

# Novel Prognostic Markers for Liver Hepatocellular Carcinoma and Their Roles in Immune Infiltration of the Tumor Microenvironment

Meihai Deng (✉ [dengmeih@mail.sysu.edu.cn](mailto:dengmeih@mail.sysu.edu.cn))

Third Affiliated Hospital of Sun Yat-Sen University <https://orcid.org/0000-0001-8495-3885>

**Hao Liang**

Department of General Surgery, The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

**Kunpeng Hu**

Department of General Surgery, The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

**Zhaozhong Zhong**

Department of Hepatobiliary surgery, The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

**Zhiyong Xiong**

Department of General Surgery, The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

**Yupeng Ren**

Department of Hepatobiliary surgery, The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

**Zhicheng Yao**

Department of General Surgery, The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

**Weiwei Deng**

Department of Dermatology and Venerology, Peking University First Hospital, Beijing, China. Research Center for Medical Mycology, Peking University, Beijing, China. Beijing Key Laboratory of Molecular Diagnosis of Dermatoses, Peking University First Hospital

---

## Research

**Keywords:** weighted gene co-expression network analysis, liver hepatocellular carcinoma, immune infiltrate, prognostic marker, TCGA

**Posted Date:** January 7th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-139862/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

# Abstract

**Background:** Liver hepatocellular carcinoma (LIHC) is one of the most common malignant cancers worldwide, the overall prognosis of LIHC remains unsatisfactory. Valuable prognostic biomarkers are still urgently needed for LIHC. This study aimed to explore hub genes associated with the prognosis of LIHC and tumor microenvironmental immune infiltration, providing potential prognostic biomarker and therapeutic target for LIHC.

**Methods:** RNA-seq counts data for LIHC samples were obtained from TCGA database. RNA-seq counts data for normal liver samples were obtained from GTEx database. Weighted gene co-expression network analysis (WGCNA) was used to cluster differentially expressed genes with similar expression profiles to form modules and significant modules and key genes were screened. Next, these genes were verified by cox analyses and overall survival analysis. Further, CIBERSORT was used to explore the relationship between these genes and tumor infiltrating immune cells.

**Results:** A total number of 2661 significant DEGs were included for consensus WGCNA analysis, which identified 6 modules. Blue module ( $r=0.85$ ,  $p<0.0001$ ) showed high relationship with LIHC, which included 400 genes. After the overall survival analyses of hub genes, *CDC20*, *CDCA5*, *CDCA8*, *KIF2C* and *KIFC1* were identified as five potential marker genes, which would result in an unfavorable prognosis in LIHC. Further CIBERSORT analysis showed these novel biomarkers expression levels in LIHC were positively correlated with activated memory CD4+ T cells, follicular helper T cells, regulatory T cells and macrophages M0. While, resting memory CD4+ T cells, monocytes, macrophages M2, resting mast cells showed a negative correlation with the 5 novel biomarkers expression levels.

**Conclusions:** The study screened 5 genes with marked prognostic capability for LIHC and found these genes were correlated with the infiltration of immune cells in LIHC tumor microenvironment. The findings might provide a more detailed molecular mechanism underlying LIHC occurrence and progression, holding promise for acting as potential biomarkers and therapeutic targets.

## Background

Liver hepatocellular carcinoma (LIHC) is one of the most common malignant cancers and is currently the third leading cause of cancer-related deaths worldwide [1]. The incidence of LIHC is increasing in regions that have conventionally been low incidence areas, such as North America and some European countries [2]. Although great advancements have been made in LIHC treatment such as liver transplantation and surgical resection [3], the overall prognosis of LIHC remains unsatisfactory, which is mainly attributed to its rapid progression, high recurrence rate and short overall survival (OS) time [4]. As a result, it is still important to identify new diagnostic and prognostic markers.

Normally, stroma maintains the tissue homeostasis and acts as a barrier toward tumor formation; however, when a cell starts to be cancerous, its surrounding matrix changes in a way to support cancer development [5, 6]. This modified stroma around the malignant cells is termed tumor microenvironment

(TEM) [7]. There are several components in the typical tumor microenvironment, including immune and inflammatory cells, stromal fibroblasts, endothelial cells, blood vessels, mesenchymal stem cells, neural cells, adipose cells, chemokines, cytokines and extracellular matrices [8]. There is a lot of evidence showing that LIHC initiation and progression benefit from its associated tumor territory [9–12]. Recent years, the field of cancer immunotherapy is moving fast because of the encouraging clinical results obtained of cytotoxic T lymphocyte-associated protein 4 (CTLA-4) and programmed cell death protein 1 (PD-1) inhibitors for the treatment of cancer [13–15]. LIHC is an inflammation-driven disease with potentially chronic liver inflammation and cirrhosis, and just has fewer chromosomal aberrations, suggesting a combination of immunological interventions may be more effective with conventional treatment of this disease [16]. In the tumor microenvironment, non-malignant cells can help tumor cells to proliferate, invade and metastasize [9]. Multiple immune cells coexist and interact in a complex series of pathways that ultimately lead to tumor carcinogenesis [10]. The immunosuppressive features of TEM not only play as one of the major roles inducing cancer progression but also a big challenge for effective immunotherapy. Therefore, it is critically important to get an in-depth understanding of its specific functions and mechanisms.

In recent years, with the development of high-throughput research technology and genome-wide microarray technology, the application of bioinformatics tools in biological and medical studies, including research regarding cancer early diagnosis, cancer grading and prognosis prediction, has increased remarkably [17]. Large numbers of changed genes can be detected by these technologies, and can be further analyzed to identify critical significant genes as novel biomarkers [18]. Weighted gene co-expression network analysis (WGCNA) is an important bioinformatics tool that can construct gene co-expression networks to explore the correlations among different gene clusters or the relationships between gene clusters and clinical features and detect hub genes that may serve as diagnostic, prognostic or therapeutic targets [19]. CIBERSORT algorithm, an online tool, can be used to analyze large-scale RNA mixtures and identify cell types in the bulk RNA-seq matrix [20]. Meanwhile, there are plenty of public free databases, such as TCGA and the GEO, from which abundant information can be excavated via various bioinformatics methods. These bioinformatic approaches and databases make it possible for us to achieve a system-level insight into the complex biological processes that underlie cancer. Although alpha-fetoprotein (AFP) and AFP mRNA have been used as potential prognosis biomarkers for LIHC [21], their applications have certain limitations due to they rely on significant tumor burden, and the evaluation of their value has been incomplete [22]. Therefore, there is an unmet need to screen and identify new prognostic markers for LIHC.

In the current study, we analyzed 226 normal liver samples from GTEx database and 371 LIHC samples from TCGA database to get differentially expressed genes (DEGs). And 2661 DEGs, including 1092 up-regulated genes and 1569 down-regulated genes, were selected. These DEGs were analyzed by gene ontology (GO) analysis, Kyoto encyclopedia of genes and genomes (KEGG) analysis and gene set enrichment analysis (GSEA). Then, we constructed a gene co-expression network based on WGCNA and identified 6 modules associated with clinical features of LIHC. Blue module showed high relationship with LIHC; the top 20 genes of the blue module were hub genes. The overall survival analyses of hub genes

revealed that five genes (*CDC20*, *CDCA5*, *CDCA8*, *KIF2C*, *KIFC1*) were associated with worse prognosis in LIHC prognosis. The unfavorably prognostic values of these 5 genes were also identified by cox analyses. Importantly, we also observed that the novel biomarkers were significantly upregulated in LIHC, thus providing highly reliable analytic results. In addition, the expression levels of these genes were found to have a strong correlation with infiltrating immune cells, suggesting that these hub genes may be through regulating the levels of immune infiltration in LIHC TEM to promote the occurrence and development of LIHC.

## Methods

### Microarray Data

RNA-seq counts data for LIHC samples were obtained from TCGA database. RNA-seq counts data for normal liver samples were obtained from GTEx database.

### DEGs Analysis

After using Limma package to normalize the gene expression matrix,  $|\log_2\text{Fold change}| \geq 2$  and  $p \leq 0.01$  were used for screening DEGs between LIHC samples and normal samples. The heatmap of all samples was generated by the pheatmap package, and the volcano map was produced by the ggplot2 package. GO analysis, a regular method in the annotation of large-scale functional enrichment studies, is normally classified into MF, BP, and CC categories [23]. The KEGG is a database resource for understanding high-level functions and utilities of the biological system [24]. GSEA is a computational method that functions to identify classes of genes that are overrepresented in a large set of genes that may have a connection with disease phenotypes [25]. ClusterProfiler package was used to analyze GO, KEGG and GSEA.

### Weighted Gene Co-Expression Network Analysis (WGCNA)

Weight co-expression network was performed in accordance to the protocol of WGCNA package in the R language. A gene coexpression similarity measure was first calculated and used to relate every pairwise gene-gene relationship. The adjacency matrix and topological overlap matrix (TOM) were then constructed utilizing a 'soft' power adjacency function [26]. "Gene modules" are groups of genes that have high topological overlap. Modules were generated using algorithms and blockwiseModules function of WGCNA. Then, the expression level of each module was calculated, which is also called module eigengene (ME). Finally, the most related genes with LIHC were identified in a module.

### Identification and Validation of Biomarkers

A key module, which including 20 hub genes was identified after weighted gene co-expression network analysis. The hub genes were validated using Kaplan-Meier Plotter database, and five novel biomarkers were finally screened. Then, translational levels between LIHC and normal specimens and survival analysis of these genes were validated by the ggstatsplot package. Also, univariate and multivariate cox analyses were conducted to explore the contributions of novel biomarkers.

## The Tumor Microenvironment Analyses

CIBERSORT is a versatile computational method for quantifying cell fractions from bulk tissue gene expression profiles (GEPs), can accurately estimate the immune composition of a tumor biopsy [27]. Using CIBERSORT to identify the infiltrating immune cell populations of the bulk RNA-seq samples from TCGA database. Kaplan-Meier curves were used to show the relationship between LIHC patients' overall survival and TEM immune infiltration. Log-rank test was used to test the relationship.

## Results

### Different Gene Expression from Data between TCGA and GTEx Was Analyzed

After data processing, a total number of 2661 significant DEGs, including 1092 up-regulated and 1569 down-regulated genes, were identified between normal liver samples and LIHC samples. The expression of DEGs shown in Fig. 1A by heatmap. Figure 1B showed the distribution of all the significantly different expressed genes on the two dimensions of  $-\log_{10}$  (false discovery rate, FDR) and  $\log_2$  (fold change, FC) through a volcano map. To test the biological function of the identified genes, information from differentially expressed genes was applied to GO and KEGG analysis. The GO results showed that leukocyte migration, organelle fission, nuclear division were significantly enriched in biological process (BP). For cellular component (CC), DEGs were particularly enriched in extracellular matrix, collagen-containing extracellular matrix and chromosome region, and so on. Besides, receptor regulator activity, receptor ligand activity and so on were significantly enriched in molecular function (MF) (Fig. 1C). KEGG analysis demonstrated that DEGs were particularly enriched in neuroactive ligand-receptor interaction, cytokine-cytokine receptor interaction, cell cycle, and so on (Fig. 1D). Moreover, Epstein-Barr virus infection, Human T-cell leukemia virus 1 infection and MicroRNAs in cancer were up-regulated while the cAMP signaling pathway, Neuroactive ligand-receptor interaction were down-regulated by KEGG-GSEA (Fig. 1E).

### Novel Biomarkers Relevant to LIHC Were Identified via WGCNA

Next, the gene co-expression networks were conducted on 2661 DEGs of all samples. As shown in Fig. 2A, softpower 6 was chosen to construct nets, and there were 6 modules were identified. Then, genes in the 6 color modules were continuously used to analyze the module-trait (LIHC and normal) co-expression similarity and adjacency. Blue module ( $r = 0.85$ ,  $p < 0.0001$ ) showed high relationship with LIHC, which included 400 genes (Fig. 2B). The results of the cluster (Fig. 2C) also revealed that blue module was closely relevant to LIHC. As shown in Fig. 2D the positive correlation between module membership (MM) and gene significance (GS) of each gene in blue module ( $r = 0.95$ ;  $p \ll 0.05$ ). The gene expression levels of blue module were shown in Fig. 2E.

### Validation of Novel Biomarkers

After the overall survival analyses of hub genes, we identified five potential marker genes, which would result in an unfavorable prognosis in LIHC (Fig. 3A). The expression levels of *CDC20*, *CDCA5*, *CDCA8*, *KIF2C* and *KIFC1* were all significantly upregulated in LIHC (Fig. 3B), indicating a potential role of these genes in LIHC occurrence and development. Next, a univariate cox proportional hazard regression analysis was conducted to further verify the association of the expression levels of these genes mentioned above with overall survival. As shown in Table 1, univariate cox analysis showed the high expression of *CDC20*, *CDCA5*, *CDCA8*, *KIF2C* and *KIFC1* would make a poor prognosis. Further multivariate cox proportional hazard regression analysis revealed that high *CDCA8* expression was related to inferior prognosis, which indicated that *CDCA8* could serve as an independent risk factor to predict the prognosis of LIHC. These results above further validated the potential of these genes to be taken as new biomarkers for prognosis.

Table 1  
Correlation between the hub genes and infiltrating immune cells and prognosis.

Factors	Univariate cox analysis		Multivariate cox analysis	
	HR (95% CI)	P	HR	P
<b>CDC20</b>	1.529(1.305–1.792)	0.000	1.302	0.372
<b>CDCA5</b>	1.555 (1.267–1.909)	0.000	0.667	0.138
<b>CDCA8</b>	1.715(1.410–2.087)	0.000	2.017	0.023
<b>KIF2C</b>	1.658(1.381–1.992)	0.000	1.182	0.682
<b>KIFC1</b>	1.465(1.227–1.751)	0.000	0.703	0.184
<b>Macrophages.M0</b>	19.023(5.353–67.606)	0.000	8.682	0.002
<b>Mast.cells.resting</b>	0.025(0.001–0.623)	0.025	0.378	0.566
<b>T.cells.follicular.helper</b>	92.329(0.039-216504.185)	0.253		

## The Expression Levels of Novel Biomarkers Were Significantly Correlated with Infiltrating Immune Cell in TEM

Since TEM has been revealed as a critical role in the initiation and progression of tumors, we decided to investigate whether the novel biomarkers have an influence in LIHC TEM. In order to calculate the relative proportions of the infiltrating immune cell in LIHC TEM, we first employed CIBERSORT to deconvolute the bulk RNA-seq samples. The results suggested that naive B cells, resting dendritic cells, macrophages M0, macrophages M1, macrophages M2, resting mast cells, monocytes, activated NK cells, plasma cells, resting memory CD4 + T cells, CD8 + T cells, follicular helper T cells, gamma delta T cells and regulatory T cells have participated in the immune infiltration of TEM in LIHC. And it is notably that macrophages M0,

macrophages M1, macrophages M2, resting memory CD4 + T cell, CD8 + T cells and regulatory T cells made up the majority of these infiltrating immune cells (Fig. 4). Next, Pearson's correlation analyses were conducted to identify the relationship between novel biomarkers and infiltrating immune cells. As showed in Fig. 5, the novel biomarkers expression levels in LIHC were positively correlated with activated memory CD4 + T cells, follicular helper T cells, regulatory T cells and macrophages M0. While, resting memory CD4 + T cells, monocytes, macrophages M2, resting mast cells showed a negative correlation with the 5 novel biomarkers expression levels. Moreover, Kaplan-Meier survival analyses revealed that high level infiltration of follicular helper T cells and macrophages M0 was related with poor prognosis, but less activated mast cells got a worse prognosis (Fig. 6A, B, D), which were consistent with their relationship with the 5 genes above (Figure S1-S4). Although infiltrating resting mast cells also showed positive relationship with patient prognosis, the P value was not statistically significant (Fig. 6C). Together, these results indicated that the novel biomarkers may involve in regulating the immune infiltration of LIHC TEM, further promoting the development of LIHC.

## Discussion

Among all cancers, Liver hepatocellular carcinoma is the fifth most frequently diagnosed cancer, ranking as the third leading cause of cancer-related death, seriously impacting human health [28]. Currently, the main LIHC treatment strategies include surgical resection, microwave ablation, radiofrequency ablation and transcatheter arterial chemoembolization (TACE) [29, 30]. Although surgical resection is believed to have a definitive curative effect for LIHC, the clinical outcome is still poor due to frequent recurrence and metastasis [31–33]. Additionally, most LIHC cases are detected in advanced stages with the invasion of major blood vessels, obvious extrahepatic metastases or poor liver function, making them unfit for surgical resection [34]. Better biomarkers for predictive and prognostic molecules for LIHC are still required.

In this study, bioinformatics and comprehensive analyses of multiple datasets were used to screen genes that proved to be novel prognosis factors for LIHC. Firstly, we compared the translational profiles between normal and cancerous samples to obtain DEGs. GO and KEGG enrichment analysis was performed to interpret the functions and pathways of DEGs. The enriched BPs included leukocyte migration, organelle fission, nuclear division, chromosome segregation, humoral immune response, and so on. These processes had strong relationship with LIHC progression. The processes of GO CCs and GO MFs are also typically representative features of LIHC progression. The enriched KEGG pathways included neuroactive ligand-receptor interaction, cytokine-cytokine receptor interaction, cell cycle, and so on, which are also associated with LIHC. Then, co-expression network analysis by WGCNA identified that blue module was significantly associated with LIHC traits. The genes in these modules were taken intersection with DEGs between pathology stages I and IV. Finally, *CDC20*, *CDCA5*, *CDCA8*, *KIF2C* and *KIFC1* were screened out. Using GEPIA based on the TCGA database, we proved the expression levels of *CDC20*, *CDCA5*, *CDCA8*, *KIF2C* and *KIFC1* were significantly up-regulated in LIHC. And the survival analysis showed that high expression of these genes was associated with poor prognosis. Moreover, univariate cox analysis showed that the high expression of *CDC20*, *CDCA5*, *CDCA8*, *KIF2C* and *KIFC1* would make a poor prognosis.

Further multivariate cox proportional hazard regression analysis revealed that high *CDCA8* expression was related to inferior prognosis, which indicated that *CDCA8* could serve as an independent risk factor to predict the prognosis of LIHC.

Strikingly, the 5 genes have been reported correlated with clinical outcomes of a huge number of solid tumors. Cell division cycle 20 homologue (*CDC20*), also called Fizzy, specifically activating the anaphase-promoting complex-cyclosome, promoting ubiquitination and proteolysis of cell-cycle-regulatory proteins, which is essential for anaphase-promoting complex activity, initiation of anaphase, and cyclin proteolysis during mitosis [35]. In recent years, mounting evidence has revealed that *CDC20* plays an oncogenic role in human tumorigenesis. Overexpression of *CDC20* was observed in a variety of human tumors, including pancreatic cancer [36], breast cancer [37], prostate cancer [38], lung cancer [39], colorectal cancer [40] and liver hepatocellular carcinoma [41]. Li et al. reported that overexpression of *CDC20* was observed in 68% liver hepatocellular carcinoma tissues compared to adjacent non-tumor liver tissues. Moreover, high levels of *CDC20* were positively correlated with gender, tumor differentiation, and TNM stage [41]. In our study, we found the overexpression of *CDC20* may result in an unfavorable prognosis and promote TEM immune infiltration in LIHC patients. Cell division cycle associated 5 (*CDCA5*), is an important element for the interaction between cohesin and chromatin in interphase. As reported, the expression of *CDCA5* was upregulated in LIHC tissues compared to paracancerous tissues and had a negative correlation with patient survival. Further study showed that *CDCA5* promotes oncogenesis by enhancing cell proliferation and inhibiting apoptosis via the AKT pathway in liver hepatocellular carcinoma, plays an important role in LIHC progression [42]. These results were consistent with our findings. The cell division cycle associated 8 (*CDCA8*) plays an important role in mitosis [43]. *CDCA8* is a putative oncogene that is up-regulated in many types of cancer tissues, such as lung cancer [44] and gastric cancer [45]. While, little is known about its role in liver hepatocellular carcinoma. Our study revealed that *CDCA8* was significantly upregulated in LIHC and linked to poor prognosis. Besides, *CDCA8* plays a vital role in regulating immune cell of TEM and could serve as an independent risk factor to predict the prognosis of LIHC. Kinesin family member 2C (*KIF2C*), the kinesin-like protein functioning as a microtubule-dependent molecular motor, participates in spindle assembly and microtubule disaggregation, thereby determinately regulating cell cycle during mitosis [46–49]. A recent study firstly demonstrated *KIF2C* was substantially higher expression in tumor tissues than adjacent nontumor tissues, upregulated *PCNA* and *CDC20* expression and was significantly involved in growth promoting pathways, fully confirming our results that it could serve as a prognostic indicator and confer a novel target for clinical treatment. Kinesin family member C1 (*KIFC1*), also known as *HSET*, is a minus end-directed motor protein [50], which is a critical role in centrosome clustering in cancer cells [51]. Fu x et al. investigated the expression of *KIFC1* in paired liver hepatocellular carcinoma tissues and adjacent non-cancerous tissues and found that the expression of *KIF2C* was upregulated and connected with a poor prognosis of LIHC [52]. Our analyses further revealed that the overexpression of *KIFC1* was a novel predictive marker in patients with LIHC and involved the regulation of immune cell in LIHC TEM.

The association of inflammation and cancer was firstly hypothesized by Rudolf Virchow observations in 1863 as chronic irritation theory. Virchow found that certain cancers are associated with inflammatory

macrophages [53]. In recent years, more and more studies have proved the important role inflammatory cells played in the development of cancer. Neutrophils, monocytes, lymphocytes, dendritic cells, eosinophils and mast cells are the commonly observed cells in tumor stroma although their count depends on cancer type [53, 54]. In this study, we illustrated the percentage distribution of immune cells in LIHC TEM, finding macrophages M0, macrophages M1, macrophages M2, resting memory CD4 + T cell and CD8 + T cells accounted for most of the immune cells. Remarkably, increasing number of studies showed that macrophages facilitate cell proliferation, angiogenesis, metastasis and invasion, which were important in the initiation, development and metastasis of primary hepatocarcinoma [55]. In the early stage of tumorigenesis, M1 macrophages eliminate the tumor cells as soldiers of adaptive immunity. However, in advanced stages, M1 macrophages replaced with M2-type. M2 macrophages suppress the adaptive immune system and promote the cancer proliferation, angiogenesis and extracellular matrix remodeling as well. Eventually, tumor cells escape from the immune barriers and invade [56, 57]. Additionally, CD4 + T cells and CD8 + T cells were confirmed to participate in immune recognition and escape of LIHC [58]. Actually, tumour infiltrating lymphocytes (TILs) form a large component in solid tumours, in an attempt by the host to mediate an antitumour reaction [59]. However, this cellular response can be dysfunctional with a higher proportion of CD4+ (helper or T regulatory cells) to CD8 + cells. This promotes immune tolerance and has been shown to confer a worse prognosis [60]. E.A. Said et al found that the binding of PD-L1 (expressed on other cells) to its receptor PD-1 on macrophages promotes IL-10 release and thereby CD4 + T cell repression [61]. Effector CD8 + cells within LIHC tumors show intense PD-1 expression and the number of PD-1 + CD8 + cells was found to be related to disease progression and post-operative recurrence [62]. Since the existing treatment strategies for LIHC have great limitations, immunotherapy is a promising therapeutic option. Our results revealed the 5 novel biomarkers expression levels had a crucial association with parts of immune cells, which suggested these genes may involve in the level regulation of the immune cells of TEM in LIHC, may help to promote the application of immunotherapy in liver cancer.

## Conclusion

In conclusion, this study aimed to identify hub genes involved in LIHC by WGCNA of data from the TCGA database and we screened 5 genes with marked prognostic capability for LIHC. These genes are highly differentially expressed in tumor and non-tumor tissues. Further bioinformatic analysis showed that these genes can act as biomarkers of LIHC prognosis. Besides, these genes were correlated with the infiltration of immune cells in LIHC TEM. The findings might provide a more detailed molecular mechanism underlying LIHC occurrence and progression, holding promise for acting as potential biomarkers and therapeutic targets. However, further molecular experiments on these genes in LIHC, in vivo and in vitro, need to be carried out.

## Abbreviations

LIHC, Liver hepatocellular carcinoma; WGCNA, weighted gene co-expression network analysis; TEM, tumor microenvironment; *CDC20*, cell division cycle 20 homologue; *CDCA5*, cell division cycle associated 5; *CDCA8*, cell division cycle associated 8; *KIF2C*, kinesin family member 2C; *KIFC1*, kinesin family member C1; OS, overall survival, DEGs, differentially expressed genes; ME, module eigengene; BP, biological process; CC, cellular component; MF, molecular function; MM, module membership; GS, gene significance;

## Declarations

### Ethics approval and consent to participate

This study was an analysis of the third-party anonymized databases with preexisting IRB approval.

### Consent for publication

Not applicable.

### Availability of data and materials

The datasets supporting the conclusions of this article were available in the TCGA (<https://portal.gdc.cancer.gov>), GTEx database (<https://gtexportal.org/home/datasets>).

### Competing interests

The authors declare that they have no competing interests.

### Funding

This work was supported by National Natural Science Foundation of China (grant numbers 81702375); the Natural Science Foundation of Guangdong Province, China (grant numbers 2018A030313641, 2016A030313848); the Science and Technology Planning Project of Guangzhou city (grant numbers 201804010211).

### Authors' Contributions

MD, ZC and WD conceived the idea; WD and ZX performed manifold learning algorithm; HL, KP and ZZ performed analysis and drafted the manuscript; HL and YR performed data acquisition and figure preparations; MD and WD interpreted the results and help to revise the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

Not applicable

## References

1. Mortality, G.B.D.; Causes of Death, C. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **2016**, *388*, 1459-1544, doi:10.1016/S0140-6736(16)31012-1.
2. Kulik, L.; El-Serag, H.B. Epidemiology and Management of Hepatocellular Carcinoma. *Gastroenterology* **2019**, *156*, 477-491 e471, doi:10.1053/j.gastro.2018.08.065.
3. Yu, W.B.; Rao, A.; Vu, V.; Xu, L.; Rao, J.Y.; Wu, J.X. Management of centrally located hepatocellular carcinoma: Update 2016. *World J Hepatol* **2017**, *9*, 627-634, doi:10.4254/wjh.v9.i13.627.
4. Zheng, J.; Kuk, D.; Gonen, M.; Balachandran, V.P.; Kingham, T.P.; Allen, P.J.; D'Angelica, M.I.; Jarnagin, W.R.; DeMatteo, R.P. Actual 10-Year Survivors After Resection of Hepatocellular Carcinoma. *Ann Surg Oncol* **2017**, *24*, 1358-1366, doi:10.1245/s10434-016-5713-2.
5. Chen, F.; Zhuang, X.; Lin, L.; Yu, P.; Wang, Y.; Shi, Y.; Hu, G.; Sun, Y. New horizons in tumor microenvironment biology: challenges and opportunities. *BMC Med* **2015**, *13*, 45, doi:10.1186/s12916-015-0278-7.
6. Shimoda, M.; Mellody, K.T.; Orimo, A. Carcinoma-associated fibroblasts are a rate-limiting determinant for tumour progression. *Semin Cell Dev Biol* **2010**, *21*, 19-25, doi:10.1016/j.semcdb.2009.10.002.
7. Singh, S.R.; Rameshwar, P.; Siegel, P. Targeting tumor microenvironment in cancer therapy. *Cancer Lett* **2016**, *380*, 203-204, doi:10.1016/j.canlet.2016.04.009.
8. Shiao, S.L.; Chu, G.C.; Chung, L.W. Regulation of prostate cancer progression by the tumor microenvironment. *Cancer Lett* **2016**, *380*, 340-348, doi:10.1016/j.canlet.2015.12.022.
9. Lu, C.; Rong, D.; Zhang, B.; Zheng, W.; Wang, X.; Chen, Z.; Tang, W. Current perspectives on the immunosuppressive tumor microenvironment in hepatocellular carcinoma: challenges and opportunities. *Mol Cancer* **2019**, *18*, 130, doi:10.1186/s12943-019-1047-6.
10. Tahmasebi Birgani, M.; Carloni, V. Tumor Microenvironment, a Paradigm in Hepatocellular Carcinoma Progression and Therapy. *Int J Mol Sci* **2017**, *18*, doi:10.3390/ijms18020405.
11. Jiang, X.; Wang, J.; Deng, X.; Xiong, F.; Ge, J.; Xiang, B.; Wu, X.; Ma, J.; Zhou, M.; Li, X., et al. Role of the tumor microenvironment in PD-L1/PD-1-mediated tumor immune escape. *Mol Cancer* **2019**, *18*, 10, doi:10.1186/s12943-018-0928-4.
12. Liu, B.; Salgado, O.C.; Singh, S.; Hippen, K.L.; Maynard, J.C.; Burlingame, A.L.; Ball, L.E.; Blazar, B.R.; Farrar, M.A.; Hogquist, K.A., et al. The lineage stability and suppressive program of regulatory T cells require protein O-GlcNAcylation. *Nat Commun* **2019**, *10*, 354, doi:10.1038/s41467-019-08300-3.
13. Hodi, F.S.; O'Day, S.J.; McDermott, D.F.; Weber, R.W.; Sosman, J.A.; Haanen, J.B.; Gonzalez, R.; Robert, C.; Schadendorf, D.; Hassel, J.C., et al. Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med* **2010**, *363*, 711-723, doi:10.1056/NEJMoa1003466.

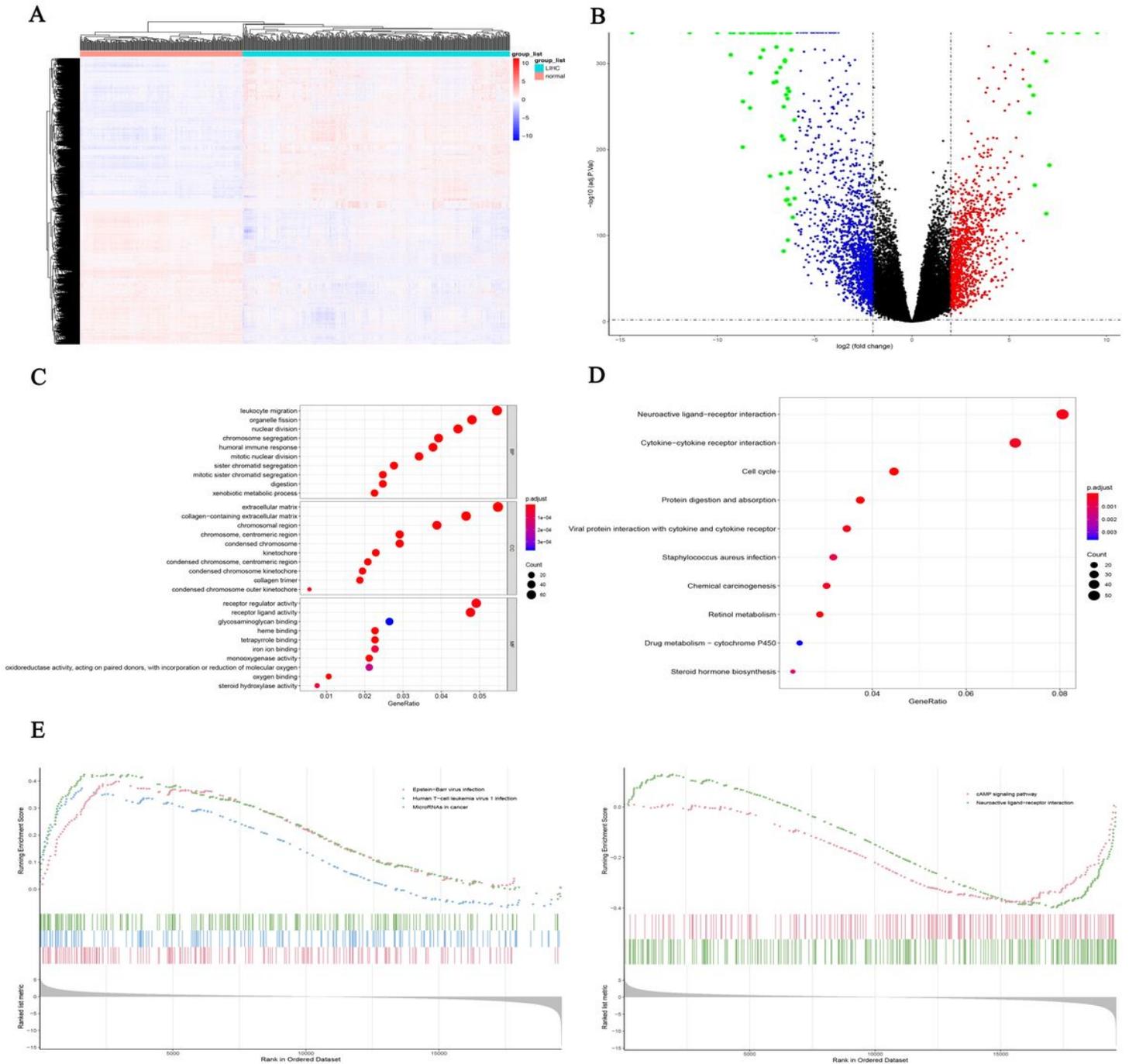
14. Powles, T.; Eder, J.P.; Fine, G.D.; Braiteh, F.S.; Loria, Y.; Cruz, C.; Bellmunt, J.; Burris, H.A.; Petrylak, D.P.; Teng, S.L., et al. MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. *Nature* **2014**, *515*, 558-562, doi:10.1038/nature13904.
15. Topalian, S.L.; Hodi, F.S.; Brahmer, J.R.; Gettinger, S.N.; Smith, D.C.; McDermott, D.F.; Powderly, J.D.; Carvajal, R.D.; Sosman, J.A.; Atkins, M.B., et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med* **2012**, *366*, 2443-2454, doi:10.1056/NEJMoa1200690.
16. Okusaka, T.; Ikeda, M. Immunotherapy for hepatocellular carcinoma: current status and future perspectives. *ESMO Open* **2018**, *3*, e000455, doi:10.1136/esmoopen-2018-000455.
17. Kulasingam, V.; Diamandis, E.P. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nat Clin Pract Oncol* **2008**, *5*, 588-599, doi:10.1038/ncponc1187.
18. Tang, J.; Kong, D.; Cui, Q.; Wang, K.; Zhang, D.; Gong, Y.; Wu, G. Prognostic Genes of Breast Cancer Identified by Gene Co-expression Network Analysis. *Front Oncol* **2018**, *8*, 374, doi:10.3389/fonc.2018.00374.
19. Langfelder, P.; Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **2008**, *9*, 559, doi:10.1186/1471-2105-9-559.
20. Newman, A.M.; Liu, C.L.; Green, M.R.; Gentles, A.J.; Feng, W.; Xu, Y.; Hoang, C.D.; Diehn, M.; Alizadeh, A.A. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **2015**, *12*, 453-457, doi:10.1038/nmeth.3337.
21. Hanazaki, K.; Kajikawa, S.; Koide, N.; Adachi, W.; Amano, J. Prognostic factors after hepatic resection for hepatocellular carcinoma with hepatitis C viral infection: univariate and multivariate analysis. *Am J Gastroenterol* **2001**, *96*, 1243-1250, doi:10.1111/j.1572-0241.2001.03634.x.
22. Tangkijvanich, P.; Anukulkarnkusol, N.; Suwangool, P.; Lertmaharit, S.; Hanvivatvong, O.; Kullavanijaya, P.; Poovorawan, Y. Clinical characteristics and prognosis of hepatocellular carcinoma: analysis based on serum alpha-fetoprotein levels. *J Clin Gastroenterol* **2000**, *31*, 302-308, doi:10.1097/00004836-200012000-00007.
23. Harris, M.A.; Clark, J.; Ireland, A.; Lomax, J.; Ashburner, M.; Foulger, R.; Eilbeck, K.; Lewis, S.; Marshall, B.; Mungall, C., et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **2004**, *32*, D258-261, doi:10.1093/nar/gkh036.
24. Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **2016**, *44*, D457-462, doi:10.1093/nar/gkv1070.
25. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S., et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **2005**, *102*, 15545-15550, doi:10.1073/pnas.0506580102.
26. Zhang, B.; Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **2005**, *4*, Article17, doi:10.2202/1544-6115.1128.

27. Chen, B.; Khodadoust, M.S.; Liu, C.L.; Newman, A.M.; Alizadeh, A.A. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol* **2018**, *1711*, 243-259, doi:10.1007/978-1-4939-7493-1\_12.
28. Hajighasemlou, S.; Pakzad, S.; Ai, J.; Muhammadnejad, S.; Mirmoghtadaei, M.; Hosseinzadeh, F.; Gharibzadeh, S.; Kamali, A.; Ahmadi, A.; Verdi, J. Characterization and Validation of Hepatocellular Carcinoma (HCC) Xenograft tumor as a Suitable Liver Cancer Model for Preclinical Mesenchymal Stem Cell Studies. *Asian Pac J Cancer Prev* **2018**, *19*, 1627-1631, doi:10.22034/APJCP.2018.19.6.1627.
29. Vietti Violi, N.; Duran, R.; Guiu, B.; Cercueil, J.P.; Aube, C.; Digkolia, A.; Pache, I.; Deltenre, P.; Knebel, J.F.; Denys, A. Efficacy of microwave ablation versus radiofrequency ablation for the treatment of hepatocellular carcinoma in patients with chronic liver disease: a randomised controlled phase 2 trial. *Lancet Gastroenterol Hepatol* **2018**, *3*, 317-325, doi:10.1016/S2468-1253(18)30029-3.
30. Yamashita, T.; Kaneko, S. Treatment strategies for hepatocellular carcinoma in Japan. *Hepatol Res* **2013**, *43*, 44-50, doi:10.1111/j.1872-034X.2012.01029.x.
31. Guo, Z.; Zhong, J.H.; Jiang, J.H.; Zhang, J.; Xiang, B.D.; Li, L.Q. Comparison of survival of patients with BCLC stage A hepatocellular carcinoma after hepatic resection or transarterial chemoembolization: a propensity score-based analysis. *Ann Surg Oncol* **2014**, *21*, 3069-3076, doi:10.1245/s10434-014-3704-8.
32. Moriguchi, M.; Takayama, T.; Higaki, T.; Kimura, Y.; Yamazaki, S.; Nakayama, H.; Ohkubo, T.; Aramaki, O. Early cancer-related death after resection of hepatocellular carcinoma. *Surgery* **2012**, *151*, 232-237, doi:10.1016/j.surg.2010.10.017.
33. Yang, N.; Li, S.; Li, G.; Zhang, S.; Tang, X.; Ni, S.; Jian, X.; Xu, C.; Zhu, J.; Lu, M. The role of extracellular vesicles in mediating progression, metastasis and potential treatment of hepatocellular carcinoma. *Oncotarget* **2017**, *8*, 3683-3695, doi:10.18632/oncotarget.12465.
34. Forner, A.; Reig, M.; Bruix, J. Hepatocellular carcinoma. *Lancet* **2018**, *391*, 1301-1314, doi:10.1016/S0140-6736(18)30010-2.
35. Chang, D.Z.; Ma, Y.; Ji, B.; Liu, Y.; Hwu, P.; Abbruzzese, J.L.; Logsdon, C.; Wang, H. Increased CDC20 expression is associated with pancreatic ductal adenocarcinoma differentiation and progression. *J Hematol Oncol* **2012**, *5*, 15, doi:10.1186/1756-8722-5-15.
36. Wang, L.; Zhang, J.; Wan, L.; Zhou, X.; Wang, Z.; Wei, W. Targeting Cdc20 as a novel cancer therapeutic strategy. *Pharmacol Ther* **2015**, *151*, 141-151, doi:10.1016/j.pharmthera.2015.04.002.
37. Karra, H.; Repo, H.; Ahonen, I.; Loyttyneimi, E.; Pitkanen, R.; Lintunen, M.; Kuopio, T.; Soderstrom, M.; Kronqvist, P. Cdc20 and securin overexpression predict short-term breast cancer survival. *Br J Cancer* **2014**, *110*, 2905-2913, doi:10.1038/bjc.2014.252.
38. Kwan, P.S.; Lau, C.C.; Chiu, Y.T.; Man, C.; Liu, J.; Tang, K.D.; Wong, Y.C.; Ling, M.T. Daxx regulates mitotic progression and prostate cancer predisposition. *Carcinogenesis* **2013**, *34*, 750-759, doi:10.1093/carcin/bgs391.

39. Kato, T.; Daigo, Y.; Aragaki, M.; Ishikawa, K.; Sato, M.; Kaji, M. Overexpression of CDC20 predicts poor prognosis in primary non-small cell lung cancer patients. *J Surg Oncol* **2012**, *106*, 423-430, doi:10.1002/jso.23109.
40. Wu, W.J.; Hu, K.S.; Wang, D.S.; Zeng, Z.L.; Zhang, D.S.; Chen, D.L.; Bai, L.; Xu, R.H. CDC20 overexpression predicts a poor prognosis for patients with colorectal cancer. *J Transl Med* **2013**, *11*, 142, doi:10.1186/1479-5876-11-142.
41. Li, J.; Gao, J.Z.; Du, J.L.; Huang, Z.X.; Wei, L.X. Increased CDC20 expression is associated with development and progression of hepatocellular carcinoma. *Int J Oncol* **2014**, *45*, 1547-1555, doi:10.3892/ijo.2014.2559.
42. Chen, H.; Chen, J.; Zhao, L.; Song, W.; Xuan, Z.; Chen, J.; Li, Z.; Song, G.; Hong, L.; Song, P., et al. CDCA5, Transcribed by E2F1, Promotes Oncogenesis by Enhancing Cell Proliferation and Inhibiting Apoptosis via the AKT Pathway in Hepatocellular Carcinoma. *J Cancer* **2019**, *10*, 1846-1854, doi:10.7150/jca.28809.
43. Dai, C.; Miao, C.X.; Xu, X.M.; Liu, L.J.; Gu, Y.F.; Zhou, D.; Chen, L.S.; Lin, G.; Lu, G.X. Transcriptional activation of human CDCA8 gene regulated by transcription factor NF-Y in embryonic stem cells and cancer cells. *J Biol Chem* **2015**, *290*, 22423-22434, doi:10.1074/jbc.M115.642710.
44. Hayama, S.; Daigo, Y.; Yamabuki, T.; Hirata, D.; Kato, T.; Miyamoto, M.; Ito, T.; Tsuchiya, E.; Kondo, S.; Nakamura, Y. Phosphorylation and activation of cell division cycle associated 8 by aurora kinase B plays a significant role in human lung carcinogenesis. *Cancer Res* **2007**, *67*, 4113-4122, doi:10.1158/0008-5472.CAN-06-4705.
45. Chang, J.L.; Chen, T.H.; Wang, C.F.; Chiang, Y.H.; Huang, Y.L.; Wong, F.H.; Chou, C.K.; Chen, C.M. Borealin/Dasra B is a cell cycle-regulated chromosomal passenger protein and its nuclear accumulation is linked to poor prognosis for human gastric cancer. *Exp Cell Res* **2006**, *312*, 962-973, doi:10.1016/j.yexcr.2005.12.015.
46. Zhang, L.; Shao, H.; Huang, Y.; Yan, F.; Chu, Y.; Hou, H.; Zhu, M.; Fu, C.; Aikhionbare, F.; Fang, G., et al. PLK1 phosphorylates mitotic centromere-associated kinesin and promotes its depolymerase activity. *J Biol Chem* **2011**, *286*, 3033-3046, doi:10.1074/jbc.M110.165340.
47. Howard, J.; Hyman, A.A. Microtubule polymerases and depolymerases. *Curr Opin Cell Biol* **2007**, *19*, 31-35, doi:10.1016/j.ceb.2006.12.009.
48. Kline-Smith, S.L.; Khodjakov, A.; Hergert, P.; Walczak, C.E. Depletion of centromeric MCAK leads to chromosome congression and segregation defects due to improper kinetochore attachments. *Mol Biol Cell* **2004**, *15*, 1146-1159, doi:10.1091/mbc.e03-08-0581.
49. Moore, A.; Wordeman, L. The mechanism, function and regulation of depolymerizing kinesins during mitosis. *Trends Cell Biol* **2004**, *14*, 537-546, doi:10.1016/j.tcb.2004.09.001.
50. DeLuca, J.G.; Newton, C.N.; Himes, R.H.; Jordan, M.A.; Wilson, L. Purification and characterization of native conventional kinesin, HSET, and CENP-E from mitotic hela cells. *J Biol Chem* **2001**, *276*, 28014-28021, doi:10.1074/jbc.M102801200.

51. Kleylein-Sohn, J.; Pollinger, B.; Ohmer, M.; Hofmann, F.; Nigg, E.A.; Hemmings, B.A.; Wartmann, M. Acentrosomal spindle organization renders cancer cells dependent on the kinesin HSET. *J Cell Sci* **2012**, *125*, 5391-5402, doi:10.1242/jcs.107474.
52. Zhang, G.P.; Shen, S.L.; Yu, Y.; Yue, X.; Hu, W.J.; Li, S.Q. Kinesin family member 2C aggravates the progression of hepatocellular carcinoma and interacts with competing endogenous RNA. *J Cell Biochem* **2020**, 10.1002/jcb.29665, doi:10.1002/jcb.29665.
53. Balkwill, F.; Mantovani, A. Inflammation and cancer: back to Virchow? *Lancet* **2001**, *357*, 539-545, doi:10.1016/S0140-6736(00)04046-0.
54. Coussens, L.M.; Werb, Z. Inflammation and cancer. *Nature* **2002**, *420*, 860-867, doi:10.1038/nature01322.
55. Liu, Y.; Cao, X. The origin and function of tumor-associated macrophages. *Cell Mol Immunol* **2015**, *12*, 1-4, doi:10.1038/cmi.2014.83.
56. Amann, T.; Bataille, F.; Spruss, T.; Muhlbauer, M.; Gabele, E.; Scholmerich, J.; Kiefer, P.; Bosserhoff, A.K.; Hellerbrand, C. Activated hepatic stellate cells promote tumorigenicity of hepatocellular carcinoma. *Cancer Sci* **2009**, *100*, 646-653, doi:10.1111/j.1349-7006.2009.01087.x.
57. Mantovani, A.; Allavena, P.; Sica, A.; Balkwill, F. Cancer-related inflammation. *Nature* **2008**, *454*, 436-444, doi:10.1038/nature07205.
58. Prieto, J.; Melero, I.; Sangro, B. Immunological landscape and immunotherapy of hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol* **2015**, *12*, 681-700, doi:10.1038/nrgastro.2015.173.
59. Qin, L.X. Inflammatory immune responses in tumor microenvironment and metastasis of hepatocellular carcinoma. *Cancer Microenviron* **2012**, *5*, 203-209, doi:10.1007/s12307-012-0111-1.
60. Fu, J.; Xu, D.; Liu, Z.; Shi, M.; Zhao, P.; Fu, B.; Zhang, Z.; Yang, H.; Zhang, H.; Zhou, C., et al. Increased regulatory T cells correlate with CD8 T-cell impairment and poor survival in hepatocellular carcinoma patients. *Gastroenterology* **2007**, *132*, 2328-2339, doi:10.1053/j.gastro.2007.03.102.
61. Said, E.A.; Dupuy, F.P.; Trautmann, L.; Zhang, Y.; Shi, Y.; El-Far, M.; Hill, B.J.; Noto, A.; Ancuta, P.; Peretz, Y., et al. Programmed death-1-induced interleukin-10 production by monocytes impairs CD4+ T cell activation during HIV infection. *Nat Med* **2010**, *16*, 452-459, doi:10.1038/nm.2106.
62. Shi, F.; Shi, M.; Zeng, Z.; Qi, R.Z.; Liu, Z.W.; Zhang, J.Y.; Yang, Y.P.; Tien, P.; Wang, F.S. PD-1 and PD-L1 upregulation promotes CD8(+) T-cell apoptosis and postoperative recurrence in hepatocellular carcinoma patients. *Int J Cancer* **2011**, *128*, 887-896, doi:10.1002/ijc.25397.

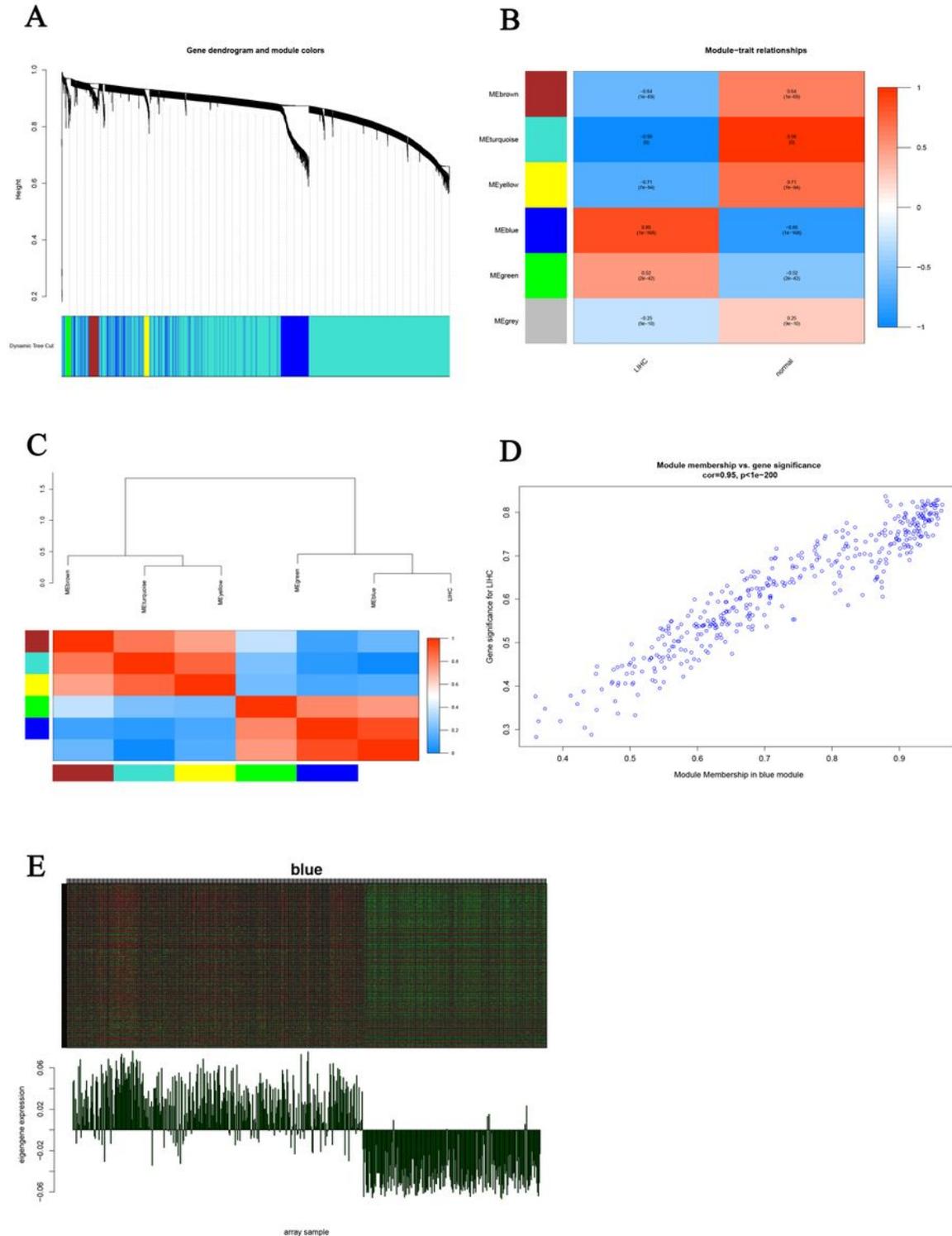
## Figures



**Figure 1**

Translational analysis between normal samples and liver hepatocellular carcinoma (LIHC) samples. (a) Heatmap plots of all genes. (b) Volcano plot of all genes in LIHC. Red dots: up-regulation; blue dots: down-regulation; black dots: non-differentially expressed genes. (c) Gene ontology (GO) analysis of DEGs. The biological process (BP), cellular component (CC), and molecular function (MF) of potential targets were clustered based on ClusterProfiler package in R software. The node size represent the numbers of genes. Adjust p-value was shown in the color intensity. Adjust p-values of all terms were less than 0.05. (d) The enriched KEGG signaling pathways were selected to demonstrate the primary biological actions of differential expression genes (DEGs). The abscissa indicates gene ratio and the enriched pathways

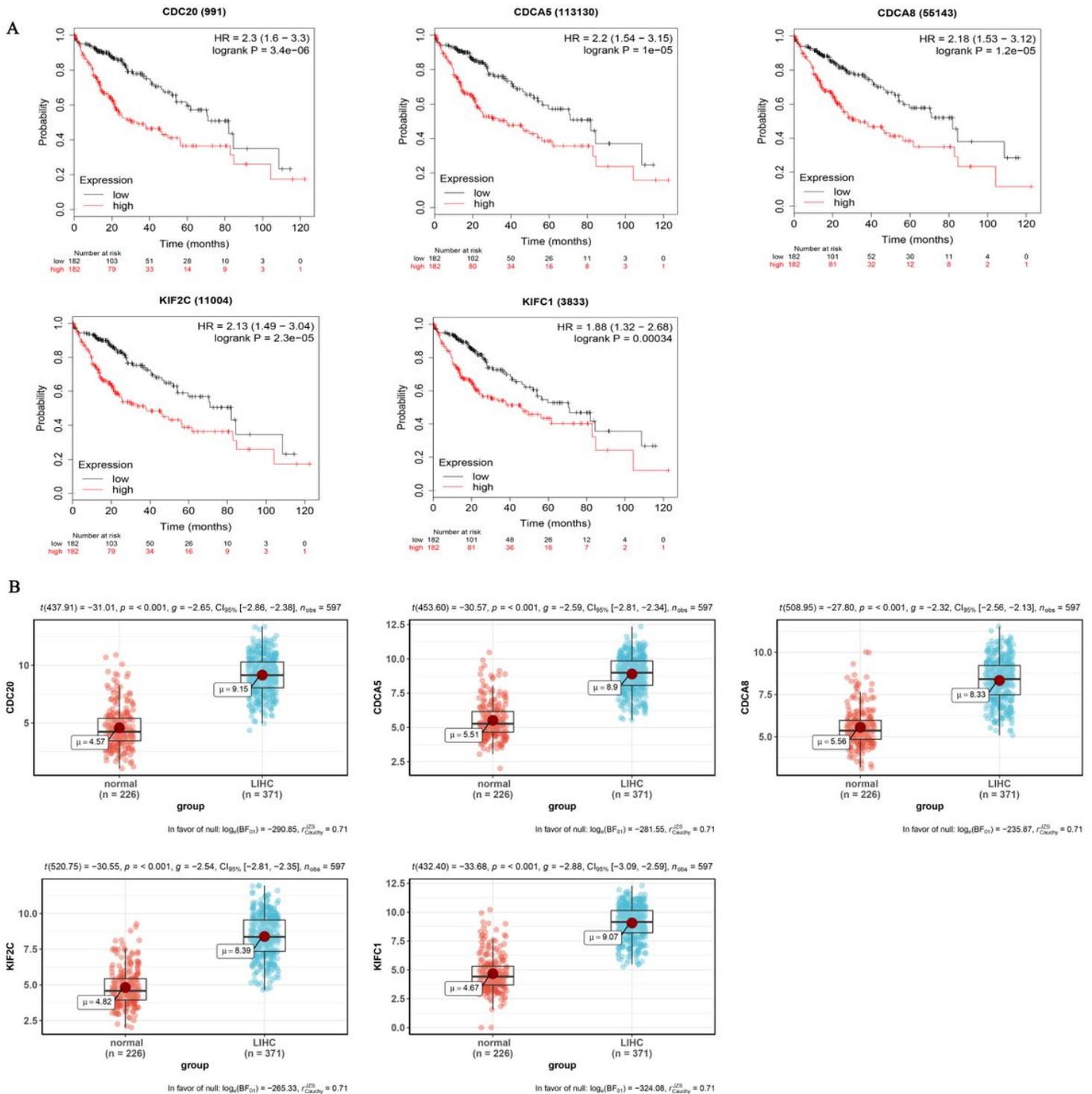
were presented in the ordinate. Node size is based on the numbers of genes, and the color intensity corresponds to the adjusted p-value. (e) KEGG-GSEA of DEGs. Adjust p-values of enriched pathways were less than 0.05.



**Figure 2**

Weighted gene co-expression network analysis (WGCNA) was performed to analyze differential expression genes. (a) Cluster dendrogram and module assignment for modules from WGCNA. The

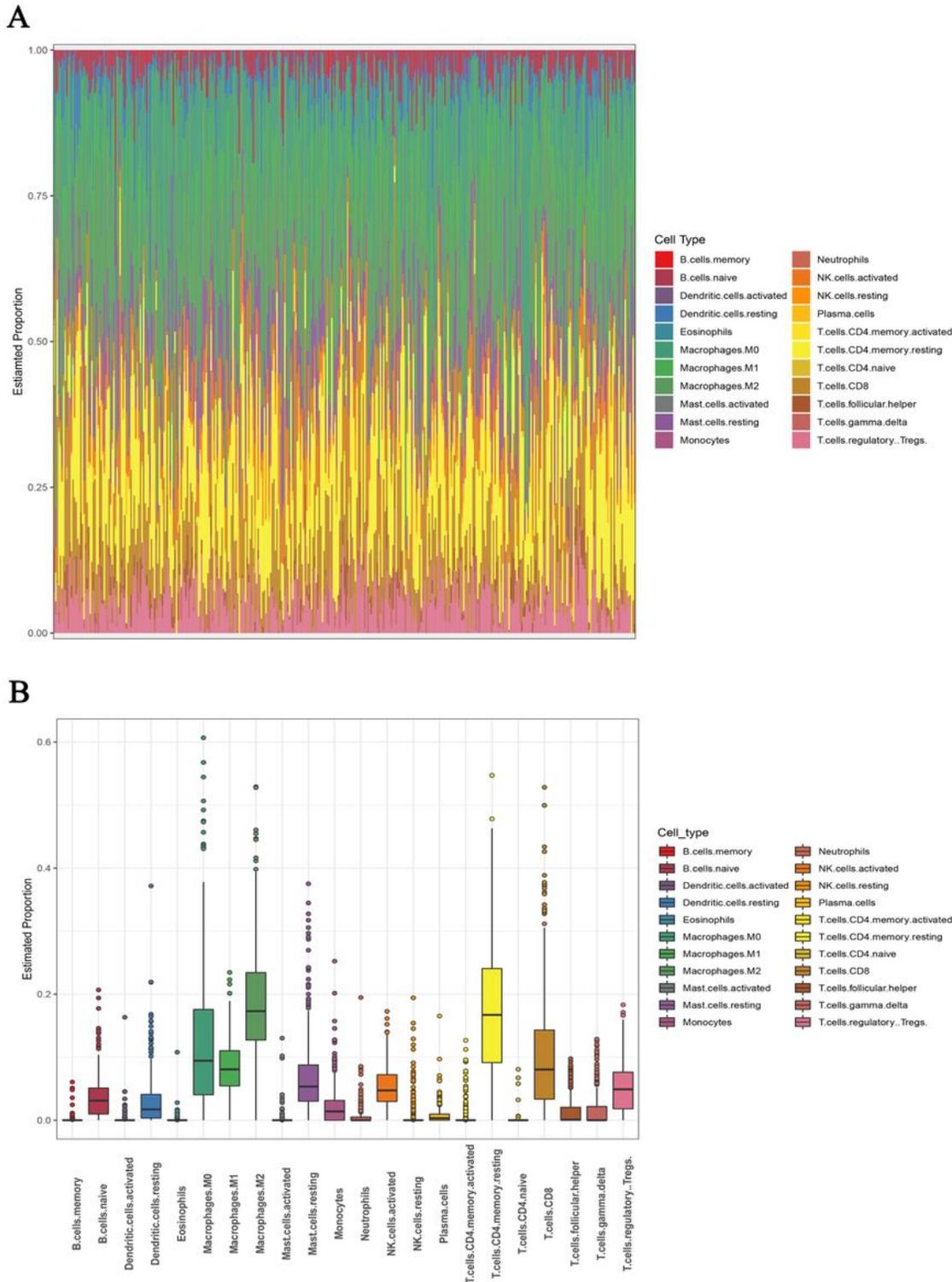
colored horizontal bar below the dendrogram represent the modules. (b) The correlation between various modules and LIHC. The correlation coefficient and P-value were presented in the heatmap. (c) Cluster between module eigengenes and IA was shown. (d) Scatter plot of module membership in the blue module. (e) Heatmap represented translational levels of all genes in the blue module.



**Figure 3**

Validation of novel biomarkers. (a) Kaplan–Meier survival plots for overall survival in hub genes. Red line and black line was for the samples with gene highly expressed and lowly

expressed. The X and Y axes represent survival time (months) and percent of survival, respectively. (b) Validation of the genes expression levels between normal samples and liver hepatocellular carcinoma.



**Figure 4**

CIBERSORT analyzed liver hepatocellular carcinoma (LIHC) RNA-seq samples. (a) The proportion of immune cell types was showed in bar plots. (b) Box plots summarizing immune cell proportions in all LIHC samples.

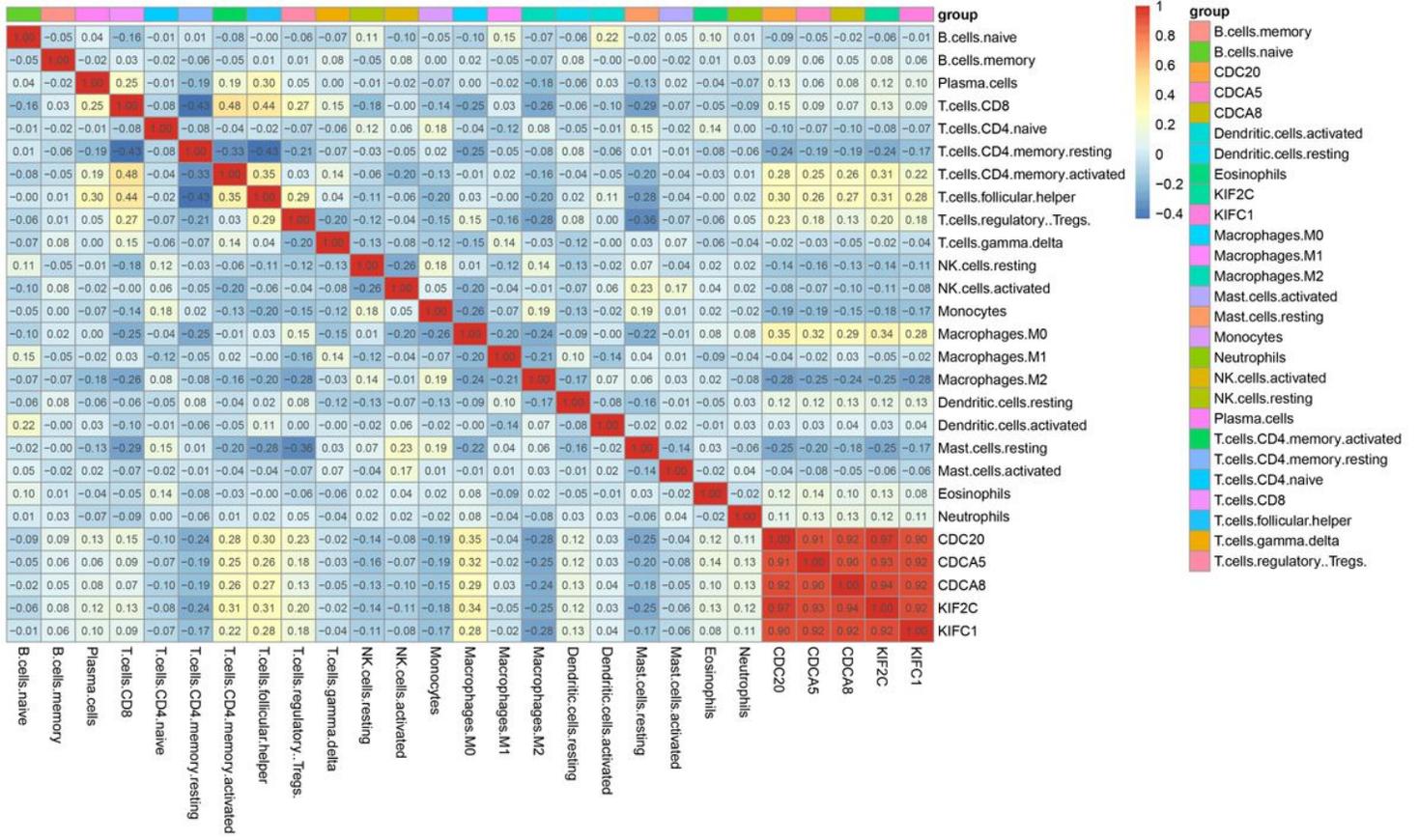
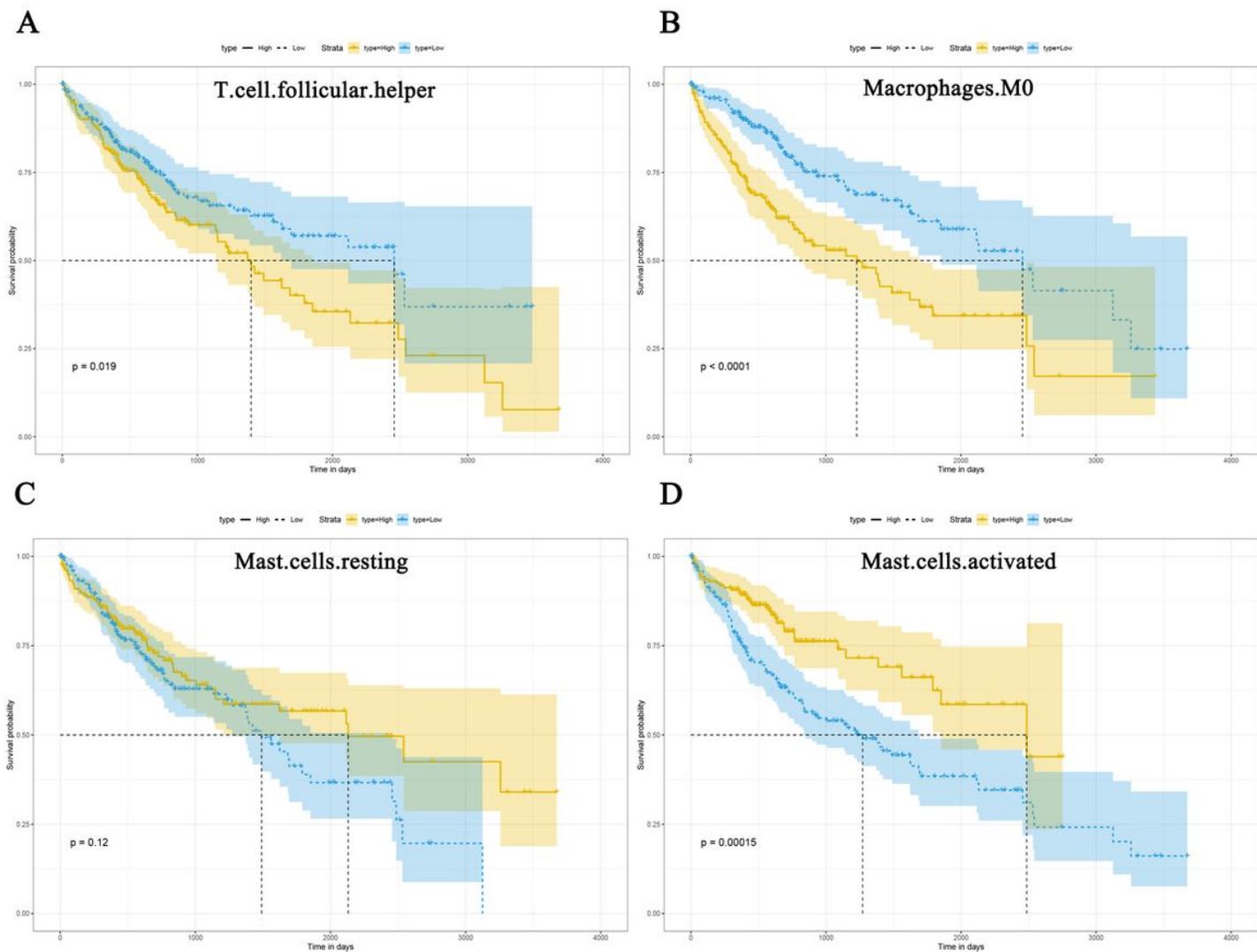


Figure 5

The heatmap of pearson's correlation. Pearson's correlation between infiltrating immune cells and the novel biomarkers. Pearson's r values were presented in the heatmap.



**Figure 6**

Prognostic value of immune cells in liver hepatocellular carcinoma tumor microenvironment (TEM). Kaplan–Meier survival curves for infiltrating follicular helper T cells (a), macrophages M0 (b), resting mast cells (c), activated mast cells (d). Yellow lines represented high immune infiltration or score in TEM and blue lines represented low immune infiltration or score in TEM.

## Supplementary Files

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

- [FigureS1.jpg](#)
- [FigureS2.jpg](#)
- [FigureS3.jpg](#)
- [FigureS4.jpg](#)