Abstract. – OBJECTIVE: The liver is a unique organ containing large populations of immune cells. Immunotherapy for liver cancer is a promising yet particularly challenging method. Therefore, it harbors great significance for the identification of immune-related subtypes and the potential therapeutic targets for hepatocellular carcinoma (HCC).

MATERIALS AND METHODS: Firstly, we classified the HCC samples downloaded from the dataset of Cancer Genome Atlas (TCGA) into two clusters based on the immune cell infiltration. Thereafter, we identified the significant module and regulatory factors using the weighted gene co-expression network analysis (WGCNA). The immune competence of the regulatory factors was delineated through the ESTIMATE algorithm, the analysis of the tumor microenvironment, and pan-cancer analysis. In the single-cell RNA sequencing analysis, we further explored the immune competence of regulatory factors. We also collected the potential drugs targeting the regulatory factors. In addition, we constructed lncRNA-miRNA-mRNA interaction regulatory networks. Finally, western blot and quantitative real-time polymerase chain reaction (qRT-PCR) were conducted to verify the protein expression of regulatory genes in HCC cell lines and tissues.

RESULTS: According to the immune cell infiltration, two immune-related subtypes-cluster 1 and cluster 2-were found. Patients in cluster 2 had a more significant immune infiltration than in cluster 1. Afterward, six significant regulatory genes were identified through WGCNA, and the expression in cluster 2 was high in cluster 1. We performed a comprehensive analysis to clarify the immune signature. The results showed that the six genes had significant immunological competence. Moreover, the expression of the six genes was similar to the subtypes’ classification. In the analysis of the prognosis value, patients in cluster 2 had a better prognosis. In addition, the IncRNA in the IncRNA-miRNA-mRNA interaction regulatory networks was located in the nucleus and cytoplasm. In the single-cell RNA sequencing analysis, the six genes were related to the immune cell. We also identified potential drugs for CD6 and CLEC12A, which may provide potential therapeutic drugs. Finally, the regulatory genes were verified in the western blot and quantitative real-time polymerase chain reaction.

CONCLUSIONS: The classification into two clusters based on the immune cell infiltration may provide a promising prospect for HCC through immunotherapy. The six regulatory genes may be potential therapeutic targets in the treatment of HCC.

Key Words: Hepatocellular carcinoma, Immune status, Prognosis value, Biomarkers, Immunotherapy.

Introduction

Hepatocellular carcinoma (HCC) is the second leading cause of cancer-related death worldwide, leading to approximately 800,000 cases in 20181. Most patients have backgrounds of chronically inflamed liver (hepatitis B, hepatitis C, alcoholic and nonalcoholic liver disease2-5, and liver fibrosis/cirrhosis. Most patients are diagnosed when their tumors are already too advanced, which leads to poor prognosis for them. Therefore, palliative systemic treatment is necessary for those patients. Although the results of several plethoras of late-stage clinical trials invigorating, such as Lenvatinib and regorafenib, the objective response rates are only 15-20%6.
In recent years, the results of much clinical research found that the immune cell composition of HCC tumors is closely associated with the overall prognosis and the effect of therapy\(^7\)\(^-\)\(^13\). The liver contains some immune cells which have a strong antitumor potential including natural killer cells (NK cells), helper innate lymphoid cells (ILCs), natural killer T cells (NKT cells), mucosal-associated invariant T cells (MAITs), gamma delta T cells (γδ T cells), macrophages, granulocytes, and dendritic cells (DCs), patrols the liver to maintain homeostasis\(^14\). Many studies\(^15\)\(^,\)\(^16\) have found that the tumor microenvironment (TME) is complex and diverse, and different aspects of the TME play significant impacts on prognosis and the efficacy of immunotherapy. Therefore, it is significant to identify the reasonable subtypes of patients potentially responsive to immunotherapy and identify the possible regulatory factors to exert anti-tumor properties.

In this study, we first identified the two clusters based on the immune cell infiltration and validated the classification from the immune status. Next, we identified the significant regulatory factors (BIN2, ARHGAP9, CD6, DOK2, C1orf162, CLEC12A), and the expression and immune competence in cluster 2 were higher in cluster 1. In the analysis of the prognosis value, patients in cluster 2 had a better prognosis. We also constructed lncRNA-miRNA-mRNA interaction regulatory networks, and the lncRNA was located in the nucleus and cytoplasm. The findings in the present work show that the classification of two subtypes and the identification of the regulatory factors may lay a basis for future studies on immunotherapy for the treatment of HCC patients.

**Materials and Methods**

**Ethics Statement**

The studies were approved by the Medical Ethical Committee of the Beijing Institute of Radiation Medicine (Beijing, China). Written informed consent was obtained from each patient.

**HCC Samples Data Collection and Identification of the Immune-Related Subtypes**

The flow chart of the study was presented in Figure 1. The RNA-seq profiles and clinical...
information of the HCC cohort consisting of 374 patients were downloaded from The Cancer Genome Atlas (TCGA) data portal (https://cancergenome.nih.gov/). All biomarkers for each immune cell were collected. Then, we evaluated the level of immune cell infiltration using the GSVA package and identified the immune subtypes based on the immune-related signatures for immune cells. Principal component analysis (PCA) was used to verify the reliability of the consensus clusters.

**Evaluation of the Immune Microenvironment and Prognosis Value Between Immune Subtypes**

First, we performed several kinds of different methods to evaluate the level of immune cell infiltration between different immune subtypes. Afterward, we analyzed the characteristics of the immune-related pathways in different immune subtypes using the GSVA package. The list of immune-related genes was collected from the database of Immunology Database and the Analysis Portal (ImmPort) (https://www.immport.org/shared/genelists). ImmPort includes the gene datasets of 17 immune-related pathways. ESTIMATE is employed to measure the immune cell infiltration degree (immune score) and tumor purity among the clusters. We calculated the immune score using ESTIMATE packages for the immune-related clusters. In addition, we calculated the mean value of perforin (PRF1) and granzyme A (GZMA) and obtained the cytolytic activity (CYT) score. The expression of the immunotherapy genes in different clusters also was evaluated.

Overall survival, disease-free survival, and progression-free survival between different subtypes were compared through the Kaplan-Meier method. Immune-related genes include the gene datasets of the chemokine, human leukocyte antigen (HLA), receptor, immunostimulatory and immunosuppressive. The differential expression of those genes in different immune subtypes was also evaluated.

**The Analysis of the Functional Enrichment for the Immune Subtypes**

Firstly, the edgeR package was used to screen out the differential expression genes, and $|\log_{2}(\text{FoldChange})|>1$ and $q<0.05$ ($p<0.05$, Benjamini & Hochberg) were selected as the thresholds for statistical significance. Then, we employed Gene Ontology (GO) as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis using the ClusterProfile package to reveal their potential functions ($p<0.05$). Hallmark gene sets were abstracted from the Molecular Signature Database of Broad Institute (MSigDB) (https://www.gsea-msigdb.org/gsea/index.jsp). We analyzed the hallmark for different clusters using the GSVA package and collected the differential hallmark pathways (false discovery rate <0.05) using the limma package.

**The Identification of the Most Significant Modules and Regulatory Genes**

A weighted gene co-expression network analysis (WGCNA) was used to identify the most significant consensus gene module, and the functional enrichment for the module was further conducted. Then, we screened out the significant regulatory genes, which should meet the criterion of the correlation with a module greater than 0.8, the correlation with clusters greater than 0.5, and the largest degree in the top 25% of the module.

**The Expression of Significant Regulatory Genes and Survival Characteristics in Different Clusters**

The expression of the genes in different clusters has also been analyzed. In addition, we explored the survival characteristics of the genes in the subtypes.

**The Analysis of the Association Between the Significant Regulatory Genes and Immune Cells**

Firstly, we evaluated the association among the expression of the 6 genes themselves. Then, we explored the association between those regulatory genes and immune cells.

**Immune Status Analysis of the Regulatory Genes Between Different Clusters**

An estimate package was used to calculate the immune score, estimate score, and stromal score, which were utilized to explore the association with the regulatory genes. Afterward, we further explored the association between immunotherapy genes, immune gene sets, immune cell infiltration, and the significant regulatory genes. The expression of the regulatory genes in immune cells was analyzed through the Seurat package.
The Mutation and Pan-Cancer Prognosis Analysis of the Significant Factors

The significant genes were explored in HCC. Moreover, the disease-free survival and progression-free survival analysis of pan-cancer were also performed.

The Analysis of Common Genes in the Intersection of the Genes in the PPI Network and the DEGs of the two Clusters

A list of cancer-related genes and immune-related genes was collected from COSMIC databases (http://cancer.sanger.ac.uk) and ImmPort databases, respectively. The Search Tool for the Retrieval of Interacting Genes (STRING, https://string-db.org/cgi/input.pl) is an online database resource that includes approximately 24.6 million proteins and more than 3.1 billion interactions. We constructed a PPI network for the regulatory genes using the STRING with a confidence score >0.900. Then, we collected the common genes at the intersection of the genes in the PPI network and DEGs of the two clusters.

The Construction of the IncRNA-miRNA-mRNA Competitive Endogenous RNA (ceRNA) Network

The miRcode database (http://www.mircode.org/) was used to match DEIncRNAs and DEMiRNAs. The miRNA and mRNA pairs were collected from miRcode databases (http://miRcode.nctub.tw/), miRTarBase databases (http://www.mirdb.org/), and TargetScan databases (http://www.targetscan.org/). Subsequently, we integrated the interaction between miRNAs and IncRNAs or mRNAs to construct a ceRNA regulatory network using the Cytoscape (version 3.7.2).

Drug-gene Interaction Analysis of Significant Regulatory Genes

The Therapeutic Target Database (http://db.idrblab.net/) provides information on drugs with the corresponding proteins, pathways, and diseases. The drug-gene interaction database (DGIdb, www.dgidb.org) consolidates disparate data sources and describes drug-gene interactions and gene druggability. DrugBank (www.drugbank.ca) is a web-enabled database containing comprehensive molecular information about drugs, their mechanisms, their interactions, and their targets. We identified the interactions between the drugs and significant regulatory genes from these databases.

The protein was extracted from the HCC cell line, and the protein concentration was determined according to the introduction of the BCA Kit (Boster Biological Technology Co. Ltd, Wuhan, Hubei, China). The extracted protein was added to the loading buffer and boiled at 95°C for 10 minutes, and each well was loaded at 30 μg. The protein was isolated by 10% polyacrylamide gel electrophoresis (Boster Biological Technology Co. Ltd, Wuhan, Hubei, China) with voltage transferred from 80 v to 120 v. The wet transfer was conducted with the transmembrane voltage of 100 mv for 45-70 minutes. Then, the protein was transferred to polyvinylidene fluoride (PVDF) membrane, and the membrane was sealed by 5% bovine serum albumin (BSA) at room temperature, added with the primary antibody BIN2 (1:1,000), Polyclonal Antibody to Cluster Of Differentiation 6 (CD6) (1:1,000), C-type lectin domain family 12 member A (CLEC12A) (1:1,000), Clorf162 (1:500), [all purchased from Cell Signaling Technologies (CST), Beverly, MA, USA], and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (1:1,500) (Becton and Dickinson Company, Bioscience, San Jose, CA, USA) for incubation at 4°C overnight. The membrane was washed with Tris-buffered saline tween (TBST) 3 times (5 minutes for each), incubated with corresponding secondary antibody at room temperature for 1 hour, washed 3 times (5 minutes for each), and developed by chemiluminescence reagent. With GAPDH serving as an internal reference, Bio-rad Gel Doc EZ imager (GEL DOC EZ IMAGER, Bio-rad, California, USA) was applied to develop. The gray value of the target band was analyzed by Image J software, protein expression = target band gray value / internal reference gray value.

The total RNA was extracted using a Trizol kit (Invitrogen). Complementary DNA (cDNA) was synthesized using a TaqMan™ MicroRNA reverse transcript kit (Applied Biosystems Company, Shanghai, China). The PCR assay was performed by ABI Prism® 7300 real-time fluorescence quantitative PCR instrument (Applied Biosystems, Shanghai, China). U6 and GAPDH were set as endogenous controls. The relative levels were calculated by the 2^(-ΔΔCt) method.

Statistical Analysis

All statistical analyses were performed by R software, version 4.1.2 (www.r-project.org). Kaplan-Meier survival curves were drawn by the R package “Survival” and the log-rank test was
used for comparisons. The Wilcoxon test was used to compare continuous variables. For all statistical analyses, the threshold of statistical significance was defined as \( p < 0.05 \).

**Results**

**Identification of two Immune Subtypes for HCC**

We first downloaded the HCC cohort from TCGA databases using the TCGAbiolinks package\(^{28}\) in R software, which included 374 HCC samples and 50 normal samples. The ssGSEA algorithm from GSVA packages was used to evaluate the level of infiltration for 28 immune cells in the HCC samples. Then, the two immune subtypes, the optimal classification in patients with HCC, were identified based on the immune cell infiltration score (Figure 2A-C). Cluster 1 included 230 patients with HCC, while cluster 2 consisted of 144. There was more immune cell infiltration in cluster 2 than in cluster 1. This immune-related classification for HCC was also well presented in PCA. To compare the level of immune cell infiltration between two clusters, we calculated the GSVA score using the GSVA packages. The results showed that almost all immune cells except Memory B cells were particularly enriched in cluster 2 (\( p < 0.05 \)) (Supplementary Figure 1).

**The Difference in Tumor Microenvironment and Prognosis Value Between Two Immune-Related Subtypes**

To further evaluate the difference in tumor microenvironment between two immune subtypes, we took several common methods to judge the level of immune cell infiltration, including CIBERSORT, TIMER, quanTIseq, EPIC, and xCell. Although the analysis for immune cell infiltration between the two immune subtypes revealed that only four types of the 22 immune cells inferred by the CIBERSORT analysis had significantly different degrees of infiltration, the fraction of these immune cells was significantly higher in cluster 2 than in cluster 1 (Figure 2D). The infiltration analysis between the two immune subtypes inferred by the TIMER analysis demonstrated that five types of six immune cells in cluster 2 were significantly higher degrees of infiltration than in cluster 1 (Figure 2E). In the quanTIseq analysis, the infiltration of eight types of 11 immune cells was significantly higher in cluster 2 than in cluster 1 (Figure 2F). The significantly different fractions of B cell (\( p < 0.001 \)), cancer-associated fibroblast (\( p < 0.001 \)), T cells CD4+ (\( p < 0.01 \)), and endothelial cell (\( p < 0.05 \)) in the EPIC analysis were higher in cluster 2 than in the cluster 1 (Figure 2G). In the xCell analysis, compared within cluster 1, the immune cell infiltration in the cluster 2 group had a higher xCell score. The immune score (\( p < 0.001 \)),stroma score (\( p < 0.001 \)), and microenvironment score (\( p < 0.001 \)) had the same situation (Figure 2H). Therefore, the level of immune cell infiltration in cluster 2 was higher in cluster 1 after the analysis of the immune cell infiltration using five methods, which could suggest that cluster 2 had a hotter tumor microenvironment.

To compare the difference in immune-related pathways between two immune subtypes, we collected the gene datasets of 17 immune-related pathways from ImmPort databases. Then, we calculated the score of immune-related pathways for each sample using the ssGSEA algorithm in the GSVA package and compared the score of immune-related pathways for two clusters. The results showed that 14 of 17 immune-related pathways (except interferons, TGFβ Family Member, and TGFβ Family Member Receptor) were remarkably enriched in the two clusters, and the enrichment in cluster 2 was higher than in cluster 1. We also computed the immune score for two subtypes using the ESTIMATE package in the R software, and the result showed that the degree of enrichment in cluster 2 was higher than in cluster 1. The cytolytic activity score (CYT) is a quantitative measure of immune cytolytic activity based on transcript levels of GZMA and PRF1\(^{26}\). We collected the CYT score using the mean of GZMA and PRF1 expression for each group and found that the CYT score in cluster 2 was higher in cluster 1. To evaluate the expression of the immunotherapy genes between two clusters, we collected 12 immunotherapy genes (\textit{CD274, PDCD1, CTLA4, CD80, CD86, C3D38, TIGIT, ICOS, IDO1, LAG3, CCL5, CXCL10})\(^{29}\). The result showed that the expression of twelve immunotherapy genes in cluster 2 was significantly higher than in cluster 1 (Supplementary Figure 2). Those indicated that the patients in cluster 2 were more sensitive to immunotherapy than those in cluster 1.

In addition, we analyzed the expression of the immune-related genes between two groups using the gene sets of chemokine, HLA, receptor, immunostimulatory, and immunosuppressive (Supplementary Figure 3). The results showed...
Figure 2. The classification of immune-related subtypes based on the immune cell infiltration score and the comparison of the immune cell infiltration score in different subtypes through the analysis of CIBERSORT(A), TIMER(B) quanTIseq, EPIC, and xCell. A. Cumulative distribution function (CDF) curve of the consistency score for different subtype numbers (k = 2-9). B. Delta area curves represent the relative change in area under the cumulative distribution function (CDF) curve for each category number k compared to k-1. C. The consensus score matrix for HCC samples when k = 2. A higher consensus score between two samples indicates that they are more likely to be assigned to the same cluster in different iterations, and the classification effect is the best. The comparison of the immune cell infiltration score in different subtypes through the analysis of CIBERSORT (D), TIMER (E), quanTIseq (F), EPIC (G), xCell (H). p values were showed as: p≤0.05; **; p≤0.01; ***; p≤0.001; ****; p≤0.0001; ns: not significant.
that the expression of those genes in cluster 2 was higher in cluster 1.

In the prognosis analysis, although no difference between the two clusters in the overall survival was not found, the patients with HCC in cluster 2 had a longer disease-free survival and progression-free survival than patients in cluster 1 (Supplementary Figure 2E-F). Patients with HCC in cluster 2 had better prognostic benefits than those in cluster 1. Collectively, we thought that patients in cluster 2 had a hotter tumor microenvironment and better prognosis than those in cluster 1 from the comprehensive analysis of the immune landscape, including immune cell infiltration, immune-related pathways, and immunotherapy genes.

### The Functional Enrichment of the two Immune-Related Subtypes

We collected the differential expression genes (DEGs) between the two subtypes using the edgeR package \((q<0.05\) and \(|\log(FoldChange)|>1\), Benjamini and Hochberg procedure). A total of 1,233 DEGs were collected (False Discovery Rates (FDR)<0.05 and fold change>2 or fold change <1/2), 1,055 DEGs of which were up-regulated, while 178 DEGs had down-regulated. Then, we performed the functional enrichment analysis of the KEGG and GO using the Cluster Profile package (Figure 3A). The results of the top 30 pathways in the KEGG enrichment analysis showed that pathways for DEGs were remarkably enriched in the immune-related pathways such as chemokine signaling pathway, Th17 cell, Th1 cell, Th2 cell, B cell pathways, and immune network (Figure 3B). The results in the GO enrichment analysis demonstrated that many pathways were mainly enriched in the biological process of immunity, such as humoral immune response and the activation of T cells (Figure 3C). These results demonstrated that the significant difference in tumor microenvironment between the two immune subtypes existed, and the DEGs between the two subtypes were involved in many pathways related to immunity. We also downloaded the Hallmark gene sets from MsigDB databases and evaluated the difference of hallmarks between immune-related subtypes using GSVA and limma package (FDR<0.05). A total of 34 differential hallmarks were obtained, and the result showed that those hallmarks were mainly involved in apoptosis, inflammation, B2M checkpoint, and WNT signal pathways (Supplementary Figure 4).

### The Identification of Significant Regulators

We did a co-expression analysis based on the differential expression of mRNA between the two subtypes using the module-trait relationships approach of WGCNA. Among the significant modules, the turquoise module \((R=0.52)\) was significantly related to subtypes (Figure 4A). Then, we performed a functional enrichment annotation for the module, and the results showed that the module was highly enriched in the KEGG pathways related to immune biological processes (Figure 4B-C). This suggested that the genes of the module were more representative. To identify the significant regulatory genes, we set the relationship score of the module to more than 0.8 and the subtypes to more than 0.5 and collected the genes located in the top 25% of this module. We finally identified six significant regulatory genes \((BIN2, ARHGAP9, CD6, DOK2, C1orf162, CLEC12A)\).

### The Expression and Prognosis Value of the Significant Regulatory Genes

The expression of the six significant regulatory genes between two immune-related subtypes in cluster 2 was significantly higher in cluster 1 (Figure 5A-B). Due to the reason that the six regulatory genes originated from the same module and expression patterns were highly consistent with prognostic value, we performed the analysis of expression for the six genes themselves. The results showed that the relationship among six genes was positively correlated (the correlation coefficient was more than 0.7) (Figure 5C). We also found that the expression of six genes was positively correlated with the immune-activated cells and negatively associated with immunosuppressed cells (Figure 5D). Those results indicated that the six regulatory genes had important immune competence.

In the prognosis value analysis of disease-free survival and progression-free survival, we found that significant differences existed between the high expression and the low expression. Moreover, the survival analysis showed that patients with high expression had a better prognosis than patients with low expression (Supplementary Figure 5), which was consistent with the expression of six regulatory genes between subtypes.

### The Immunological Competence of the six Important Regulators

To clarify the immune function of the six important regulatory genes, we performed an analysis of immunological competence. Firstly, we did a Pearson correlation after collecting the
The result showed that six genes were significantly associated with the immune score, especially DOK2 (the correlation value = 0.89, p<0.05) (Figure 5E). Then, we collected the immunotherapy genes, the expression of which was significantly different between the two immune subtypes. Moreover, those immunotherapy genes were particularly related to the six important regulators, especially BIN2 and CD86 (the correlation value = 0.9, p<0.05) (Figure 5F).

To evaluate the association between six important regulators and immune cell infiltration, we performed the analysis of TIMER and xCell. In the analysis of TIMER, the six important regulators were significantly associated with the immune cell infiltration score, especially the association between BIN2 and Myeloid dendritic cell (Pear-
son correlation analysis, the correlation value = 0.81 and p<0.05) (Figure 5G). In the analysis of xCell, the results showed that CD6 had the most significant association with Myeloid dendritic cells (Pearson correlation analysis, the correlation value = 0.8, p<0.05) (Figure 5H). Those results showed that the genes BIN2 and CD6 had the most significant immunological competence, while DOK2 had the most significant relationship with the immune score. There also had internal correlations among the six important regulators. Therefore, we speculated that all six important regulators may have significant immunological competence.

To further elucidate the immunological competence of the six important regulators, we collected the single-cell sequence data of immune cells for the liver from scRNA-HCC databases (https://www.omic.tech/scrna-hcc#). It includes more than 70,000 single-cell transcriptomes for 10 HCC patients from four relevant sites: primary tumor, portal vein tu-

Figure 4. The identification and analysis of the significant module and regulatory factors. A, The identification of the significant module in the analysis of module-trait relationships. The red and blue colors indicate a strong positive correlation and strong negative correlation, respectively. The functional enrichment analysis of the significant module in the analysis of the Kyoto encyclopedia of genes and genomes (KEGG) (B) and Gene Ontology (GO) (C). p values were showed as: ns: not significant; *: p≤0.01; **: p≤0.001; ***: p≤0.0001; ****: p≤0.00001;
Figure 5. The correlation analysis and immunological competence of the significant regulatory factors. Comparison of the expression of the significant regulatory genes between two clusters (A-B) (p<0.05). C. The analysis of the self-expression of the regulatory genes. D. The association between the regulatory genes and the status of the immune cells. E. The association between the six regulatory factors and the immune score, the estimated score, and the stromal score. F. The association between the six regulatory factors and immunotherapy genes. The association between the six regulatory factors and immune cell infiltration through the TIMER algorithm and xCell algorithm (G and H). I. Visualization of clustering and annotation in the tSNE plot. J. The expression of the regulatory genes in the immune cells. ****: p≤0.001.
Potential biomarkers for the prognosis and treatment of HCC immunotherapy

We then explored the expression of the six genes. The results showed that all significant genes had expression in some clusters, while some genes were expressed in almost all clusters (Figure 5I). The expression level of CLEC12A, DOK2, ARHGAP9, BIN2, and C1orf162 was high in the Myeloid dendritic cell, which was consistent with the analysis result of the xCell and TIMER. Therefore, those results suggested that the six important regulators had significant immunological competence.

We also explored the relationship between the gene sets of chemokine, HLA, receptor, immunostimulatory and immunosuppressive, and six important regulators. The results showed that the six genes were significantly related to those gene sets, and the correlation patterns of these genes were similar. In the chemokine analysis, CD6 and CCL5 had the strongest correlation (the correlation value = 0.88, p<0.05). BIN2 and HLA-DA had the strongest correlation (the correlation value = 0.87, p<0.05). In HLA, the strongest correlation was BIN2 and CCR5 (the correlation value = 0.87, p<0.05). In immunostimulatory analysis, CD6 and CD48 had the strongest correlation (the correlation value= 0.94, p<0.05). In immunosuppressive (the correlation value = 0.94, p<0.05). In immunosuppressive analysis, CD6 and CD96 had the strongest correlation (the correlation value = 0.94, p<0.05) (Supplementary Figure 6). Those results showed that the six important regulators had significant immunological competence, especially CD6 and BIN2.

The Mutation and Pan-Cancer Prognosis Analysis of the Six Important Regulators

In the expression analysis of the six important regulators, we found that the mutation frequency of the six genes in HCC was relatively low (Figure 6A). Those results indicated that these genes might influence the two immune subtypes of HCC by the expression level of gene expression. To explore the function of the six important regulators in pan-cancer, we analyzed disease-free survival and progression-free survival. The result showed that BIN2 could influence disease-free survival in 12 cancers and progression-free survival in 19 cancers (Figure 6B). CD6 could play a role in 10 cancers and affect the progression-free survival in 19 cancers (Figure 6C). DOK2 could contribute to disease-free survival in 9 cancers and progression-free survival in 15 cancers (Figure 6D). ARHGAP9, C1orf162, and CLEC12A had a similar situation. Those results indicated the significant prognosis value of the six regulatory genes (Supplementary Figure 7).

PPI Network Construction Between Immune-Related Genes and Significant Regulatory Genes

A PPI network was constructed for the six significant regulatory genes using STRING with a confidence score >0.900. The result showed that except for BIN2, the other five regulatory genes were constructed into a close network, which had a total of 2,062 nodes. Then, we obtained the common genes using the intersections of genes between the DEGs of two clusters and the genes in the network. We found that 17 cancer-related genes, 133 immune-related genes, 22 immune-related oncogenes, and the other 318 genes were closely related to the regulatory genes (Figure 7A). There had 295 proteins associated with CLEC12A, 280 with DOK2, 271 with ARHGAP9, 184 with CD6, and 115 with C1orf162, respectively.

The Construction of the IncRNA-miRNA-mRNA ceRNA Network

To establish the IncRNA-miRNA-mRNA ceRNA network, we collected the interactions between miRNA and mRNA from the targetscan database, mircode database, and mirtarbase database, and miRNA and IncRNA from mircode databases. We identified 107 miRNAs that had an association with six significant mRNA. In addition, we collected the IncRNA related to the 107 miRNAs. Finally, we constructed a IncRNA-miRNA-mRNA ceRNA network including five IncRNA 21 miRNA, and six mRNA (Figure 7B), which were related to the five regulatory genes. Meanwhile, we used IncLocato to predict the subcellular localization of 5 IncRNAs in the construction of ceRNAs. The results showed that MIAT and HCG11 were predicted to be located in the nucleus, and DN-M3OS, HLA–DQB1–AS1, and LINC00426 were predicted to be located in the cytoplasm.

Drug-gene Interaction Analysis of Significant Regulatory Genes

The six significant regulatory genes were used as significant regulatory genes in a drug–gene interaction analysis. We identified a list of four potential drugs for CD6 and CLEC12A (Figure 7C).

Validation of the Gene Signature in Clinical Tissue Samples

To validate the expression of C1orf162 and CD6, we carried out Western blot and qRT-PCR using three and twelve pairs of HCC tissues and paracancerous tissues, respectively.
The results showed that their expression in the tumor tissues was downregulated compared with those in the paracancerous tissues (p<0.05) (Figure 8A-D).

Figure 6. The mutation and the pan-cancer analysis of the regulatory genes. A. The mutation frequency of the regulatory genes in HCC. The analysis of disease-free survival and progression-free survival of the regulatory gene BIN2, CD6, and DOK2 in pan-cancer analysis (B-D).
Discussion

In recent years, we have achieved a revolutionary breakthrough in therapy for most patients with advanced stages, such as the application of immune checkpoint inhibitor therapy targeting either the programmed cell death1 (PD-1)/programmed cell death ligand 1 (PD-L1) or cytotoxic T-lymphocyte (CTLA-4) pathways. Unfortunately, the clinical benefits of the objective response for patients were extremely limited. The liver is a unique organ that contains large populations of immune cells, some of which have strong antitumor potential. Nevertheless, the liver is also an organ of immune tolerance, which may lead to the suppression of the anticancer immune response and further influence the effect of immune immunotherapy. The survival of HCC patients was closely related to the increased innate immune and inflammatory gene expres-

Figure 7. A, the common genes using the intersections of genes between the DEGs of two clusters and the genes in the network. B, The lncRNA-miRNA-mRNA competitive endogenous RNA (ceRNA) network. C, The drug-gene interaction analysis of significant regulatory genes.
The association between immunosuppressive factors, the inflammation or immune-related gene expression, and the possible role of the neutrophils in the tumor microenvironment and prognosis value have been demonstrated preliminarily in HCC. Taken together, it is significant to make a reasonable classification and identify the significant regulatory factors to activate the anti-tumor function.

In our study, we first classified the HCC datasets downloaded from TCGA into two immune-related clusters based on the immune cell infiltration score and explored the difference in immune status and clinical prognosis between clusters. The immune status and clinical prognosis in cluster 2 were better than in cluster 1, which suggests that cluster 2 was an immune-hot phenotype compared with cluster 1. Based on those results, functional enrichment analysis based on the difference of the differential gene expression showed that immune-related pathways in cluster 2 were higher in cluster 1. Then, we identified the six significant regulatory factors, the analysis of which showed that the prognosis of patients in cluster 2 was better than that in cluster 1. Moreover, those genes were closely related to immune status, and their expression in cluster 2 was higher in cluster 1. In the single-cell RNA sequence analysis, the expression of the six genes was correlated to the immune cell. Both the analysis of patients from TCGA and the results of single-cell RNA sequence analysis indicated that the six genes were related to the immunology. We also identified potential drugs for CD6 and CLEC12A, which were helpful to explore the potential therapeutic drugs to treat HCC. Moreover, we constructed lncRNA-miRNA-mRNA interaction regulatory networks, and the lncRNA was located in the nucleus and cytoplasm. In addition, the result of the pan-cancer analysis showed those genes contributed to an important prognosis effect, which highlighted the potential antitumor effect. Finally, we validated the expression of genes in the experiment.

Our study was different from recent immune subtype reports on HCC. The study reported classified the HCC data sets into five clusters.
based on the expression of immune-related genes profiles and showed that patients from Cluster 3 had the worst prognosis and lower immune score. However, the immune score and prognosis in Cluster 4 and Cluster 2 were inconsistent. In our study, we identified the two subtypes for HCC and further verified the rational classification using bioinformatics analysis of immune status. We found that patients in cluster 2 were immune-hot subtype and had a better prognosis. In the analysis of the functional enrichment, the result showed that the DEGs were enriched in the immune-related pathways, including the GO, KEGG, and MsigDB analysis. In addition, we also collected significant regulatory factors, which were closely related to the immune status and could be regarded as possible immunotherapy targets.

Our current study showed that patients in cluster 2 had more types of immune cell infiltration. These results were verified using the CIBERSORT, TIMER, quanTIseq, EPIC, and xCell. The infiltration of CD4+ and B cell in cluster 2 in almost all analyses had a higher level in cluster 1. In the immune cell population, B cells are the main effector indirectly killing cancer cells. CCL19/CCR7 axis, CCL21/CCR7 axis, and CXCL13/CXCR5 axis are considered to be the main B cell-related pathways in tumor immune reaction. Since the TME plays a role in both beneficial and adverse consequences for tumor initiation, progression, and response to therapy, the immune cell composition of TME in the HCC has a major impact on cancer biology. However, the different immune cell subsets induce either inflammation or restrict antitumor immunity. The immune-related pathways in the Cluster 2 were higher enriched than in Cluster 1. The CYT score had a higher level in cluster 2 compared to cluster 1. In addition, the analysis of expression immunotherapy genes and immune-related genes showed the patients in cluster 2 had a better immunotherapy response and a better prognosis. In the functional enrichment analysis based on the differential gene expression, the enrichment of the immune-related pathways in cluster 2 was higher in cluster 1. These analyses indicated that cluster 2 was an immune-hot subtype with a better immunotherapy response.

Here, six genes were identified as significant regulatory factors in HCC. Moreover, we found that their expression in cluster 2 was higher compared to cluster 1, and patients in cluster 2 had a better prognosis than in cluster 1 in the survival analysis. We also verified the expression of the CD6 and C1orf162 in the experiment. BIN2, also called BRAP1, is a bona fide member of the N-BAR protein family, which can influence podosome formation, motility, and phagocytosis in leukocytes. In endometriosis, BIN2 is regarded as a novel Mφ2 macrophage-related biomarker. BIN2 was identified as the hub gene in HCC, and the expression in HCC tissues was higher than in normal tissues. Rho GTPase activating protein 9 (ARHGAP9), a member of the RhOGAP family, had a down-regulated level in HCC. However, ARHGAP9 can be identified as a potential tumor suppressor by affecting migration and invasion of HCC cells via regulating FOXJ2 and its target gene CDHT. In our study, we found that ARHGAP9 was closely associated with immune cell infiltration. In an analysis of the cSNP chip on hepatocellular carcinoma-related genes, the expression of DOK2 shows polymorphisms, and the DOK2 has a strong association with the immune score. The hypomethylated C1orf162 is associated with the mRNA expression level comparison between gastric cancer and normal gastric tissues and may be regarded as a novel marker. C-type lectin domain family 12 member A (CLEC12A), as an Ag delivery receptor, is broadly expressed by all human DC subsets and is an attractive candidate for in vivo boosting tumor-reactive T cell immunity in cancer patients. As shown in our study the CLEC12A had a strong association with the myeloid dendritic cell activated in HCC. The expression of CLEC12A also was downregulated in the HCC than in the normal tissues. In the analysis of the immune status, we found that the six regulatory genes had a close relationship with chemokines, HLA, receptors, immunostimulatory molecules, and immunosuppressive molecules, especially for CD6 and BIN2. The six genes had a different function in the tumorigenesis and development of HCC. In our study, the results will help provide a novel direction to study the HCC. Dendritic cells play an important role in initiating, regulating, and maintaining immune responses and inducing anti-tumor immune responses. The characteristics of relatively low cost, the easy obtainment, safety, and the absence of sequence exclusion and gene integration make the mRNA vaccine to be regarded as highly feasible for targeting tumor-specific antigens and promising immunotherapy strategies. Recently, the mRNA vaccine has presented significant potential in the treatment of COVID-19. Moreover, most patients with HCC had a background of viral hepatitis.
In our study, we found that CD6 and BIN2 were closely associated with the dendritic cells, which reflected the great potential in the development of the mRNA vaccine. In the single-cell RNA sequence analysis, we found that the expression of the six regulatory genes was a significant association with the immune cell, and the six genes in HCC were mainly expressed in the Myeloid cells, Mature B cells, and T/NK cells. Myeloid cells in tumor tissues consisted of a dynamic immune population characterized by a non-uniform phenotype and diverse functional activities. However, some of the most successful treatments have now been applied in the clinic and are being investigated in clinical trials, such as tumor-associated macrophages (TAMs) and tumor-associated neutrophils (TANs). Recent studies demonstrated the specificity and cytotoxicity of T and, NK cells towards tumor-specific or associated target antigens generated by genetic engineering of the immune cells to express a chimeric antigen receptor. Those indicated that the six genes may change the status of Mature B cells and T/NK cells to perform the anti-cancer property. Those genes in the pan-cancer analysis also affected the disease-free survival and progression-free survival in most cancers. Therefore, the identification of these significant immune-related genes may help upregulate their expression in immune-cold HCC and enhance the outcome of immunotherapy.

LncRNA has been regarded as a molecular sponge of miRNA to adsorb miRNA to remove the inhibition of miRNA on its target genes. The network formed by mRNAs, miRNAs, and LncRNAs can jointly play a role in regulating gene expression at the transcription or posttranscriptional stage. In our study, we constructed a significant regulatory factor-related ceRNA network and predicted the subcellular localization of the selected LncRNAs. The results showed that LncRNA was located in the LncRNA nucleus and cytoplasm. In the ceRNA networks, we found that CD6 was the core gene, and was related to almost all miRNA. mRNA vaccines have a great advantage to respond rapidly to the global explosion of the coronavirus disease 2019 (COVID-19) U.S., and the mRNA vaccine field will encompass a dramatic rise in the market value and will attract widespread interest in anticancer. FDA has recently approved prophylactic vaccines for the hepatitis B virus that can cause liver cancer. In future studies, the vaccine development of mRNA for HCC may offer another treatment.

Potential drugs identified by drug-gene interaction for CD6 and CLEC12A were classified into CAR-T cells targeting CLL1, itolizumab, oncolysin CD6, tepotitamab. Itolizumab, a humanized IgG1 kappa anti-CD6 monoclonal antibody, binds to domain 1 of human CD6, which is used to treat autoimmune and inflammatory diseases. In 2020, Itolizumab was approved for emergency use in India to treat cytokine release syndrome in COVID-19 patients with moderate to severe acute respiratory distress syndrome, and the results showed better survival and recovery benefits. ticOncolytic CD6 and CAR-T cells targeting CLL1 have a clear indication for the treatment of Hematologic tumors such as Leukemia, Lymphoid, Autoimmune Disease, and Lymphoma. Tepotitamab can bind to CLEC12A leading to a potent cytotoxic T-lymphocyte (CTL) response against CLEC12A-expressing tumor cells. All these drugs may provide promising prospects for the treatment of HCC.

Our study provides the potential therapeutic target for the clinical treatment through the identification of the six genes, which may provide new insights and novel directions to utilize and activate the immune response for HCC treatment. These can explore the direct antitumor function and cytotoxic capacities of the immune system. Firstly, we performed a comprehensive analysis of the six significant genes, which may provide promising prospects for the treatment of HCC. We identified the six genes in patients with HCC from TCGA and further verified them in the single-cell RNA sequence data. Moreover, we also validated the expression of the genes in the experiment. Secondly, we used the immune cell infiltration analysis, immune-related pathways, immune competence analysis, immune targeting gene analysis, and immune gene analysis to clarify the association between the six genes and immunology in immune tumor microenvironment analysis. Because hepatocellular carcinoma (HCC) is the second leading cause of cancer-related death worldwide, we further explored the universal significance of the six significant genes in the pan-cancer analysis, in which the results indicated the universal significance.

**Limitations**

Several limitations in this study should be mentioned. Firstly, we did only construct clusters for HCC from the TCGA database. Secondly, although we experiment to explain the expression of the six significant genes, our study lacks the elucidation between the six significant genes and immunology. In the future, we will do some ex-
Potential biomarkers for the prognosis and treatment of HCC immunotherapy

Experiments to solve the problems through RNA sequence in animals and treat the HCC based on the association between six genes and immunology. Therefore, our consequent works will do some basic experiments to verify the reasonability of the classification in HCC.

**Conclusions**

Collectively, we identified the two clusters based on immune cell infiltration. The analysis of immune status found that cluster 2 was an immune-hot subtype and had a better prognosis compared to cluster 1. Moreover, we collected six regulatory factors, which may be regarded as potential biomarkers for the prognosis and treatment of HCC immunotherapy.

**Conflict of Interest**
The Authors declare that they have no conflict of interests.

**Authors’ Contribution**
ZW designed the study. HL, QH, GZ, WW, DJ, and SJ collected the data, performed the analysis, and wrote the manuscript. HL and QH participated in the analysis and prepared the manuscript draft. GZ experimented. All authors approved the final manuscript.

**Informed Consent**
All patients provided written informed consent.

**Ethics Approval**
The studies were approved by the Medical Ethical Committee of the Beijing Institute of Radiation Medicine (20200924-34MT).

**Funding**
No funding was requested for this study.

**ORCID ID**
Zi-Hao Wang: 0000-0001-7065-3048.

**References**


Potential biomarkers for the prognosis and treatment of HCC immunotherapy


38) Hu B, Yang XB, Sang XT. Molecular subtypes based on immune-related genes predict the prognosis for hepatocellular carcinoma patients. Int Immunopharmacol 2021; 90: 107164.


