, 2561 (2018).
6. Russo, F. P., Imondi, A., Lynch, E. N. & Farinati, F. When and how should we perform a biopsy for HCC in patients with liver cirrhosis in 2018? A review. *Dig. Liver Dis.* **50**, 640–646 (2018).
7. Di Tommaso, L. *et al.* Role of liver biopsy in hepatocellular carcinoma. *World J. Gastroenterol.* **25**, 6041 (2019).
8. Ye, Z., Zeng, Z., Shen, Y. & Chen, Z. Identification of hub genes in peripheral blood mononuclear cells for the diagnosis of hepatocellular carcinoma using a weighted gene co-expression network analysis. *Exp. Ther. Med.* **20**, 890–900 (2020).
9. Wang, X. *et al.* Extracellular mRNA detected by molecular beacons in tethered lipoplex nanoparticles for diagnosis of human hepatocellular carcinoma. *PLoS ONE* **13**, e0198552 (2018).
10. Higashi, M. *et al.* A potential serum N-glycan biomarker for hepatitis C virus-related early-stage hepatocellular carcinoma with liver cirrhosis. *Int. J. Mol. Sci.* **21**, 8913 (2020).
11. Showe, M. K. *et al.* Gene expression profiles in peripheral blood mononuclear cells can distinguish patients with non-small cell lung cancer from patients with nonmalignant lung disease. *Cancer Res.* **69**, 9202–9210 (2009).
12. Kinouchi, T. *et al.* Expression level of CXCL 7 in peripheral blood cells is a potential biomarker for the diagnosis of renal cell carcinoma. *Cancer Sci.* **108**, 2495–2502 (2017).
13. Sharma, P. *et al.* Early detection of breast cancer based on gene-expression patterns in peripheral blood cells. *Breast Cancer Res.* **7**, R634. <https://doi.org/10.1186/bcr1203> (2005).
14. Kim, J., Patel, K., Jung, H., Kuo, W. P. & Ohno-Machado, L. AnyExpress: Integrated toolkit for analysis of cross-platform gene expression data using a fast interval matching algorithm. *BMC Bioinform.* **12**, 1–14 (2011).
15. Kim, E. & Viatour, P. Hepatocellular carcinoma: Old friends and new tricks. *Exp. Mol. Med.* **52**, 1898–1907 (2020).
16. Mosallaei, M. *et al.* PBMCs: A new source of diagnostic and prognostic biomarkers. *Arch. Physiol. Biochem.* **128**, 1081–1087 (2022).
17. Kunadirek, P. *et al.* Identification of BHLHE40 expression in peripheral blood mononuclear cells as a novel biomarker for diagnosis and prognosis of hepatocellular carcinoma. *Sci. Rep.* **11**, 11201 (2021).
18. Daniele, B., Bencivenga, A., Megna, A. S. & Tinessa, V.  $\alpha$ -fetoprotein and ultrasonography screening for hepatocellular carcinoma. *Gastroenterology* **127**, S108–S112 (2004).
19. Tzartzeva, K. *et al.* Surveillance imaging and alpha fetoprotein for early detection of hepatocellular carcinoma in patients with cirrhosis: A meta-analysis. *Gastroenterology* **154**, 1706–1718 (2018).
20. Chen, G., Zhang, W. & Ben, Y. Identification of key regulators of hepatitis C virus-induced hepatocellular carcinoma by integrating whole-genome and transcriptome sequencing data. *Front. Genet.* **12**, 741608 (2021).
21. Repsold, L. & Joubert, A. M. Platelet function, role in thrombosis, inflammation, and consequences in chronic myeloproliferative disorders. *Cells* **10**, 3034 (2021).
22. Zanetto, A. *et al.* Influence of hepatocellular carcinoma on platelet aggregation in cirrhosis. *Cancers* **13**, 1150 (2021).
23. Kullaya, V. I. *et al.* Platelets modulate innate immune response against human respiratory syncytial virus in vitro. *Viral Immunol.* **30**, 576–581. <https://doi.org/10.1089/vim.2016.0161> (2017).
24. Rachidi, S. *et al.* Platelets subvert T cell immunity against cancer via GARP-TGF $\beta$  axis. *Sci. Immunol.* **2**, eaai7911. <https://doi.org/10.1126/sciimmunol.aai7911> (2017).
25. Totani, L. & Evangelista, V. Platelet–leukocyte interactions in cardiovascular disease and beyond. *Arterioscler. Thromb. Vasc. Biol.* **30**, 2357–2361 (2010).
26. Stoiber, D. & Assinger, A. Platelet-leukocyte interplay in cancer development and progression. *Cells* **9**, 855 (2020).
27. Patel, S. R., Hartwig, J. H. & Italiano, J. E. The biogenesis of platelets from megakaryocyte proplatelets. *J. Clin. Investig.* **115**, 3348–3354 (2005).
28. Yun, S.-H., Sim, E.-H., Goh, R.-Y., Park, J.-I. & Han, J.-Y. Platelet activation: The mechanisms and potential biomarkers. *BioMed Res. Int.* **2016**, 15 (2016).
29. Pavlovic, N., Rani, B., Gerwins, P. & Heindryckx, F. Platelets as key factors in hepatocellular carcinoma. *Cancers* **11**, 1022 (2019).
30. Ramadori, P., Klag, T., Malek, N. P. & Heikenwalder, M. Platelets in chronic liver disease, from bench to bedside. *JHEP Rep.* **1**, 448–459 (2019).
31. Xu, W., Liu, L. Z., Loizidou, M., Ahmed, M. & Charles, I. G. The role of nitric oxide in cancer. *Cell Res.* **12**, 311–320 (2002).
32. Kozasa, K. *et al.* Estrogen stimulates female cancer progression by inducing myeloid-derived suppressive cells: Investigations on pregnant and non-pregnant experimental models. *Oncotarget* **10**, 1887 (2019).
33. Zeng, X. & Tao, H. Diagnostic and prognostic serum marker of cholangiocarcinoma. *Oncol. Lett.* **9**, 3–8 (2015).
34. Carruba, G. Estrogens in hepatocellular carcinoma: Friends or foes?. *Cancers* **13**, 2085 (2021).
35. Lima, L. G. & Monteiro, R. Q. Activation of blood coagulation in cancer: Implications for tumour progression. *Biosci. Rep.* **33**, e00064 (2013).
36. Li, X. *et al.* Intrahepatic metastases may be specific to hepatocellular carcinoma due to the coagulation and fibrinolytic systems. *Oncol. Rep.* **44**, 2345–2352 (2020).
37. Jang, M.-K., Su Kim, H. & Chung, Y.-H. Clinical aspects of tumor necrosis factor- $\alpha$  signaling in hepatocellular carcinoma. *Curr. Pharm. Design* **20**, 2799–2808 (2014).
38. Stanilov, N. S., Dobрева, Z. G. & Stanilova, S. A. Higher TNF- $\alpha$  production detected in colorectal cancer patients monocytes. *Biotechnol. Biotechnol. Equip.* **26**, 107–110. <https://doi.org/10.5504/50YRTIMB.2011.0020> (2012).
39. Litmanovich, A., Khazim, K. & Cohen, I. The role of interleukin-1 in the pathogenesis of cancer and its potential as a therapeutic target in clinical practice. *Oncol. Ther.* **6**, 109–127 (2018).
40. Zekri, A.-R.N. *et al.* Role of relevant immune-modulators and cytokines in hepatocellular carcinoma and premalignant hepatic lesions. *World J. Gastroenterol.* **24**, 1228 (2018).
41. Cai, H. *et al.* Overexpressed integrin  $\alpha$ 2 inhibits the activation of the transforming growth factor  $\beta$  pathway in pancreatic cancer via the TFCP2-SMAD2 axis. *J. Exp. Clin. Cancer Res.* **41**, 1–18 (2022).
42. Zhang, L. *et al.* Is integrin subunit  $\alpha$ 2 expression a prognostic factor for liver carcinoma? A validation experiment based on bioinformatics analysis. *Pathol. Oncol. Res.* **25**, 1545–1552 (2019).
43. Shang, L. *et al.* Prognostic value of integrin variants and expression in post-operative patients with HBV-related hepatocellular carcinoma. *Oncotarget* **8**, 76816 (2017).
44. Schrottmaier, W. C., Kral, J. B., Badrnya, S. & Assinger, A. Aspirin and P2Y12 Inhibitors in platelet-mediated activation of neutrophils and monocytes. *Thromb. Haemost.* **114**, 478–489 (2015).
45. Van Kolen, K., Gilany, K., Moens, L., Esmans, E. L. & Slegers, H. P2Y12 receptor signalling towards PKB proceeds through IGF-1 receptor cross-talk and requires activation of Src, Pyk2 and Rap1. *Cell. Signal.* **18**, 1169–1181 (2006).
46. Pavlović, N., Kopsida, M., Gerwins, P. & Heindryckx, F. Inhibiting P2Y12 in macrophages induces endoplasmic reticulum stress and promotes an anti-tumoral phenotype. *Int. J. Mol. Sci.* **21**, 8177 (2020).
47. Zhu, Y. *et al.* RNASE2 mediates age-associated B cell expansion through monocyte derived IL-10 in patients with systemic lupus erythematosus. *Front. Immunol.* **13**, 752189 (2022).
48. Wu, T. *et al.* Ribonuclease A family member 2 promotes the malignant progression of glioma through the PI3K/Akt signaling pathway. *Front. Oncol.* **12**, 921083 (2022).
49. Wan, B., Liu, B., Huang, Y., Yu, G. & Lv, C. Prognostic value of immune-related genes in clear cell renal cell carcinoma. *Aging* **11**, 11474 (2019).

50. Li, Y. *et al.* Identification of a novel immune prognostic model in gastric cancer. *Clin. Transl. Oncol.* **23**, 846–855 (2021).
51. Peng, Y. *et al.* Identification of a prognostic and therapeutic immune signature associated with hepatocellular carcinoma. *Cancer cell Int.* **21**, 1–21 (2021).
52. Shen, B., Zhang, G., Liu, Y., Wang, J. & Jiang, J. Identification and analysis of immune-related gene signature in hepatocellular carcinoma. *Genes* **13**, 1834 (2022).
53. Cheng, J.-S. *et al.* The MAP3K7-mTOR axis promotes the proliferation and malignancy of hepatocellular carcinoma cells. *Front. Oncol.* **9**, 474 (2019).
54. Jin, X. *et al.* High expression of LINC01268 is positively associated with hepatocellular carcinoma progression via regulating MAP3K7. *OncoTargets Ther.* **14**, 1753–1769 (2021).
55. He, X. *et al.* Cyclin A1 is associated with poor prognosis in oesophageal squamous cell carcinoma. *Oncol. Lett.* **18**, 706–712 (2019).
56. da Silva, R. M. *et al.* CCNA1 gene as a potential diagnostic marker in papillary thyroid cancer. *Acta Histochem.* **122**, 151635 (2020).
57. Sonntag, R. *et al.* Cyclin E1 and cyclin-dependent kinase 2 are critical for initiation, but not for progression of hepatocellular carcinoma. *Proc. Natl. Acad. Sci.* **115**, 9282–9287 (2018).
58. Huang, Q., Li, J. & Wei, A. Identification of potential therapeutic targets in hepatocellular carcinoma using an integrated bioinformatics approach. *Transl. Cancer Res.* **7**, 849–858 (2018).
59. Shen, M. *et al.* KCNJ2/HIF1 $\alpha$  positive-feedback loop promotes the metastasis of osteosarcoma. *Cell Commun. Signal.* **21**, 46 (2023).
60. Chen, S., Huang, M. & Hu, X. Interference with KCNJ2 inhibits proliferation, migration and EMT progression of apillary thyroid carcinoma cells by upregulating GNG2 expression. *Mol. Med. Rep.* **24**, 1–10 (2021).
61. Lee, I., Lee, S.-J., Kang, T. M., Kang, W. K. & Park, C. Unconventional role of the inwardly rectifying potassium channel Kir2. 2 as a constitutive activator of RelA in cancer. *Cancer Res.* **73**, 1056–1062 (2013).
62. Kim, H.-Y. & Hong, S. Multi-faceted roles of DNAJB protein in cancer metastasis and clinical implications. *Int. J. Mol. Sci.* **23**, 14970 (2022).
63. Shaker, M. K. *et al.* Evaluation of golgi protein 73 (GP73) as a potential biomarkers for hepatocellular carcinoma. *Clin. Lab.* <https://doi.org/10.7754/Clin.Lab.2020.190911> (2020).
64. Yin, L. *et al.* Identification of novel blood-based HCC-specific diagnostic biomarkers for human hepatocellular carcinoma. *Artif. Cells Nanomed. Biotechnol.* **47**, 1908–1916 (2019).
65. Peng, J.-L. *et al.* Identification of potential biomarkers of peripheral blood mononuclear cell in hepatocellular carcinoma using bioinformatic analysis: A protocol for systematic review and meta-analysis. *Medicine* **100**, e24172 (2021).
66. Jiang, J. X. *et al.* Insights into significant pathways and gene interaction networks in peripheral blood mononuclear cells for early diagnosis of hepatocellular carcinoma. *J. Cancer Res. Ther.* **12**, 981–989 (2016).

## Acknowledgements

We would like to thank the Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tarbiat Modares university, and also Raheleh Heydari and Mahmoud Amiri whose support made it possible to carry out this study.

## Author contributions

S.F. and A.M. contributed to data collection, methodology, analysis and interpretation of the results, writing and editing the manuscript. M.G. contributed to performing experiments, investigation and writing the manuscript. A.R. contributed to clinical data and sample collection. J.Z. advisor. A.M. contributed to Conceptualization, project administration, supervision, writing—review and editing. All authors participated in the finalization of the manuscript, and approved the final draft.

## Funding

This study was partly supported by the Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-58158-9>.

**Correspondence** and requests for materials should be addressed to A.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024