Discover Oncology

Analysis

Identification of novel potential biomarkers using bulk RNA and single cells to build a neural network model for diagnosis of liver cancer

Yingzheng Gao[1] · Jiahao Chen[1] · Weidong Du[2]

Received: 17 November 2024 / Accepted: 17 April 2025
Published online: 12 May 2025

Abstract

Background As a common cancer, liver cancer imposes an unacceptable burden on patients, but its underlying molecular mechanisms are still not fully understood. Therefore, there is an urgent need to potential biomarkers and diagnostic models for liver cancer.

Methods In this study, transcriptome and single-cell datasets related to liver cancer were downloaded from the UCSC Xena database and the Mendeley database, and differential analysis and weighted gene co-expression network analysis were used to find differentially expressed genes related to liver cancer. We used multiple machine algorithms to find hub genes related to liver cancer, and constructed new artificial neural network models based on their transcriptome expression patterns to assist in the diagnosis of liver cancer. Subsequently, we conducted survival analysis and immune infiltration analysis to explore the correlation between hub genes and immune cells, and used single-cell data to verify hub genes related to liver cancer.

Results This study identified *MARCO*, *KCNN2*, *NTS*, *TERT* and *SFRP4* as central genes associated with liver cancer, and constructed a new artificial neural network model for molecular diagnosis of liver cancer. The diagnostic performance of the training cohort and the validation cohort was good, with the areas under the ROC curves of 1.000 and 0.986, respectively. Immune infiltration analysis determined that these central genes were closely associated with different types of immune cells. The results of immunohistochemistry and the results at the single cell level were consistent with those at the transcriptome level, and also showed obvious differences between different cell types in liver cancer and healthy states.

Conclusion This study identified *MARCO*, *KCNN2*, *NTS*, *TERT*, and *SFRP4* from multiple dimensions and highlighted their key roles in the diagnosis and treatment of liver cancer from multiple dimensions, providing promising biomarkers for the diagnosis of liver cancer.

Keywords Liver cancer · Molecular marker · Machine learning algorithm · Artificial neural network · Single cell analysis

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-02420-7.

✉ Weidong Du, doctordu20@163.com | [1]The First School of Clinical Medicine, Zhejiang Chinese Medical University, Hangzhou 310006, China. [2]The First Affiliated Hospital of Zhejiang, Zhejiang Provincial Hospital of Traditional Chinese Medicine, Chinese Medical University, Hangzhou 310006, China.

1 Introduction

Liver cancer (LC) is the third leading cause of cancer death [1]. Surgical resection or liver transplantation can treat resectable LC, while sorafenib is used for advanced, unresectable cases, but recurrence is common [2]. LC is often caused by chronic liver disease, primarily hepatitis B (accounting for 50% of cases), hepatitis C, alcohol abuse, and metabolic syndrome [3]. Hepatitis B and C infections are global health problems, causing approximately 1 million new cases and 450,000 deaths each year [4].

The pathogenesis of liver cancer is a complex multi-step process with a variety of histological features and mutations [5–8]. In 80% of LC cases, telomerase activation occurs, usually driven by mutations in the telomerase reverse transcriptase promoter [6, 9]. Solid liver tumors are composed of malignant cells and stromal cells, and stromal cells play a key role in tumor initiation and progression [10, 11]. Tumor-associated macrophages (TAMs), neutrophils (TANs), and dendritic cells within the tumor microenvironment (TME) promote tumor growth, metastasis, and invasion. In addition, immunosuppression in chronic liver disease (especially T cell immunosuppression) is associated with the development of liver cancer [12].

LC is highly heterogeneous, and recent advances in nucleic acid profiling have identified several candidate genes involved in its development, particularly those that affect cell cycle regulation, the Wnt/β-catenin pathway, and epigenetic mechanisms [13, 14]. Some of these genes have shown promise as therapeutic targets. Studies of the tumor microenvironment (TME) have shown that it plays a crucial role in promoting HCC growth by providing a supportive environment for cancer cells and suppressing immune activity. The TME appears to have a more consistent gene expression profile across patients than LC itself, making it an attractive target for novel therapeutics. Although hepatectomy and liver transplantation remain the only curative treatments, immunotherapy regimens have extended the median survival of patients with intermediate or advanced LC from 3 months with tyrosine kinase inhibitors (TKIs) to 20 months [15]. Although the incidence of LC has decreased due to HBV vaccination and HCV antiviral treatment, rising rates of alcohol abuse and obesity have led to an increase in cases [17]. This highlights the urgent need to explore new single-cell biomarkers to better understand the mechanisms that determine the efficacy of LC immunotherapy [16].

To systematically identify potential molecular markers for LC, we applied multiple machine learning algorithms to transcriptomics and single-cell transcriptomics, identifying *MARCO*, *KCNN2*, *NTS*, *TERT*, and SFRP4 as hub genes associated with LC. Furthermore, we classified different molecular subtypes of LC based on the expression patterns of these genes. Survival analysis and immune invasion analysis revealed significant differences in survival and immune cell infiltration among different subtypes. Using these hub genes, we developed a novel artificial neural network (ANN) model that showed strong diagnostic performance for LC in both training and testing cohorts. Furthermore, we analyzed the expression of these hub genes in single-cell transcriptomes and identified the cell types significantly affected by them. Collectively, these findings enhance our understanding of the functional roles of LC genes and provide potential biomarkers and therapeutic targets.

To systematically identify potential molecular markers of LC, we combined multiple machine learning algorithms with transcriptomics and single-cell transcriptomics and identified *MARCO*, *KCNN2*, *NTS*, *TERT*, and *SFRP4* as central genes associated with LC. Furthermore, we classified LC into different molecular subtypes based on the expression patterns of these genes. Survival analysis and immune invasion analysis revealed significant differences in survival and immune cell infiltration between subtypes. Using these central genes, we developed a novel artificial neural network (ANN) model that demonstrated strong diagnostic performance for LC in both training and testing cohorts. Furthermore, we analyzed the expression of these hub genes in single-cell transcriptomes and identified the cell types significantly affected by them. Finally, we validated our results by comparing immunohistochemical staining images of LC-HGs in healthy and LC tissues. Taken together, these findings improve our understanding of the functional roles of LC genes and provide potential biomarkers and therapeutic targets.

2 Materials and methods

2.1 Data downloading and processing

We downloaded a liver transcriptome dataset consisting of 419 samples from the UCSC Xena database (https://xena.ucsc.edu/), including 369 liver cancer samples and 50 healthy controls. Additionally, we obtained a single-cell dataset

[18] from Mendeley Data (https://data.mendeley.com/datasets/6wmzcskt6k/1), which includes 18 samples—8 liver cancer samples (57,254 cells) and 10 normal samples (77,189 cells). Furthermore, we downloaded independent liver cancer cohorts GSE272166 and GSE265834 from the GEO database for validation.

2.2 Differential analysis of gene expression

We used the "limma [19]" R package to compare the expression profiles of LC and normal samples, aiming to identify differentially expressed genes (DEGs) between the two groups. Using a threshold of |log2 FoldChange|> 2 and an adjusted $P < 0.05$, we identified 1,629 DEGs.

2.3 Weighted gene co-expression network analysis

We performed weighted gene co-expression network analysis (WGCNA) on 19,620 genes in the samples using the R package "WGCNA" [20]. To ensure the co-expression network approximated a scale-free distribution, we selected a soft power of 3 for the WGCNA analysis, which resulted in the identification of 6 modules. We then assessed the relationships of these modules with normal and LC samples. The turquoise module with the closest relationship to LC was selected. Finally, we intersected the 1,263 genes in the turquoise module with the 1629 DEGs, identifying 224 DEGs related to liver cancer.

2.4 Protein–protein interactions among differentially expressed LC genes

Protein–protein interactions (PPIs) between LC DEGs were analyzed using the STRING database (https://string-db.org/) and Cytoscape software [21].

2.5 GO and KEGG functional enrichment analysis

We used the DAVID online tool (https://david.ncifcrf.gov/) to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses on the 224 differentially expressed LC genes to parse their biological functions, where a P value less than 0.05 was considered statistically significant.

2.6 Selection of hub genes and enrichment analysis of LC-HGs

We used three machine learning algorithms to identify hub genes, including support vector machine recursive feature elimination (SVM-RFE), random forest (RF) and least absolute shrinkage and selection operator (LASSO). SVM-RFE is a backward selection method that ranks features based on model training scores and selects the optimal subset [22]. RF, an ensemble learning technique, combines multiple decision trees for classification [23]. LASSO, a linear regression technique with L1 regularization, selects features by driving some coefficients to zero, enhancing sparsity [24]. We analyzed 224 LC DEGs using the SVM-RFE algorithm in the "e1071" package [25], the RF algorithm in the "randomForest" package [26], and the LASSO algorithm in the "glmnet" package [27] to identify candidate genes. Venn diagram analysis of the candidate genes from these algorithms revealed five intersecting hub genes (LC-HGs).

2.7 Construction of a new ANN model for diagnosing LC

To diagnose LC more effectively, we developed an artificial neural network (ANN) model based on LC-HG using the R package "neuralnet" [28] to improve the accuracy of diagnosing LC. The model consists of three main parts: 1) an input layer containing information of five LC-HGs; 2) three hidden layers—the first layer contains the gene expression and weights of the five LC-HGs, the second layer contains the weights from the first hidden layer, and the third layer contains the weights from the second hidden layer; 3) an output layer that classifies the sample as"healthy"or"LC". To improve the convergence speed and accuracy, we set the number of neurons in the first hidden layer to 9, the number of neurons in the second hidden layer to 4, and the number of neurons in the third hidden layer to 2. And the ROC curves on the training and test sets were used to evaluate the prediction performance of the ANN.

2.8 Assessing the diagnostic value of LC-HGs

To further determine the diagnostic value of LC-HGs for LC, we evaluated the performance of these genes in the combined dataset using ROC curves. The expression data and disease status information of LC-HGs were extracted from the samples. The “roc” function in the “pROC” package was used to perform ROC analysis [29], and the “ci” function was used to calculate the area under the curve (AUC).

2.9 Consensus clustering analysis

We applied the “ConsensusClusterPlus” package [30] to classify LC into two molecular subtypes based on the previously identified LC-HGs. We used the PAM algorithm with Euclidean distance for 1000 iterations on the samples. The k value ranged from 2 to 9 to determine the optimal number of clusters. Two-sided log-rank test and Kaplan–Meier survival analysis were then performed using the R package “survminer” [31] to evaluate the overall survival difference between the two clusters.

2.10 Immune infiltration analysis

We performed single-sample gene set enrichment analysis (ssGSEA) to assess immune cell infiltration using the R packages “GSVA” [32] and “GSEABase” [33] to examine the immunological characteristics of LC patients. We retrieved the gene information of 28 immune cell gene sets (Table S1) from the TISIDB database (http://cis.hku.hk/TISIDB/) for ssGSEA and calculated the ssGSEA score for each sample, representing the abundance of specific immune cell types. Wilcox test was used to perform pairwise comparisons between healthy and LC samples to determine differences in immune cell infiltration. Finally, we visualized the correlation between immune cell infiltration and LC-HGs expression.

2.11 Single-cell sequencing analysis

We analyzed the single-cell data using the"Seurat"package [33], selecting the top 2,000 highly variable genes. The"Harmony"package [34] was employed to correct batch effects. We used the first 30 principal components for clustering and applied the Uniform Manifold Approximation and Projection (UMAP) algorithm for visualization. LC-HG expression was retrieved for each cell type, followed by a differential expression analysis.

2.12 Clinical validation of LC-HGs

To verify the RNA expression level of LC-HGs in healthy and tumor samples, we downloaded the immunofluorescence data of LC-HGs from the Human Protein Atlas database (HPA, https://www.proteinatlas.org/) [35] to indicate their subcellular localization.

2.13 Statistical analysis

We used the “limma [19]” R package to compare the expression of LC-HGs extracted from the validation set. All statistical analyses and visualizations in this study were performed using R software. Analysis of variance (ANOVA) [36] was used for multiple group comparisons, while the Wilcoxon rank sum test was used for pairwise comparisons. Figure 1 shows the complete technical pipeline of this study.

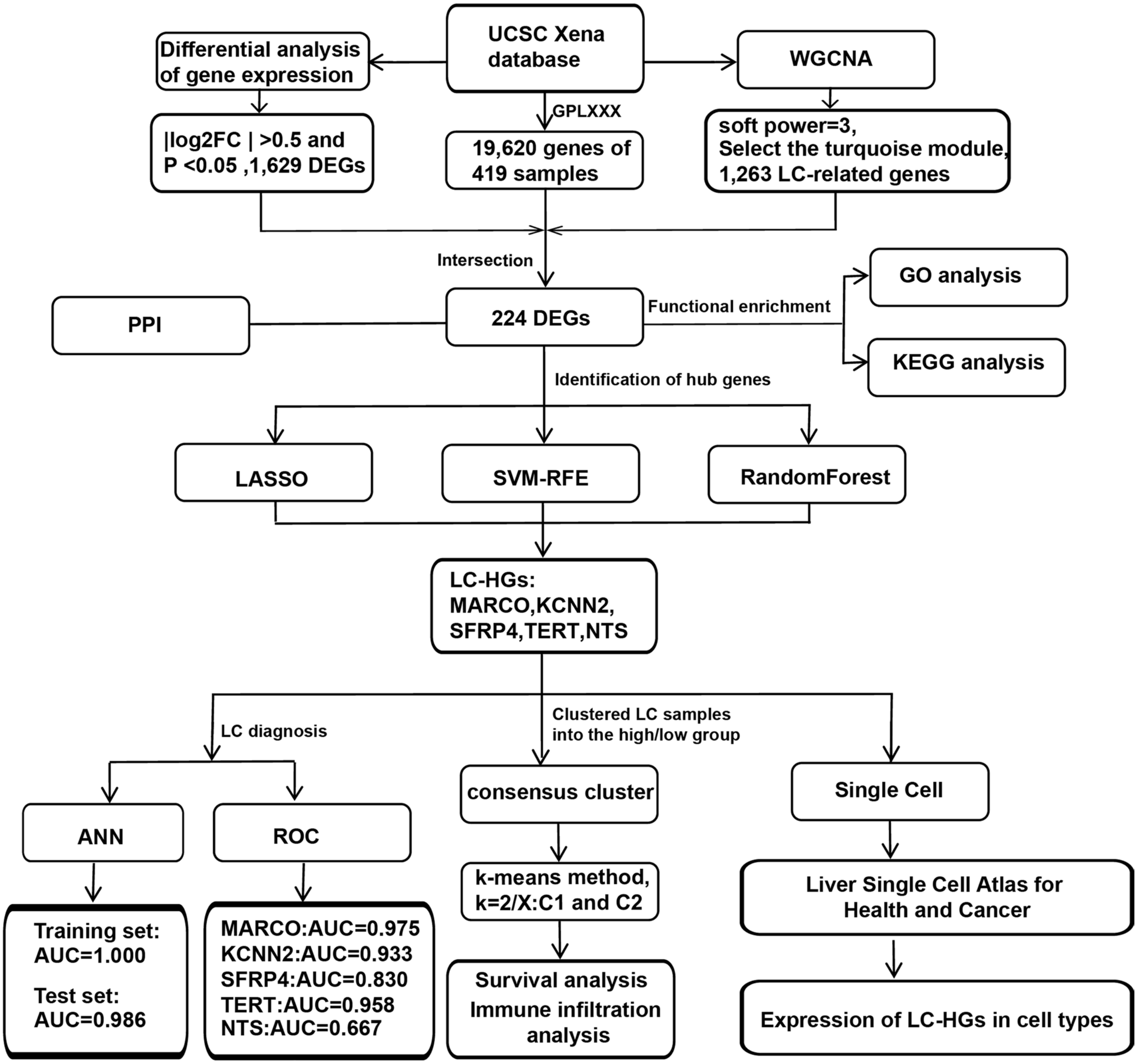
UCSC Xena database
Differential analysis of gene expression
WGCNA
GPLXXX
|log2FC | >0.5 and P <0.05 ,1,629 DEGs
19,620 genes of 419 samples
soft power=3, Select the turquoise module, 1,263 LC-related genes
Intersection
224 DEGs
PPI
Functional enrichment
GO analysis
KEGG analysis
Identification of hub genes
LASSO
SVM-RFE
RandomForest
LC-HGs: MARCO,KCNN2, SFRP4,TERT,NTS
LC diagnosis
Clustered LC samples into the high/low group
ANN
ROC
consensus cluster
Single Cell
Training set: AUC=1.000
Test set: AUC=0.986
MARCO:AUC=0.975
KCNN2:AUC=0.933
SFRP4:AUC=0.830
TERT:AUC=0.958
NTS:AUC=0.667
k-means method, k=2/X:C1 and C2
Survival analysis Immune infiltration analysis
Liver Single Cell Atlas for Health and Cancer
Expression of LC-HGs in cell types

Fig. 1 The complete technical route of this research

3 Results

3.1 Screening for DEGs between LC samples and healthy samples

First, we performed differential gene expression analysis by comparing 369 LCs with 50 normal tissues in the combined dataset and obtained 1629 DEGs (Table S2) (Fig. 2A). These 1629 DEGs were significantly differentially expressed between LC and normal samples (Figure S1 A).

3.2 Identification of LC module genes

We performed WGCNA analysis on 19,620 genes in LC samples to identify LC module genes. We first selected 3 as appropriate soft intensities (Fig. 2B), and then clustered them using the sample dendrogram and LC feature heat

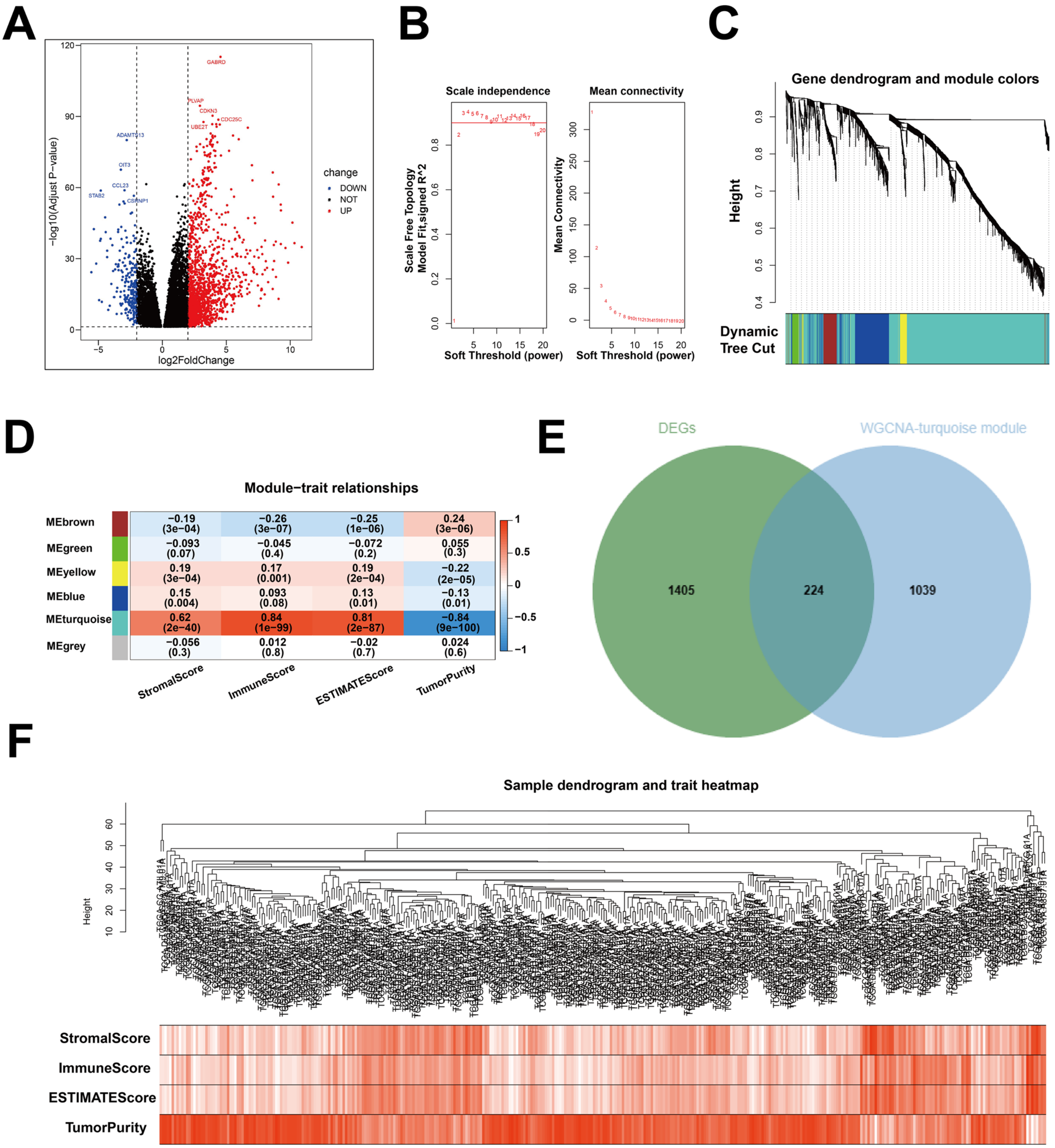

Fig. 2 Differential analysis and WGCNA identified LC DEGs. **A** Volcano plot of DEGs between LC and healthy samples, red represents up-regulation and blue represents down-regulation. **B** Plot of soft threshold power with scale-free topological model fit index and average connectivity. Three was selected as the appropriate soft power. **C** Gene dendrogram showing differential metric clustering. **D** Turquoise module is the module most specific to LC traits in the samples. **E** Venn diagram showing the intersection of DEGs with genes in the turquoise module. **F** Sample dendrogram and heat map of the traits shown

map (Fig. 2C) to view the overall situation of the normal group and LC group in the sample (Fig. 2F). The cyan module was determined to be the most strongly associated with LC through module gene clustering (Fig. 2D). Finally, 1263 genes associated with LC were identified (Table S3).

3.3 LC DEGs and functional enrichment analysis

We intersected the 1629 DEGs with the 1263 turquoise module genes and obtained 224 LC DEGs (Fig. 2E). PPI showed strong interactions (Figure S1B). Subsequently, in order to better explain the pathogenesis of LC, we used GO and KEGG analysis to clarify which biological processes and functions were enriched in these 224 DEGs. GO results showed that these DEGs were mainly related to energy generation and mitochondrial autophagy, such as "antimicrobial humoral response", "genitalia development", "ventral spinal cord development", "extracellular matrix disassembly", "nervous system process involved in regulation of systemic arterial blood pressure", "catenin complex", "extrinsic component of plasma membrane", "postsynaptic endocytic zone", "extrinsic component of membrane", "muscle myosin complex", "synaptic membrane", "oligosaccharide binding", "receptor ligand activity", "signaling receptor activator activity" and "oxidoreductase activity, acting on single donors with" (Figure S2 A). Additionally, KEGG results showed that they were associated with amino acid metabolism and cancer, that is, "Calcium signaling pathway", "Neuroactive ligand-receptor interaction", "Ascorbate and aldarate metabolism", "Melanoma", "Dopaminergic synapse", "Gastric cancer" (Figure S2B).

3.4 Determination of LC-HGs by machine learning

Next, we used three machine learning algorithms to train and analyze the 224 genes to identify key LC DEGs. These algorithms have been widely used to analyze biological data and accurately identify core genes in gene expression profiles.

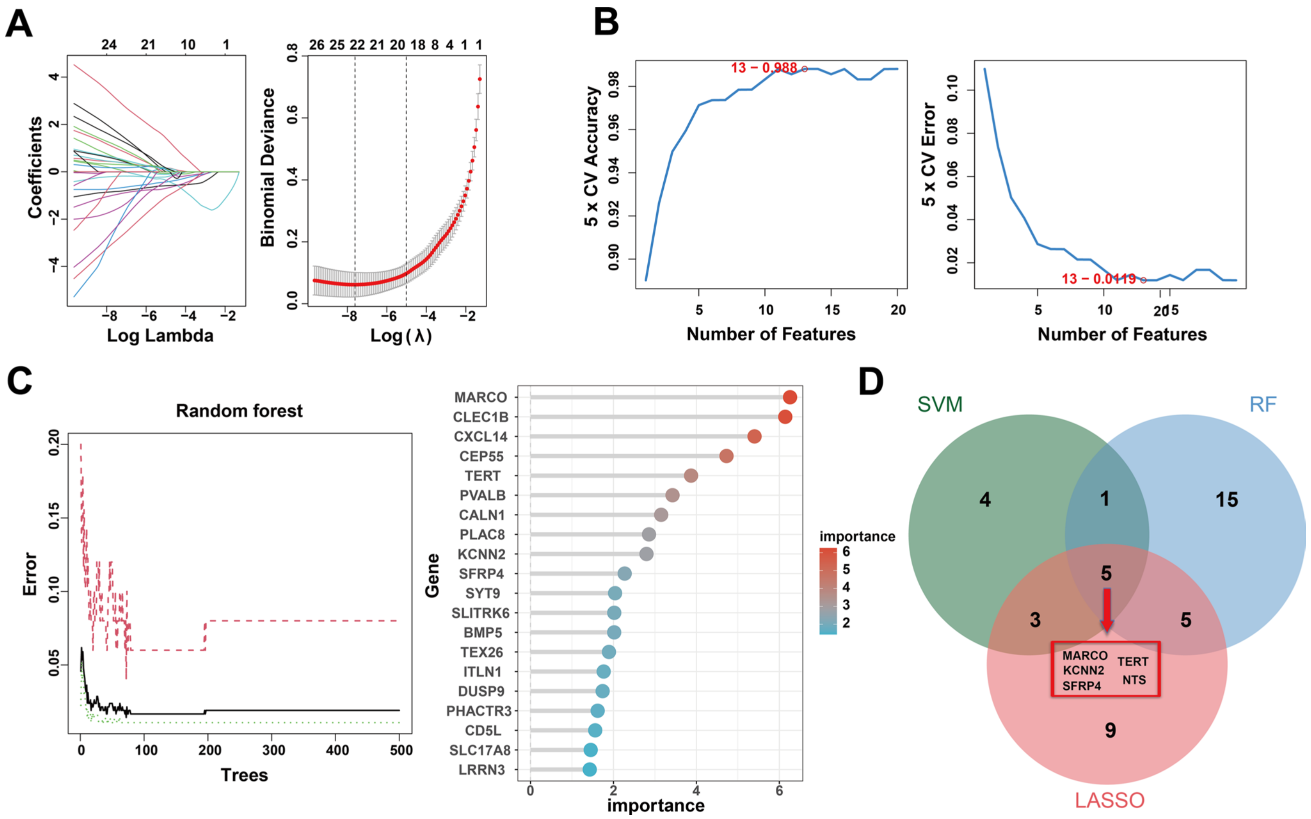

Fig. 3 Identification of LC-HGs. **A** Overview of LASSO coefficients of candidate genes and cross-validation used to adjust the selection of predictors. **B** 13 potential genes were identified by the SVM-RFE algorithm with an accuracy of 1. **C** RF model error rate vs. the number of classification trees and gene importance scores. **D** Venn diagram of hub genes screened by RF, SVM-RFE, and LASSO algorithms

Table 1 Optimal candidate genes screened by different machine learning algorithms

Methods	Genes
LASSO	TERT, MMP11, MARCO, KCNN2, SFRP4, CXCL14, CA4, TPPP2, PLAC8, CYP39 A1, NTS, ZNF541, HIST1H2 AG, ADAMDEC1, BMP5, SLC17 A8, SCUBE1, AMPD1, DEFB132, FGF23, OPRPN, PRAMEF15
RF	MARCO, CLEC1B, CXCL14, CEP55, TERT, PVALB, CALN1, PLAC8, KCNN2, SFRP4, SYT9, SLITRK6, BMP5, TEX26, ITLN1, DUSP9, PHACTR3, CD5L, SLC17 A8, LRRN3, GYS2, DRP2, FGF23, UNC13 A, HGF, NTS
SVM-REF	SFRP4, TPPP2, TERT, NTS, OPRPN, CTNNA3, KCNN2, PRAMEF15, OR7 C1, ISL2, MARCO, BPIFB4, CEP55

First, we used the LASSO algorithm to identify the changes in the regression coefficients of the 224 DEGs, and used tenfold cross-validation to select the optimal and minimum criteria for the penalty parameter (λ) (Fig. 3A), and screened a total of 22 candidate genes. We used the SVM-RFE model to screen out 9 candidate genes with the highest accuracy and lowest error rate (Fig. 3B). In addition, we included the 224 DEGs into the RF model and screened out 26 candidate genes with importance scores > 0 (Fig. 3C). Finally, we crossed the candidate genes screened by the above three algorithms (Table 1). *MARCO*, *KCNN2*, *NTS*, *TERT* and *SFRP4* were identified by all three machine learning approaches (Fig. 3D)

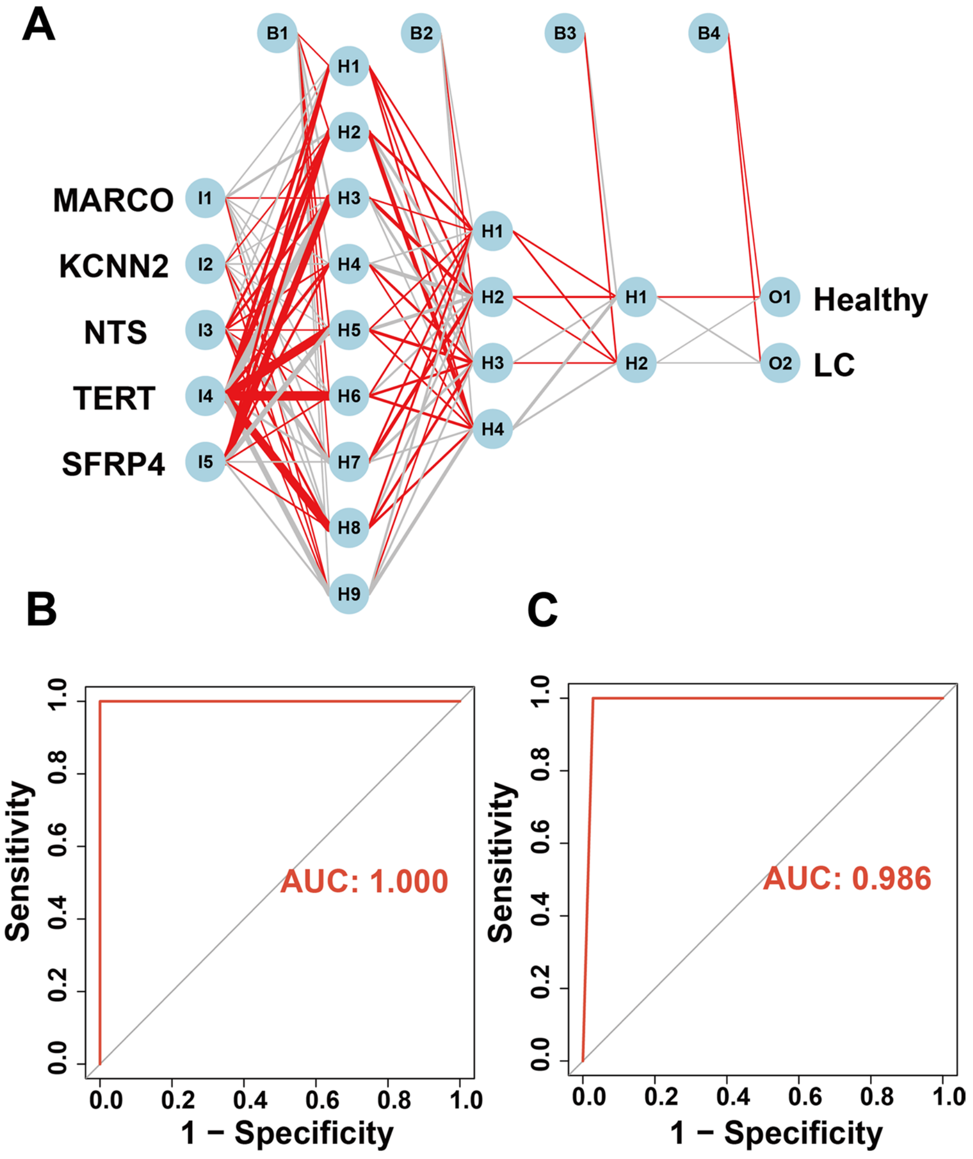

Fig. 4 An ANN model was constructed based on LC-HGs to diagnose LC. **A** ANN model. **B** Training set AUC = 1 (**C**) Test set AUC = 0.986

Table 2 Artificial neural network prediction performance for training and test sets

		Training set		Test set	
		Healthy	LC	Healthy	LC
Prediction	Healthy	42	0	8	2
	LC	0	298	0	69
	Healthy accuracy	1.000		1.000	
	LC accuracy	1.000		0.972	
	AUC	1.000		0.986	

so were defined as LC-HGs in this study. Enrichment analysis of the identified LC-HGs revealed that these genes were associated with "GnRH secretion", "Insulin secretion", "Bile secretion", "Serotonergic synapse", "Gastric cancer", "Phagosome", and "Hepatocellular carcinoma" pathways (Figure S5).

3.5 ANN model verifies the diagnostic performance of LC-HGs

We examined the significance of five LC-HGs in healthy and LC samples and found significant differences in all of them (Figure S3), ROC analysis showed that the AUC of LC-HGs performed well (Figure S4). Subsequently, we incorporated the five LC-HGs into an ANN designed to predict whether a sample has LC (Fig. 4A). We compared the ANN prediction results with the actual grouping information of the samples. The prediction accuracy of the training set was 1.000, and the prediction accuracy of the test set was 0.986 (Table 2). Finally, we used the ROC curve to evaluate the predictive ability of the ANN model for the training set and the test set. The AUC value of the training set was 1.000 (Fig. 4B) and the AUC value of the test set was 0.875 (Fig. 4C). Overall, the performance of the ANN model was convincing, and LC-HGs has the potential to be used as an independent diagnostic predictor for LC.

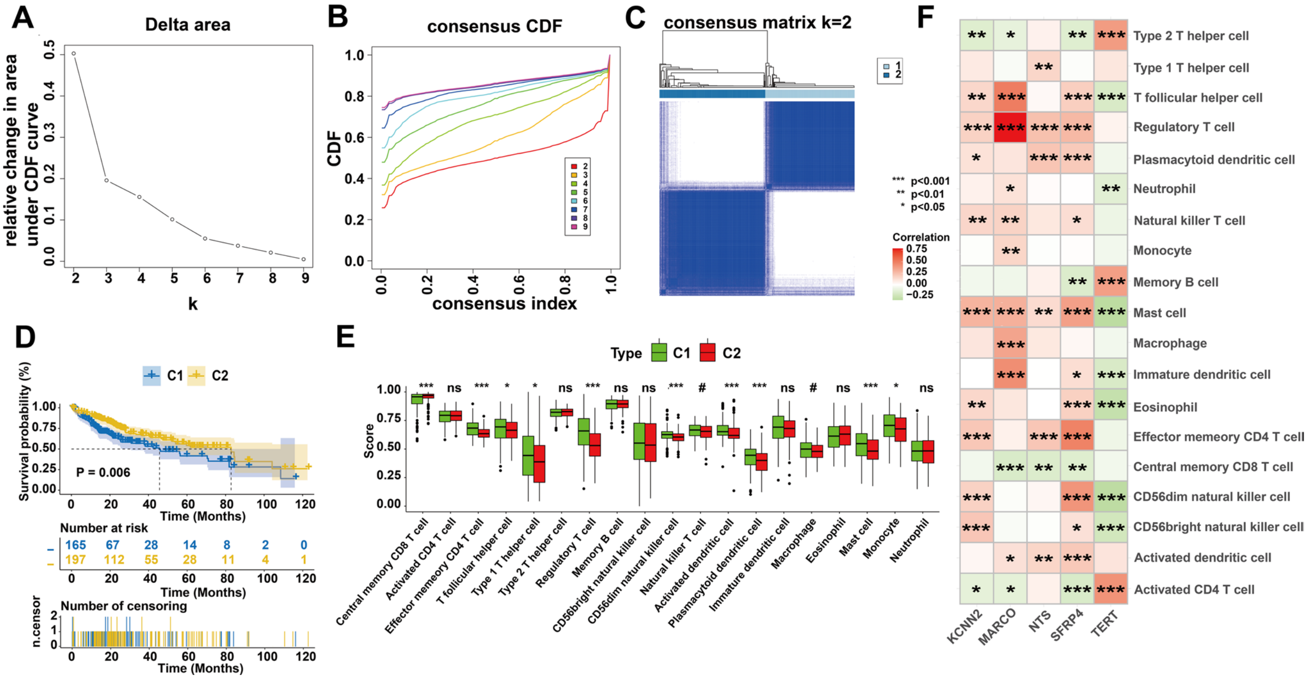

Fig. 5 Identification of LC subtypes, survival analysis and immune infiltration analysis. **A–C** Consensus clustering of LC samples, k = 2. **D** Survival status of C1 and C2 subgroups. **E** The difference of immune cell infiltration abundance between the C1 samples and C2 samples (*p < 0.05, **p < 0.01, ***p < 0.001). **F** Correlation between LCs expression and immune cell infiltration abundance (*p < 0.05, **p < 0.01, ***p < 0.001)

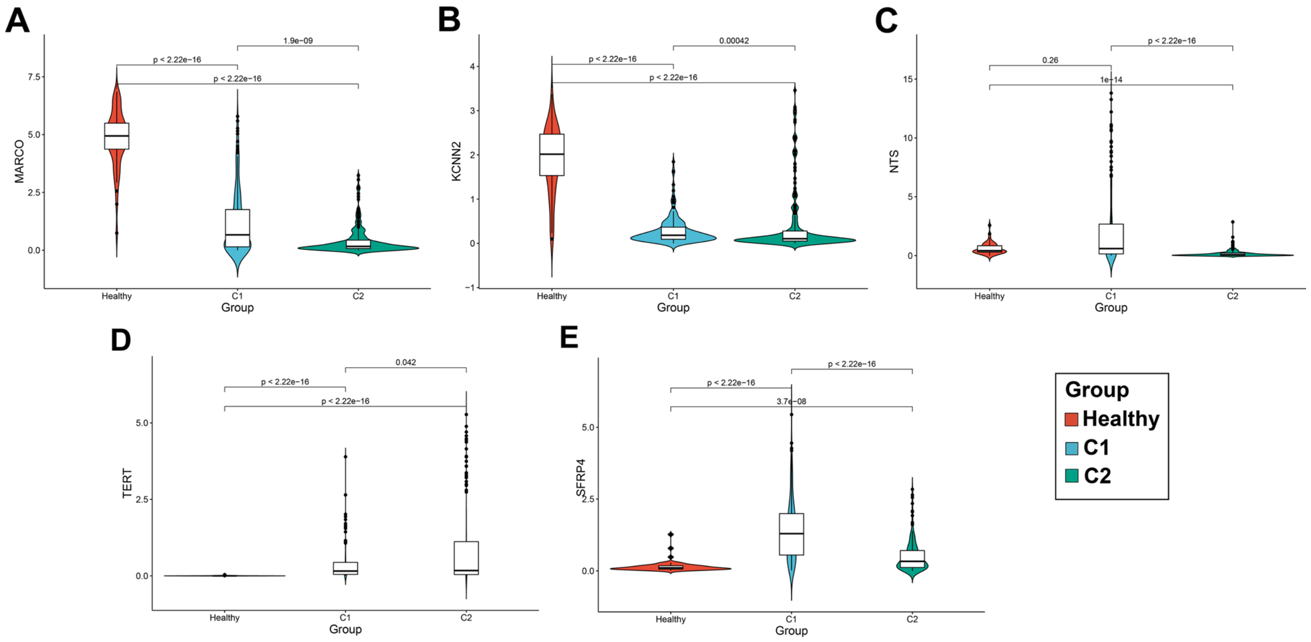

Fig. 6 Expression of LC-HGs in the healthy group, C1 and C2 subgroups

3.6 Determination of molecular subtypes, survival analysis and immune infiltration

We identified molecular subtypes of LC based on the transcriptional patterns of LC-HGs. The k-means method of unsupervised consensus clustering was used to cluster the LC samples. After comprehensive consideration, K= 2 was determined to be the optimal number of clusters (Fig. 5A, B), which was used to divide the LC samples into two subtypes (Fig. 5C), defined as C1 and C2 subtypes. The boundaries between C1 and C2 were very clear, indicating the reliability of clustering of LC samples. The results showed that the expression levels of five LC-HGs in C1 and C2 were significantly different (Fig. 6). The expression levels of *MARCO*, *NTS*, and *SFRP4* were higher in group C1, and the expression levels of *KCNN2* and *TERT* were higher in group C2.

Of course, survival analysis between different subtypes needs to be checked. The results of survival analysis showed that different subtypes (C1 and C2) identified according to the transcriptome expression pattern of LC-HGs had significant differences in survival, among which C2 had a better survival status than C1 (Fig. 5D). Since the survival difference of liver cancer patients is related to the degree of immune cell invasion, it is necessary to explore the difference of immune invasion between C1 and C2 to determine the degree of influence of specific cell types on the survival status of liver cancer patients. We used the ssGSEA algorithm to perform immunoinfiltration analysis in LC patients included in this study. The results showed that cell the C2 group had higher content of cells such as "Central memory CD8 T cells" ($P < 0.05$), and the C1 group had higher content of cells such as "CD56 dim natural killer cells", "Regulatory T cells", "Activated dendritic cell", and "Mast cell" ($P < 0.05$) (Fig. 5E), which may be contributing to the difference in survival between the C1 and C2 groups. Furthermore, to clarify the immune role of LC-HGs, we calculated the correlation between the abundance of immune cell infiltration and the expression of LC-HGs. The results showed that the expression of LC-HGs was significantly correlated with the abundance of these immune cells. For example, the expression of *MARCO*, *KCNN2*, *NTS*, and *SFRP4* was positively correlated with regulatory T cells and mast cells, while the expression of TERT was positively correlated with follicular helper T cells. Cells, CD56 dim natural killer cells and CD56bright natural killer cells were negatively correlated (Fig. 5F). This significant correlation indicates that dysregulated expression of LC-HGs may lead to the destruction of the immune microenvironment in LC patients, thereby exacerbating the inflammatory characteristics of LC patients.

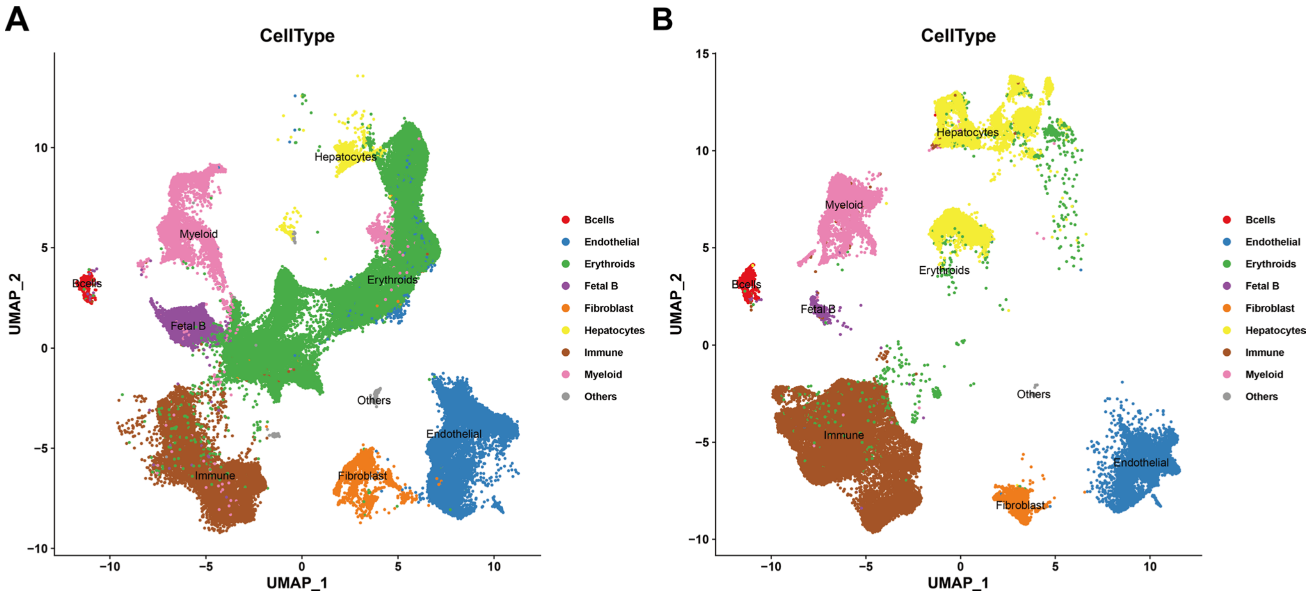

Fig. 7 Single-cell landscape of the liver in healthy and cancerous states. **A** Single-cell landscape of the liver in health. **B** Single-cell landscape of the liver in cancerous states

3.7 Validation at the single-cell level

We performed single-cell analysis on 10 normal samples (77,189 cells) and 9 liver cancer samples (57,254 cells), and we drew UMAP maps of healthy and liver cancer livers. By performing cell clustering on them, we classified 9 cell types including B cells, Endothelial cell, Erythroids cell, Fetal B cell, Fibroblast, Hepatocytes, Immune cell, Myeloid cell and other cells. As shown in the Fig. 7, there is a significant difference in the number of Fetal B cell and Immune cell.

Subsequently, we extracted the expression levels of these LC-HGs in each cell type, and analyzed the significant differences in the expression levels of LC-HGs in each cell type under healthy state and liver cancer state. The results showed that there were significant differences in the expression levels of *KCNN2* in healthy and LC Endothelial cells (Figure S6 A), and significant differences in the expression levels of *TERT* in healthy and LC Erythroids cells and Hepatocytes (Figure S6B). There were significant differences in the expression levels of *SFRP4* in healthy and LC Endothelial cells, Fibroblast and Hepatocytes (Figure S6 C). There are significant differences in the expression levels of *MARCO* in healthy and LC Endothelial cells, Erythroids cells, Fibroblast, Hepatocytes and Myeloid cells (Figure S6D). With the exception of B cells and Other cells, there were significant differences in the expression levels of *NTS* in both healthy and LC remaining 7 cell types (Figure S6E). This significant correlation suggests that dysregulation of LC-HGs expression in some specific cell types may induce the transformation of the liver from a healthy state to a liver cancer state, which may cause some cells to become cancerous or lose their regulatory role.

3.8 Clinical validation of LC-HGs

Although the subcellular localization and expression patterns of LC-HGs at the mRNA level have been investigated, information about their localization and expression at the protein level remains to be elucidated. To determine the expression difference of LC-HGs proteins, IHC staining images of LC-HGs proteins in LC tissues and normal liver tissues were obtained from the HPA database. The results showed that the protein expression levels of *MARCO*, *KCNN2*, and *NTS* in normal liver tissues were higher than those in LC, while the protein expressions of *TERT* and *SFRP4* were lower than those in LC (Fig. 8).

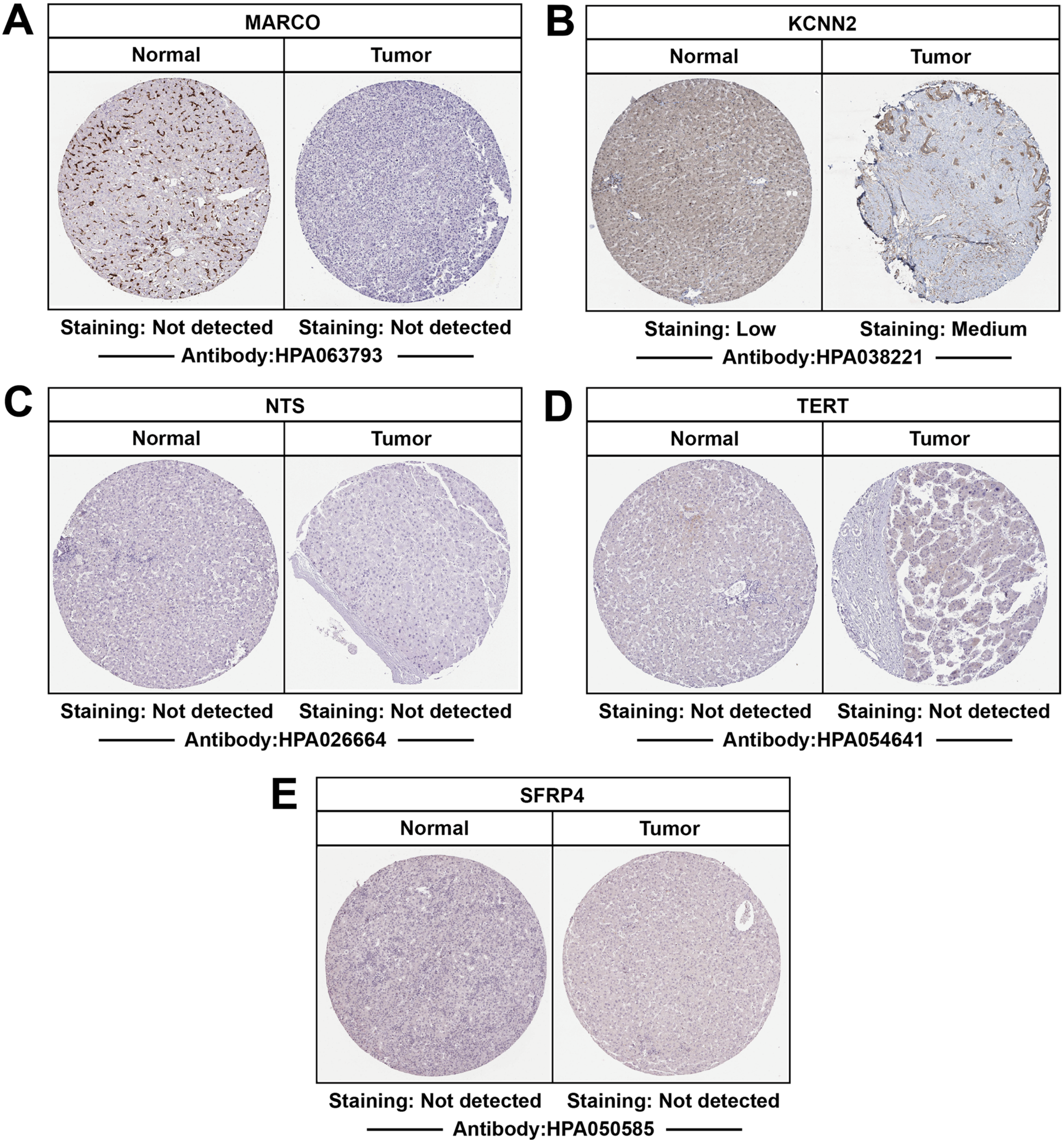

Fig. 8 Validation of LC-HGs Regulation of LC. **A**–**G** Protein expression of the indicated LC-HGs in LC tumor and normal tissues using clinical specimens from human proteome profiling

3.9 Expression of LC-HGs in the validation set

We observed the expression of LC-HGs extracted in the validation set, and the validation of the new liver cancer population showed that the expression of the identified LC hub genes was consistent with that of the original cohort (Figure S7).

4 Discussion

LC remains a global health challenge, with more than 1 million people expected to develop the disease each year by 2025 [37]. Clinically, tumors are often difficult to detect through routine screening when patients are asymptomatic. Advanced LC may present with symptoms such as abdominal pain, weight loss, a right upper abdominal mass, and unexplained disease progression, and fever may also occur [38]. In some cases, initial symptoms may include bloody ascites, shock, or peritonitis due to tumor bleeding. The high mortality rate of LC is attributed to the lack of early diagnostic methods and effective treatments, as well as the complex molecular mechanisms of liver cancer [39]. This study aimed to investigate the potential of LC-HGs as a biomarker for liver cancer to enhance clinical diagnosis and treatment.

This study utilizes extensive RNA-seq and single-cell data from LC. We first identified 1629 genes significantly different in LC through differential analysis. Weighted gene co-expression network analysis then revealed 1263 turquoise module genes associated with LC. A total of 224 LC DEGs were identified through intersection analysis, highlighting their interconnected roles and significant involvement in LC development. To validate these genes, we performed functional enrichment analysis, which revealed strong associations with pathways and processes such as "antimicrobial humoral response," "genitalia development," "ventral spinal cord development," "extracellular matrix disassembly," "calcium signaling pathway," "neuroactive ligand-receptor interaction," "ascorbate and aldarate metabolism," and "melanoma," among others, supporting the reliability of our findings. Furthermore, three machine learning methods were employed to identify LC-HGs, with all methods consistently recognizing *MARCO*, *KCNN2*, *NTS*, *TERT* and *SFRP4*.

In this study, we identified and validated the post-transcriptional regulation of LC by *MARCO*, *KCNN2*, *NTS*, *TERT*, and *SFRP4*. Studies have shown that when *MARCO* is upregulated in liver cancer, it can inhibit tumor cell migration and invasion. *MARCO* overexpression can promote liver cancer cell apoptosis and reduce proliferation in vitro and in vivo. Gene set enrichment analysis (GSEA) showed that *MARCO* may be related to the P53 signaling pathway [40]. *KCNN2* expression is significantly reduced in LC tissues, suggesting that circ*KCNN2* from the *KCNN2* gene may act as a tumor suppressor. The expression of *KCNN2* and circ*KCNN2* is regulated by the transcription factor NFYA, and high NFYA levels are associated with an increased risk of recurrence after LC surgery [41]. NFYA is known to promote various cancers. It promotes LC by increasing tumor-specific transcripts of lin-28 homolog B (ILN28) and inhibiting tumor suppressor function by recruiting transcriptional repressors [42, 43]. In addition, demethylation of the *KCNN2* promoter promotes NFYA binding, resulting in reduced *KCNN2* expression, thereby promoting LC progression [41]. Studies have found that *NTS* stimulation and increased NTR1 expression promote tumor invasion in LC by accelerating epithelial-mesenchymal transition (EMT) [44]. *NTS*-induced EMT involves upregulation of Wnt1, Wnt3, Wnt5, Axin, and p-GSK3β. The *NTR1* antagonist SR48692 also reduced the metastasis of LC xenografts overexpressing NTR1 in vivo, indicating that *NTS* is a key driver of LC invasion and metastasis through the Wnt/β-catenin pathway [45]. Therefore, *NTS* is a potential therapeutic target for preventing HCC progression. *TERT* promoter mutation is one of the most common genetic alterations in LC, highlighting that *TERT* upregulation is a key event in LC development [46].

The specific expression of these key genes in different cell types may affect the tumor microenvironment and promote the malignant progression of liver cancer by regulating the interaction between immune cells and stromal cells. *MARCO* is mainly highly expressed in M2 tumor-associated macrophages (TAMs), which may inhibit anti-tumor immunity by enhancing the secretion of immunosuppressive cytokines (such as IL-10 and TGF-β), while promoting the activation of hepatic stellate cells (HSCs), enhancing the fibrosis process, and providing a more suitable growth environment for tumor cells [47]. *KCNN2* plays an important role in T cells and endothelial cells. Its low expression in CD8 + T cells may weaken the killing ability of cytotoxic T cells, while its abnormal expression in endothelial cells affects tumor angiogenesis and further changes the oxygen and nutrient supply of the microenvironment [48]. The upregulation of *NTS* and its receptors in tumor cells and fibroblasts enhances the migration ability of tumor cells by stimulating pro-inflammatory signaling pathways, and regulates the secretion of chemokines by stromal cells to affect the recruitment of immune cells [49]. As a key factor in telomerase activation, *TERT* not only promotes cell proliferation in tumor cells, but also may reduce the immune system's ability to recognize and eliminate tumors by regulating the expression of immune checkpoint molecules (such as PD-L1) [50]. *SFRP4* is an antagonist of the Wnt signaling pathway. Its expression changes in fibroblasts and immune cells affect the plasticity of the tumor microenvironment, such as regulating the remodeling of matrix components and affecting the state of immune infiltration

[51]. Overall, the cell-specific expression differences of these genes may promote the immune escape, invasion and progression of liver cancer by affecting the interaction between immune cells and stromal cells.

The diagnostic potential of the identified LC-HGs was validated using an ANN [52]. Recent studies have demonstrated that ANN technology excels in diagnosing, predicting, and treating LC [53]. ANN models generally exhibit high accuracy and AUC values, with some achieving 100% accuracy [41, 54]. Despite the advancements, current clinical diagnostic methods for LC, particularly at the molecular level, are often limited and dependent on clinical experience [55]. We developed an ANN model using LC-HGs to predict LC occurrence. The model achieved prediction accuracies of 1.000 for the training set and 0.960 for the test set, with predictive abilities of 1.000 and 0.875, respectively. This ANN model shows promise as an independent diagnostic tool for LC.

There were significant differences in survival between group C1 and group C2, among which C2 had a better survival status than C1, which was caused by the difference in immune cell abundance, which was caused by the expression of LC-HGs. *MARCO*, *KCNN2*, *NTS* and *SFRP4* genes showed a positive correlation trend for most immune cell types. However, *TERT* showed a negative correlation with most immune cell types, indicating that LC-HGs co-regulated the immune microenvironment in liver cancer patie*NTS*. The higher the expression level of *MARCO*, *KCNN2*, *NTS* and *SFRP4*, the lower the expression level of *TERT*, and the higher the abundance of immune cells, indicating the higher the risk and degree of LC, and the more serious the degree of liver damage. Previous studies have also shown that these genes can promote and suppress cancer through different transcription factors [40; 42; 43], which may provide a new target for the diagnosis and treatment of LC. The results at the single cell level were consistent with those at the transcriptome level, and also showed significant differences between different cell types in liver cancer and healthy state, indicating that LC-HGs closely regulates different cell types. In some specific cell types, the dysregulation of LC-HGs expression may induce the liver to change from healthy state to liver cancer state. This can cause some cells to become cancerous or lose their regulatory role [56, 57].

The potential molecular markers of LC identified in this study have important clinical application value. The ANN model based on key genes can be used for risk assessment and early screening of liver cancer. Combined with RNA sequencing data, the risk of liver cancer can be calculated for individuals. For high-risk individuals need to strengthen follow-up and need regular screening (ultrasound examination every 6 months and more frequent imaging examinations) [58] and develop personalized treatment plans combining traditional Chinese and Western medicine [59–61]. In addition, this model can be used as a supplement to alpha-fetoprotein testing to improve the sensitivity and specificity of early diagnosis, thereby improving the early detection rate of high-risk individuals [62–64].

In terms of assisting clinical decision-making, the expression patterns of these genes may be closely related to the aggressiveness and prognosis of liver cancer. For example, high expression of *MARCO*, *KCNN2* and *TERT* may indicate a more aggressive subtype, which can be used for preoperative evaluation to help clinicians develop more accurate surgical or local ablation treatment plans [42]. In addition, combined with survival analysis, the model can also be used to assess the risk of postoperative recurrence, thereby optimizing follow-up strategies, such as arranging more frequent imaging examinations for high-risk patients to improve the effect of early intervention [65].

In addition, immune infiltration analysis in this study showed that these genes were associated with different types of immune cells, suggesting their potential role in immunotherapy and personalized treatment strategies. For example, *SFRP4* may affect the tumor microenvironment [51] and has predictive value for the efficacy of immune checkpoint inhibitors, providing new ideas for optimizing immunotherapy strategies for liver cancer.

TCGA (The Cancer Genome Atlas) database provided by UCSC Xena, covers transcriptome data of large-scale liver cancer patients and can be used to explore the expression patterns of key genes [66, 67]. Secondly, the Mendeley database contains single-cell RNA sequencing datasets shared by multiple published studies [68], used to analyze gene expression differences in different cell types to improve the biological credibility of the results. Both databases are widely used in multiple bioinformatics studies, and the data are publicly available, facilitating repeated verification and subsequent research [62, 63, 66]. Validation using an independent GEO database set confirmed that the expression patterns of the identified LC hub genes were consistent with those in the primary liver cancer cohort. This reliable performance in an independent population underscores the potential of LC-HGs.

The present study has several limitations that warrant acknowledgment. First, although the analysis incorporated both training cohorts and an independent validation set to enhance reliability, the sample sizes remain constrained due to the preliminary nature of this research phase. Further validation through multicenter large-scale cohorts and comprehensive in vivo/in vitro experiments is imperative to strengthen these findings. Second, the analytical framework focused exclusively on transcriptomic profiles, omitting integrative multi-omics data (e.g., epigenetics, proteomics, metabolomics), which may introduce incomplete interpretations of regulatory network dynamics. Additionally, while the newly developed ANN model demonstrates preliminary predictive utility, its clinical translation requires optimization through expanded sample sizes and rigorous multicenter trials to assess generalizability and robustness. Addressing these limitations will constitute a primary focus of subsequent investigations to advance the translational potential of this work.

5 Conclusions

In this study, we performed a series of bioinformatics analyses based on LC-associated transcriptomic data and single-cell data. Specifically, we identified 224 LC-associated DEGs through differential analysis and WGCNA, and enrichment analysis showed that these genes were associated with the development of liver cancer. Next, we used multiple machine learning algorithms (RF, Lasso, and SVM-RFE) to jointly identify *MARCO*, *KCNN2*, *NTS*, *TERT*, and *SFRP4* as key genes in hepatocellular carcinoma development, defined as LC-HGs, the biological mechanism of LC-HGs was explored by enrichment analysis. At the same time, we construct a new artificial neural network model for LC diagnosis, and the ROC curve of this model shows good performance. In addition, ssGSEA results showed that LC-HGs can modulate the immune microenvironment in patients with LC. Finally, we verified at the single-cell level that LC-HGs indeed regulate the transformation of specific cell types, verified the reliability of LC-HGs with a new cohort, and confirmed the expression of LC-HGs at the protein level by immunohistochemistry. This study verified its importance in LC-HG from multiple dimensions and provided a promising target for the clinical diagnosis and treatment of LC.

Author contributions Y.G.: Conceptualization, Software, Data curation, Formal Analysis, Writing–original draft. J.C.: Investigation, Methodology, Validation, Visualization, Writing–original draft. W.D.: Funding acquisition, Resources, Supervision, Writing–review and editing.

Funding There was no project support for this study.

Data availability Datasets related to this article are from public UCSC Xena database(https://xena.ucsc.edu/), Mendeley data(https://data.mendeley.com/datasets/6wmzcskt6k/1), GEO data(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE272166) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE265834) and HPA data(https://www.proteinatlas.org/). All data generated or analyzed during this study are included in this article/Additional files, further inquiries can be directed to the corresponding author.

Declarations

Ethics approval and consent to participate Not applicable.

Competing interests The authors declare no competing interests.

References

1. Llovet JM, Castet F, Heikenwalder M, Maini MK, Mazzaferro V, et al. Immunotherapies for hepatocellular carcinoma. Nat Rev Clin Oncol. 2022;19(3):151–72.
2. Kudo M, Finn RS, Qin S, Han KH, Ikeda K, et al. Lenvatinib versus sorafenib in first-line treatment of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial. The Lancet. 2018;391(10126):1163–73.
3. Akinyemiju T, Abera S, Ahmed M, Alam N, Alemayohu MA, et al. The burden of primary liver cancer and underlying etiologies from 1990 to 2015 at the global, regional, and national level: results from the global burden of disease study 2015. JAMA Oncol. 2017;3(12):1683–91.
4. Sanyal A, Poklepovic A, Moyneur E, Barghout V. Population-based risk factors and resource utilization for HCC: US perspective. Curr Med Res Opin. 2010;26(9):2183–91.

5. Boyault S, Rickman DS, De Reyniès A, Balabaud C, Rebouissou S, et al. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. Hepatology. 2007;45(1):42–52.
6. Zucman-Rossi J, Villanueva A, Nault JC, Llovet JM. Genetic landscape and biomarkers of hepatocellular carcinoma. Gastroenterology. 2015;149(5):1226–39.
7. Llovet JM, Zucman-Rossi J, Pikarsky E, Sangro B, Schwartz M, et al. Hepatocellular carcinoma (Primer). Nat Rev Dis Primers. 2016. https://doi.org/10.1038/nrdp.2016.18.
8. Vij M, Calderaro J. Pathologic and molecular features of hepatocellular carcinoma: An update. World J Hepatol. 2021;13(4):393.
9. Nault JC, Calderaro J, Di Tommaso L, Balabaud C, Zafrani ES, et al. Telomerase reverse transcriptase promoter mutation is an early somatic genetic alteration in the transformation of premalignant nodules in hepatocellular carcinoma on cirrhosis. Hepatology. 2014;60(6):1983–92.
10. Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? The lancet. 2001;357(9255):539–45.
11. De Visser KE, Eichten A, Coussens LM. Paradoxical roles of the immune system during cancer development. Nat Rev Cancer. 2006;6(1):24–37.
12. Behary J, Amorim N, Jiang XT, Raposo A, Gong L, et al. Gut microbiota impact on the peripheral immune response in non-alcoholic fatty liver disease related hepatocellular carcinoma. Nat Commun. 2021;12(1):187.
13. Guichard C, Amaddeo G, Imbeaud S, Ladeiro Y, Pelletier L, et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat Genet. 2012;44(6):694–8.
14. Totoki Y, Tatsuno K, Covington KR, Ueda H, Creighton CJ, et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. Nat Genet. 2014;46(12):1267–73.
15. Melero I, Yau T, Kang YK, Kim TY, Santoro A, et al. Nivolumab plus ipilimumab combination therapy in patients with advanced hepatocellular carcinoma previously treated with sorafenib: 5-year results from CheckMate 040. Ann Oncol. 2024;35(6):537–48.
16. Zhu AX, Abbas AR, De Galarreta MR, Guan Y, Lu S, et al. Molecular correlates of clinical response and resistance to atezolizumab in combination with bevacizumab in advanced hepatocellular carcinoma. Nat Med. 2022;28(8):1599–611.
17. Kanwal F, Kramer JR, Mapakshi S, Natarajan Y, Chayanupatkul M, et al. Risk of hepatocellular cancer in patients with non-alcoholic fatty liver disease. Gastroenterology. 2018;155(6):1828–37.
18. Sharma A, Seow JJW, Dutertre CA, Pai R, Blériot C, et al. Onco-fetal reprogramming of endothelial cells drives immunosuppressive macrophages in hepatocellular carcinoma. Cell. 2020;183(2):377–94.
19. Ritchie ME, Phipson B, Wu DI, Hu Y, Law CW, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47–e47.
20. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:1–13.
21. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607–13.
22. Chandrashekar G, Sahin F. A survey on feature selection methods. Comput electr eng. 2014;40(1):16–28.
23. Basu S, Kumbier K, Brown JB, Yu B. Iterative random forests to discover predictive and stable high-order interactions. Proc Natl Acad Sci USA. 2018;115(8):1943–8.
24. Hepp T, Schmid M, Gefeller O, Waldmann E, Mayr A. Approaches to regularized regression–a comparison between gradient boosting and the lasso. Methods Inf Med. 2016;55(05):422–30.
25. Noble WS. What is a support vector machine? Nat Biotechnol. 2006;24(12):1565–7.
26. Paul A, Mukherjee DP, Das P, Gangopadhyay A, Chintha AR, Kundu S. Improved random forest for classification. IEEE Trans Image Process. 2018;27(8):4012–24.
27. Vasquez MM, Hu C, Roe DJ, Chen Z, Halonen M, Guerra S. Least absolute shrinkage and selection operator type methods for the identification of serum biomarkers of overweight and obesity: simulation and application. BMC Med Res Methodol. 2016;16:1–19.
28. Beck MW. NeuralNetTools: visualization and analysis tools for neural networks. J Stat Softw. 2018;85(11):1.
29. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:1–8.
30. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26(12):1572–3.
31. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:1–15.
32. Morgan, M., Falcon, S., Gentleman, R., Maintainer, M. B. P., AnnotationDbi, et al. (2013). Package 'GSEABase'.
33. Gribov A, Sill M, Lück S, Rücker F, Döhner K, et al. SEURAT: visual analytics for the integrated analysis of microarray data. BMC Med Genomics. 2010;3:1–6.
34. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16(12):1289–96.
35. Digre A, Lindskog C. The human protein atlas—spatial localization of the human proteome in health and disease. Protein Sci. 2021;30(1):218–33.
36. St L, Wold S. Analysis of variance (ANOVA). Chemometr Intell Lab Syst. 1989;6(4):259–72.
37. Asrani SK, Devarbhavi H, Eaton J, Kamath PS. Burden of liver diseases in the world. J Hepatol. 2019;70(1):151–71.
38. Attwa MH, El-Etreby SA. Guide for diagnosis and treatment of hepatocellular carcinoma. World J Hepatol. 2015;7(12):1632.
39. Aziz H, Kwon YIC, Park A, Kwon Y, Aswani Y, Pawlik TM. Comprehensive review of clinical presentation, diagnosis, management, and prognosis of ruptured hepatocellular carcinoma. J Gastrointest Surg. 2024. https://doi.org/10.1016/j.gassur.2024.05.018.
40. Dong Q, Zhang S, Zhang H, Sun J, Lu J, Wang G, Wang X. MARCO is a potential prognostic and immunotherapy biomarker. Int Immunopharmacol. 2023;116: 109783.
41. Liu D, Liu W, Chen X, Yin J, Ma L, et al. circKCNN2 suppresses the recurrence of hepatocellular carcinoma at least partially via regulating miR-520c-3p/methyl-DNA-binding domain protein 2 axis. Clin Transl Med. 2022;12(1): e662.
42. Wang H, Ma Z, Xu M, Xiong M, Chen X, et al. Coptisine-mediated downregulation of E2F7 induces G2/M phase arrest in hepatocellular carcinoma cells through inhibition of E2F4/NFYA/NFYB transcription factors. Chem Biol Interact. 2024;397: 111063.

43. Wang Y, Weng H, Zhang Y, Long Y, Li Y, et al. The PRR11-SKA2 bidirectional transcription unit is negatively regulated by p53 through NF-Y in lung cancer cells. Int J Mol Sci. 2017;18(3):534.
44. Xiao P, Long X, Zhang L, Ye Y, Guo J, et al. Neurotensin/IL-8 pathway orchestrates local inflammatory response and tumor invasion by inducing M2 polarization of tumor-associated macrophages and epithelial-mesenchymal transition of hepatocellular carcinoma cells. Oncoimmunology. 2018;7(7): e1440166.
45. Samaržija I. Wnt signaling pathway is among the drivers of liver metastasis. Livers. 2021;1(4):180–200.
46. Amisaki M, Tsuchiya H, Sakabe T, Fujiwara Y, Shiota G. Identification of genes involved in the regulation of TERT in hepatocellular carcinoma. Cancer Sci. 2019;110(2):550–60.
47. Cheng K, Cai N, Zhu J, Yang X, Liang H, Zhang W. Tumor-associated macrophages in liver cancer: from mechanisms to therapy. Cancer Commun. 2022;42(11):1112–40.
48. Lin X, Wu JF, Wang DM, Zhang J, Zhang WJ, Xue G. The correlation and role analysis of KCNK2/4/5/15 in human papillary thyroid carcinoma microenvironment. J Cancer. 2020;11(17):5162.
49. Chen Z, Fang Y, Jiang W. Important cells and factors from tumor microenvironment participated in perineural invasion. Cancers. 2023;15(5):1360.
50. Dratwa M, Wysoczańska B, Łacina P, Kubik T, Bogunia-Kubik K. TERT—regulation and roles in cancer formation. Front Immunol. 2020;11: 589929.
51. Danieau G, Morice S, Rédini F, et al. New insights about the Wnt/β-catenin signaling pathway in primary bone tumors and their microenvironment: a promising target to develop therapeutic strategies? Int J Mol Sci. 2019;20(15):3751.
52. Mandair D, Reis-Filho JS, Ashworth A. Biological insights and novel biomarker discovery through deep learning approaches in breast cancer histopathology. NPJ Breast Cancer. 2023;9(1):21.
53. Kufel J, Bargieł-Łączek K, Kocot S, Koźlik M, Bartnikowska W, et al. What is machine learning, artificial neural networks and deep learning?—Examples of practical applications in medicine. Diagnostics. 2023;13(15):2582.
54. Popovic D, Glisic T, Milosavljevic T, Panic N, Marjanovic-Haljilji M, et al. The importance of artificial intelligence in upper gastrointestinal endoscopy. Diagnostics. 2023;13(18):2862.
55. Voelker R. What are diabetic foot ulcers? JAMA. 2023;330:2314.
56. Li X, Ramadori P, Pfister D, Seehawer M, Zender L, Heikenwalder M. The immunological and metabolic landscape in primary and metastatic liver cancer. Nat Rev Cancer. 2021;21(9):541–57.
57. Satriano L, Lewinska M, Rodrigues PM, Banales JM, Andersen JB. Metabolic rearrangements in primary liver cancers: cause and consequences. Nat Rev Dis Primers. 2019;16(12):748–66.
58. Warner E, Plewes DB, Hill KA, Causer PA, Zubovits JT, et al. Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. JAMA. 2004;292(11):1317–25.
59. Gao S, Wang W, Li J, Wang Y, Shan Y, Tan H. Unveiling polysaccharides of Houttuynia cordata Thunb.: extraction, purification, structure, bioactivities, and structure–activity relationships. Phytomedicine. 2025;28:156436.
60. Gao S, Li J, Wang W, Wang Y, Shan Y, Tan H. Rabdosia rubescens (Hemsl.) H. Hara: A Potent Anti-Tumor Herbal Remedy-Botany, Phytochemistry, and Clinical Breakthroughs. J Ethnopharmacol. 2024;3:119200.
61. Gao S, Shan Y, Wang Y, Wang W, Li J, Tan H. Polysaccharides from Lonicera japonica Thunb.: Extraction, purification, structural features and biological activities—A review. Int J Biol Macromol. 2024;15:136472.
62. Tzec-Interián JA, González-Padilla D, Góngora-Castillo EB. Bioinformatics perspectives on transcriptomics: a comprehensive review of bulk and single-cell RNA sequencing analyses. Quantitative Biology. 2025;13(2): e78.
63. Thind AS, Monga I, Thakur PK, Kumari P, Dindhoria K, Krzak M, Ashford B. Demystifying emerging bulk RNA-Seq applications: the application and utility of bioinformatic methodology. Brief Bioinform. 2021;22(6):259.
64. Gao S, Gang J, Yu M, Xin G, Tan H. Computational analysis for identification of early diagnostic biomarkers and prognostic biomarkers of liver cancer based on GEO and TCGA databases and studies on pathways and biological functions affecting the survival time of liver cancer. BMC Cancer. 2021;21:1–15.
65. Tang J, Li S, Zhou Z, Wang Y, Ni D, Zhou S. MiR-3680-3p is a novel biomarker for the diagnosis and prognosis of liver cancer and is involved in regulating the progression of liver cancer. IUBMB Life. 2024;76(10):820–31.
66. Zhang JY, Zhu X, Liu Y, Wu X. The prognostic biomarker RAB7A promotes growth and metastasis of liver cancer cells by regulating glycolysis and YAP1 activation. J Cell Biochem. 2024;125(8): e30621.
67. Hakami ZH. Biomarker discovery and validation for gastrointestinal tumors: a comprehensive review of colorectal, gastric, and liver cancers. Pathol Res Pract. 2024;255: 155216.
68. Li X, Wang CY. From bulk, single-cell to spatial RNA sequencing. Int J Oral Sci. 2021;13(1):36.

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