



# Pathological images for personal medicine in Hepatocellular carcinoma: Cross-talk of gene sequencing and pathological images

LI YANG<sup>1,2</sup>; KUN DENG<sup>3</sup>; ZHIQIANG MOU<sup>1,2</sup>; PINGFU XIONG<sup>1,2</sup>; JIAN WEN<sup>1,2</sup>; JING LI<sup>1,2,\*</sup>

<sup>1</sup> Department of General Surgery (Hepatobiliary Surgery), The Affiliated Hospital of Southwest Medical University, Luzhou, 646000, China

<sup>2</sup> Academician (Expert) Workstation of Sichuan Province, Luzhou, 646000, China

<sup>3</sup> Department of General Surgery (Gastrointestinal Surgery), The Affiliated Hospital of Southwest Medical University, Luzhou, 646000, China

**Key words:** Hepatocellular carcinoma, tumor environment, individualized medicine, predictive model, pathological model

**Abstract: Background:** Considering the great heterogeneity of Hepatocellular carcinoma (HCC), more accurate prognostic models are urgently needed. This paper combined the advantages of genomics and pathomics to construct a prognostic model. **Methods:** First, we collected data from hepatocellular carcinoma patients with complete mRNA expression profiles and clinical annotations from the TCGA database. Then, based on immune-related genes, we used random forest plots to screen prognosis-related genes and build prognostic models. Bioinformatics was used to identify biological pathways, evaluate the tumor microenvironment, and perform drug susceptibility testing. Finally, we divided the patients into different subgroups according to the gene model algorithm. Pathological models were constructed by obtaining HE-stained sections from TCGA in corresponding subgroups of patients. **Results:** In this study, we constructed a stable prognostic model that could predict overall survival in HCC patients. The signature consisted of six immune-related genes (*BX537318.1*, *TMEM147*, *CSPG4P12*, *AC015908.3*, *CEBPZOS*, and *SRD5A3*). We found increased levels of infiltration of immune cells in the tumor microenvironment in patients with low risk scores, indicating significant antitumor immunity and corresponding to better clinical outcomes. We then screened nine drugs that were more sensitive in the low-risk group than in the high-risk group. Finally, we addressed the complex cellular changes and phenotypic heterogeneity in the HCC microenvironment by combining genomics and pathomics analysis methods. **Conclusion:** Our study showed that the prognostic evaluation model of HCC based on the immune signaling pathway is feasible and provided a reference value for potential immunotherapy for HCC.

## Introduction

Globally, the incidence and mortality of primary liver cancer are the sixth and fourth highest, respectively, among which hepatocellular carcinoma (HCC) accounts for 75%–85% of primary liver cancer [1–3]. Liver is the sixth most common site of primary cancer in humans, often in the context of cirrhosis and inflammation. Cholangiocarcinoma is the second most common primary liver malignancy after HCC [4–6]. In recent decades, the incidence of cholangiocarcinoma has increased significantly, with a 5-year survival rate of <10% [7]. Although patients with early liver cancer can be surgically resected or given a liver graft, most patients with primary liver cancer are too advanced to be eligible for surgical resection. In addition, owing to its anatomical location and tissue, as well as its unique

metabolic and immunosuppressive environment, the liver is often colonized by cancer metastases from other organs [8,9]. In recent years, tumor immunotherapy has achieved good results in many cancers, but the therapeutic effect of immunotherapy in liver cancer has not been satisfactory [10,11]. Under physiological conditions, tumors accelerate cancer development by avoiding immune responses [12–14]. The key regulatory node of immunotherapy is the immunotumor microenvironment, which is divided into the T-cell infiltration-exclusion type, infiltration-inflammatory type and lymphoid structure infiltration type [13–15]. A better understanding of the immune microenvironment is an important prerequisite for immunity against liver cancer [16].

Tumors resist immune responses by activating immune checkpoints, such as programmed death-1 (PD-1) and its ligand PD-L [17–19], as well as cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) [20,21]. Immune checkpoint inhibition uses monoclonal antibodies targeting PD-1/PD-L1 and CTLA-4 to release pre-existing immunity, particularly effector CD8<sup>+</sup> T cells [13,22,23].

\*Address correspondence to: Jing Li, lijing3107623@aliyun.com

Received: 23 November 2022; Accepted: 28 December 2022



Immunotherapy has a positive effect on some patients but has no obvious effect on most patients [24,25]. Owing to the adaptability of tumors to different microenvironments, different individuals have different abilities to respond to treatment. Therefore, accurate immunotherapy and predictive biomarkers are important to achieve the best therapeutic effect and prolong the quality of life of patients. Exploring new immune-related prognostic markers is important for guiding the treatment process and prolonging the survival of patients with liver cancer.

Cancer diagnosis, prognosis, and prediction of response to therapy are often done using heterogeneous data sources, including histological sections, molecular profiles, as well as clinical data, such as patient age and comorbidities [26]. Histology-based subjective and qualitative analysis of the tumor microenvironment combined with quantitative examination with genomic testing is the standard of care for most cancers in modern clinical settings [27–29]. As the field of anatomic pathology moves from slides to digitized whole-slide images, it becomes possible to analyze pathomics and genomics in a comprehensive manner. Current modern sequencing technologies such as single-cell sequencing are capable of dissecting genomic information from individual cells in tumor specimens, and spatial transcriptomics and multiplex immunofluorescence are capable of spatially dissecting histological tissue and genomics together [30–33]. However, these techniques currently lack clinical penetration. The Cancer Genome Atlas (TCGA) contains genomically-paired whole-slide images, genotype and transcriptome data from cancer patients with ground truth survival and histological grading markers, providing a powerful tool for the combination of pathomics and genomics [34].

In our investigation, we analyzed liver cancer data from The Cancer Genome Atlas (TCGA) database, selected immune-related prognostic biomarkers, and then constructed predictive model. However, we have not stopped yet. Because pathological imaging plays an

important role in revealing the tumor microenvironment, it is limited by the complexity of previous multi-genomics studies and has not been fully applied to the analysis of liver cancer. With the rapid development of biomedical imaging applications in cancer, pathology combined with multi-genomics showed great application potential in liver cancer prediction. This is the first original study to combine pathological images and genetic groupings to complement each other.

**Methods**

Flowchart was found in Fig. 1.

*Data acquisition*

The TCGA (Cancer Genome Atlas) database was created by the National Cancer Institute and contains genomic, transcriptomic, proteomic, and methylation data from 20,000 primary cancers (<http://cancergenome.nih.gov/>). Transcriptomic data and corresponding clinical information were collected from a total of 368 HCC patients by screening samples that were not eligible for data or had missing clinical data.

*Single-sample gene set enrichment analysis (ssGSEA)*

Immune-related hallmark pathways were included in this study. The ssGSEA algorithm was used to score, and consensus cluster analysis was performed to identify immune-related patterns [35]. The R package ConsensusClusterPlus (Version 1.56.0) was used with key parameters, including maxK = 6 and repetition = 500 for stable identification [36]. A total of 368 patients with HCC were divided into three groups by consensus clustering. We further performed differential expression gene (DEG) analysis using the R package LIMMA (version 3.48.3) to compare pairwise expression of different patterns of genes using the LMFIT and EBayes functions to ensure accuracy. DEGs were selected according to the standard; the adjusted

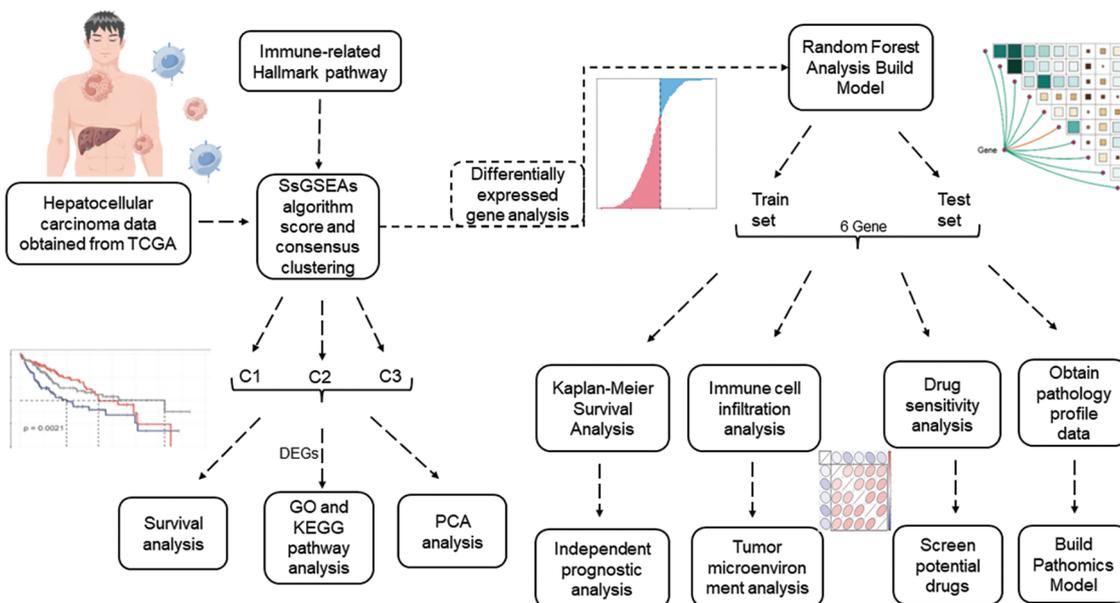


FIGURE 1. Flowchart.

$p$ -value was less than 0.001, and the absolute value of  $\log_{2}FC$  was more than 1.

#### *Immune-related signaling pathways*

Gene Ontology (GO) annotation analysis was performed using the R package clusterProfiler (version 4.0.5) with a false discovery rate (FDR) of less than 0.05 to determine significant enrichment [37]. GSEA Version 4.1.0 (Broad Institute, Cambridge, MA, USA) was used to identify a predefined genomic cohort  $p$ -value of less than 0.05, and an FDR of less than 0.25 for alterations between clusters 1-3 in a consensus cluster were considered statistically significant [38].

#### *Construction of genome models*

Based on the TCGA-LIHC cohort, we used random forest analysis to screen for genes with prognostic markers from the target gene pool and constructed a prognostic model. Next, we distinguished the high-risk group from the low-risk group according to the median risk score.

#### *Assessment of immune cell infiltration*

Two methods were used to calculate the immune infiltration score: ssGSEA and the xCell algorithm, which were visualized by stacking map correlation heat map and scatter plot, respectively.

#### *Extraction of pathological features*

The Cancer Genome Atlas (TCGA) contains genomics paired whole-slide images, genotype, and transcriptome data from cancer patients with ground truth survival and histological grading markers [34]. To observe the histological changes and pathological changes in patients at high and low risk of LIHC, we obtained hematoxylin and eosin (H&E) stained sections from TCGA from each LIHC patient, and the patients to whom each section belonged corresponded to the genomics data of LIHC patients. We used CellProfiler software (3.0.0) downloaded from the CellProfiler website ([www.cellprofiler.org](http://www.cellprofiler.org)), pathological images were analyzed, and pathomics feature extraction and data processing were performed using a special pipeline set by ourselves [39,40]. A workflow for LIHC tumor pathology feature extraction was established by adding a pre-programmed algorithm module to the pipeline. The pipeline is available from public data (<https://cellprofiler.org/published-pipelines>). The specific pipeline is set as follows: We construct eight Gaussian filters, and the extracted feature types include texture feature, wavelet feature and pipeline feature. After a series of superposition operations, the constructed pipeline can extract 6564 pathological and histological features in each pathological image. In pathological images of LIHC patients, we obtained HE staining images of five typical patients from previously identified high-risk and low-risk groups of LIHC patients for feature extraction for histopathology. For each patient with LIHC, the most typical 10 PATCH images ( $256 \times 256$  size) were selected and each pathological image was magnified 20-fold in WSI format to obtain the corresponding PATCH images. Initially, biopsy margins, tears, or gaps within the tissue were excluded and applied to masked RAW images. Masked

RAW images were then converted to grayscale to identify stained objects (DAB and nuclei). A threshold algorithm was applied to identify DAB and HTX (kernel) stained objects, respectively. Eventually, we obtained 50 images from each of the low-risk and high-risk groups for feature extraction and subsequent analysis.

#### *Construction of pathological model*

For 100 pathological images obtained from LIHC patients, we randomly selected 70 training datasets (39/31 = positive/negative) as pathomics models. The other 30 served as independent test datasets (11/19 = positive/negative). To eliminate imbalances in the training dataset, we used a variety of data normalization and regularization methods and analyzed the degree of dominance of the respective constructed models in terms of prediction level. A normalization method was applied to the feature matrix. For each feature vector, we calculated the number of l2 parameters and divided them by. The feature vector is then mapped onto the unit vector. For the normalization method, we chose the minmax, z-score, and mean methods. Because of the high dimensionality of the feature space, principal component analysis (PCA) and Pearson correlation coefficient dimensionality reduction were applied to the feature matrix. The feature vectors of the transformed feature matrices are independent of each other. Before building the model, we used ANOVA, KW, RFE, and mitigation methods to select pathomorphologic features. These methods are often used to explore salient features corresponding to labels. Ultimately, we used logistic regression to construct a pathological risk model for the selected optimal pathological features. The prediction performance of the constructed pathological feature model was analyzed and evaluated using receiver operating characteristic curves in the training, validation, cross-training, and cross-validation groups. In addition, we assessed the goodness-of-fit and clinical applicability of the pathologic signature model using the Hosmer–Lemeshow test.

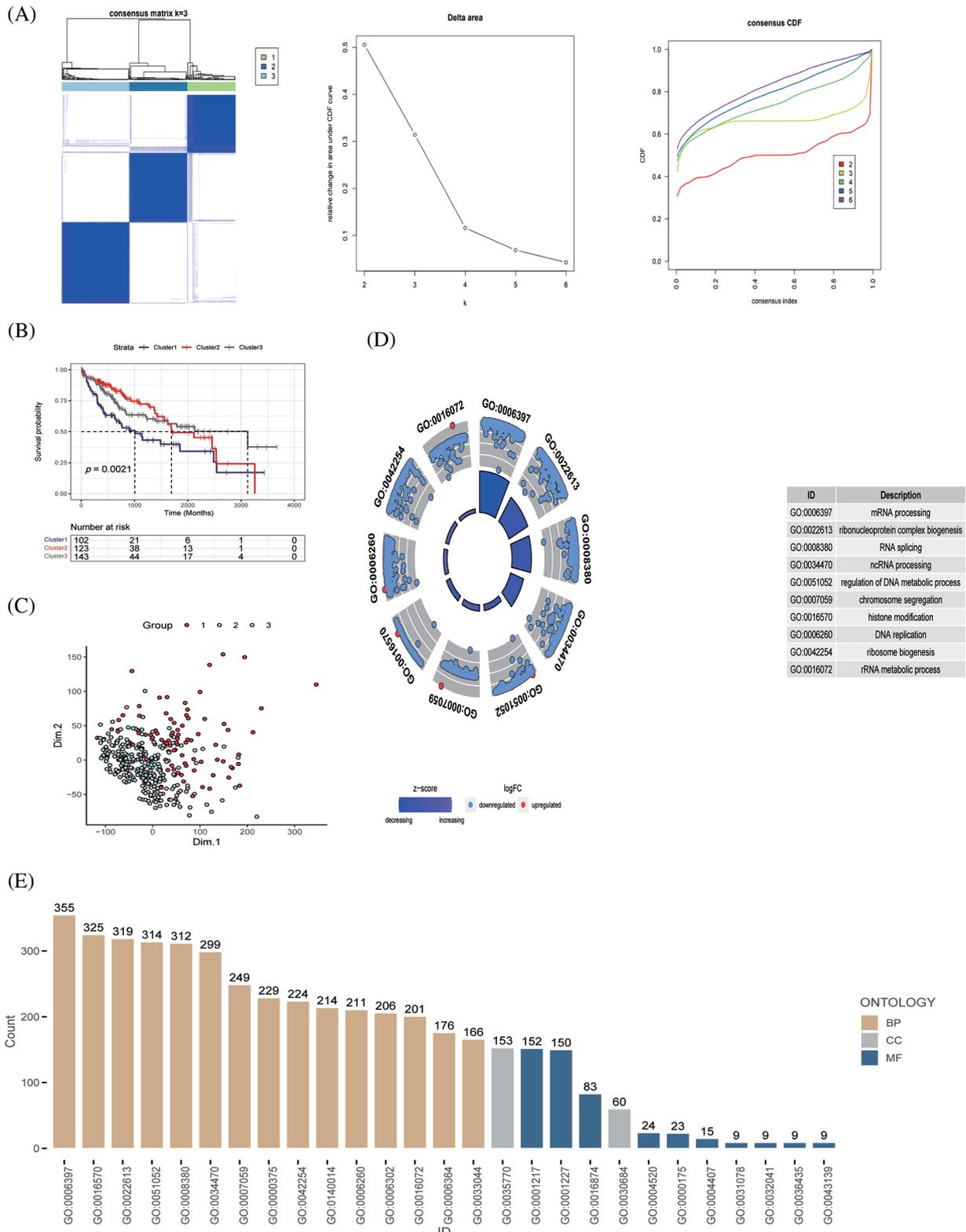
#### *Statistical analysis*

The R software was used for all analyses. Student's  $t$ -test was used for normally distributed continuous variables. Mann–Whitney U test was used for continuous variables that were not normally distributed. Limma package was used to analyze the differentially expressed threshold setting of  $\log_{2}FC > 0.5$ ,  $p < 0.05$ .

## **Results**

#### *Identification of immune pathway-related genes in HCC*

First, we used 368 patients with HCC to identify gene patterns related to immune pathways by ssGSEA algorithm scoring and consensus clustering analysis and set the K value of the consensus matrix to 3. The patients were then divided into three groups according to the optimal conditions, with significant differences between the groups (Fig. 2A). K–M analysis revealed significant differences in survival among the three groups; patients in cluster 1 had the worst survival



**FIGURE 2.** Identification of immune pathway-related genes and molecular subtype (A) Patients with HCC ( $n = 368$ ) were classified into three groups by consensus clustering ( $K = 3$ ); categorical evaluation of the three subsets formed after cluster analysis, (B) survival analysis of patients with differential subsets, and (C) distribution evaluation of subsets by PCA. (D, E) GO analysis of immune pathway-associated genes in liver cancer, discovering the cellular distribution of each gene and the functional association of gene expression.

outcomes (Fig. 2B). PCA also revealed remarkable differences in gene expression distribution among the three subsets, demonstrating the success of our classification approach (Fig. 2C). To further evaluate the differences among the three subsets, we evaluated the functional status and quality

of life of these genes *in vivo* using GO analysis based on the DEGs (Figs. 2D and 2E). From these results, we observed that the screened genes were highly correlated with the GO:0006397 pathway. These results suggest that based on the cohort information of the patients with HCC and the

expression levels of genes related to immune signaling pathways, the patients were effectively divided into different molecular subgroups by consensus clustering analysis, and the effectiveness of classification was confirmed by PCA analysis.

#### *Functional analysis of immune-related genes*

Through KEGG and functional enrichment analyses, we further confirmed the functional role and metabolically active pathways of the DEGs in tumor progression (143 genes). Fig. 3A displays that these DEGs were enriched in these pathways (cell cycle, DNA replication, herpes simplex virus 1 infection, and bacterial invasion of epithelial cells), the specific information was marked, and Fig. 3B showed the enriched gene number in the corresponding pathway. Subsequently, we further evaluated the relationship between these metabolic pathways and the genes of immune pathway, then found that hypertrophic cardiomyopathy, ribosome, retinol metabolism, complement and coagulation cascades, dilated cardiomyopathy, and the HCC immune model were significantly expressed (Fig. 3C). Fig. 3D represented an overview of the pathways involved in the genes.

#### *Identification of genes associated with HCC risk model*

Six genes of prognostic significance were selected from 143 genes in the TCGA-LIHC cohort according to a random forest algorithm: *BX537318.1*, *TMEM147*, *CSPG4P12*, *AC015908.3*, *CEBPZOS*, and *SRD5A3*, and prognostic models were constructed (Figs. 4A and 4B). Patients were divided into high-risk and low-risk groups according to the median risk score. We found that the risk score was independent of age and gender but correlated with T stage and clinical stage, and the risk value increased with disease progression (Fig. 4C). According to Kaplan–Meier (K–M) analysis, poor survival outcomes were evident for patients in the high-risk group (Fig. 4D). As the risk value increases, the patient's survival status becomes progressively worse (Fig. 4E).

#### *Preliminary evaluation of HCC risk model*

To further verify the effectiveness of our prognostic model, we needed to draw survival curves based on the other cohort (internal cohort verification), which was divided 8:2 into training and validation sets. K–M analysis and log-rank tests were used to compare the *p*-values of survival curves between the two groups (Figs. 4F and 4G). In both the training and independent validation sets, we found that patients with high-risk scores had poorer overall survival ( $p < 0.05$ ), indicating the validity and accuracy of our prognostic model.

#### *Evaluation of tumor immune microenvironment based on the two algorithms*

Subsequently, because tumor immune infiltration plays a key role in tumor development and progression, we compared the differences in infiltrating immune cells between high-risk and low-risk groups by MCPcount algorithm analysis. The results showed that neutrophil infiltration was higher in the low-risk group, indicating that neutrophils may play an important role

in anti-tumor immunity. Fig. 5B shows the correlation between the six prognostic genes. *AC015908.3* was negatively correlated with other genes, while other genes were positively correlated. Butterfly plots showed correlations between genes and immune cells (Fig. 5C). Fig. 5D shows the proportion of tumor cell infiltration in each patient. Bubble plots showed the association of six prognostic genes with immune cell infiltration (Fig. 5E). *TMEM147* was positively associated with fibroblast production, *CEBPZOS* with CD8 + T cells, and *CEBPZOS* and *TMEM147* with T cells (Fig. 5F).

The xCell algorithm was used to verify the above results. Similarly, patients in the low-risk group had an increased component of immune cell infiltration and an enhanced immuno-antitumor effect (Figs. 6A and 6B). Fig. 6C showed the percentage of immune cell infiltration in each tissue. *CEBPZOS* was positively correlated with GMP and fibroblasts, but negatively correlated with hematopoietic stem cell (HSC); *SRD5A3* was positively correlated with CLP and negatively correlated with HSC, and *TMEM137* was negatively correlated with regulatory T cells (Fig. 7A). The bubble graph shows the association of the six prognostic genes with immune cell infiltration (Fig. 7B).

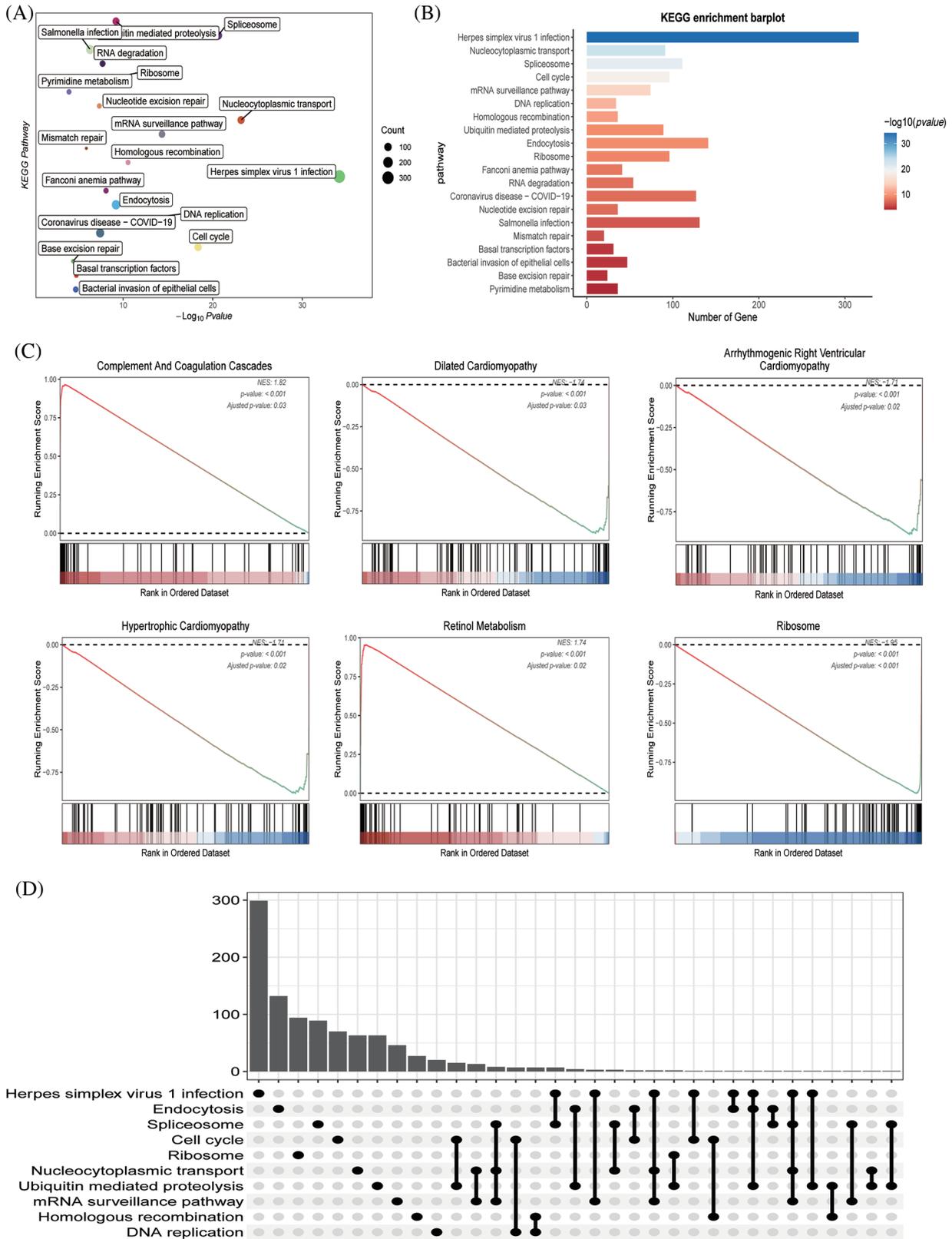
The ESTIMATE algorithm was used to calculate the ESTIMATEScore, StromalScore, and ImmuneScore of each sample, and the heat map showed the high- and low-risk grade distribution subgroup patients (Fig. 7C). Fig. 7D showed that risk is negatively correlated with the ESTIMATEScore, StromalScore, and ImmuneScore. Conclusively, we preliminarily evaluated the effectiveness of the prognostic model by drawing survival curves and correlation verification, laying a solid foundation for further verification.

#### *Drug sensitivity analysis*

To further evaluate the drug sensitivity of the immune-signaling pathway-based prognostic assessment model for HCC, we introduced drug-calculated sensitivity scores from the GDSC database to screen potential therapeutic agents for HCC based on prognostic genes. We showed the susceptibility of different drugs to six differential genes in the form of a heatmap (Fig. 8A). Finally, we screened nine drugs that were more sensitive in the low-risk group than in the high-risk group, including PD173074, PCI-34051, IWP-2, linsitinib, VE821, PD173074, PCI-34051, IWP-2, VE821, AZD4547, gefitinib, nelarabine, and AT13148 (Fig. 8B). In conclusion, our prognostic evaluation model for HCC based on immune signaling pathways has good performance for drug prediction.

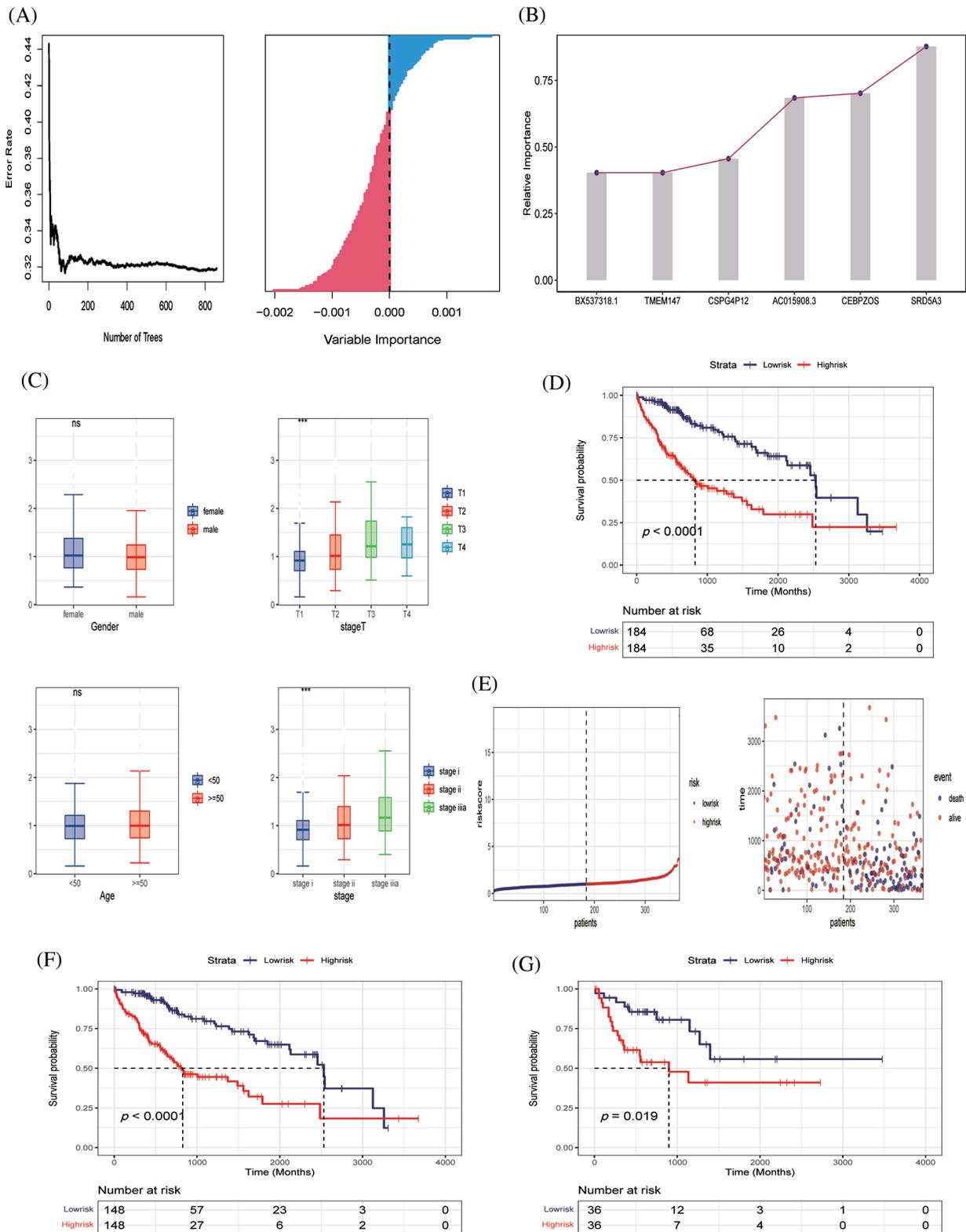
#### *Construction of pathological model*

The HE-stained sections corresponding to the high-risk and low-risk groups of patients with LIHC were screened against the pathology database; the specific pathological images and the region of interest patches are outlined in Fig. 9A. We obtained multiple pathomic features through a variety of pathology section feature extraction pipelines, including those within the DNA and skeletal regions, as shown in Fig. 9B. The weights occupied by the selected pathomic features in the machine learning model are shown in



**FIGURE 3.** The functional assessment (A) KEGG pathway analysis of immune pathway-associated genes in liver cancer. (B) The number of genes involved in the enrichment pathway. (C) KEGG analysis revealed that metabolic and tissue developmental elements, including hypertrophic cardiomyopathy, ribosome, retinol metabolism, complement and coagulation cascades, dilated cardiomyopathy and arrhythmogenic right ventricular cardiomyopathy, were significantly associated with the immune pathway-associated genes. (D) The UpSet graph showing the genes involved in the pathway.

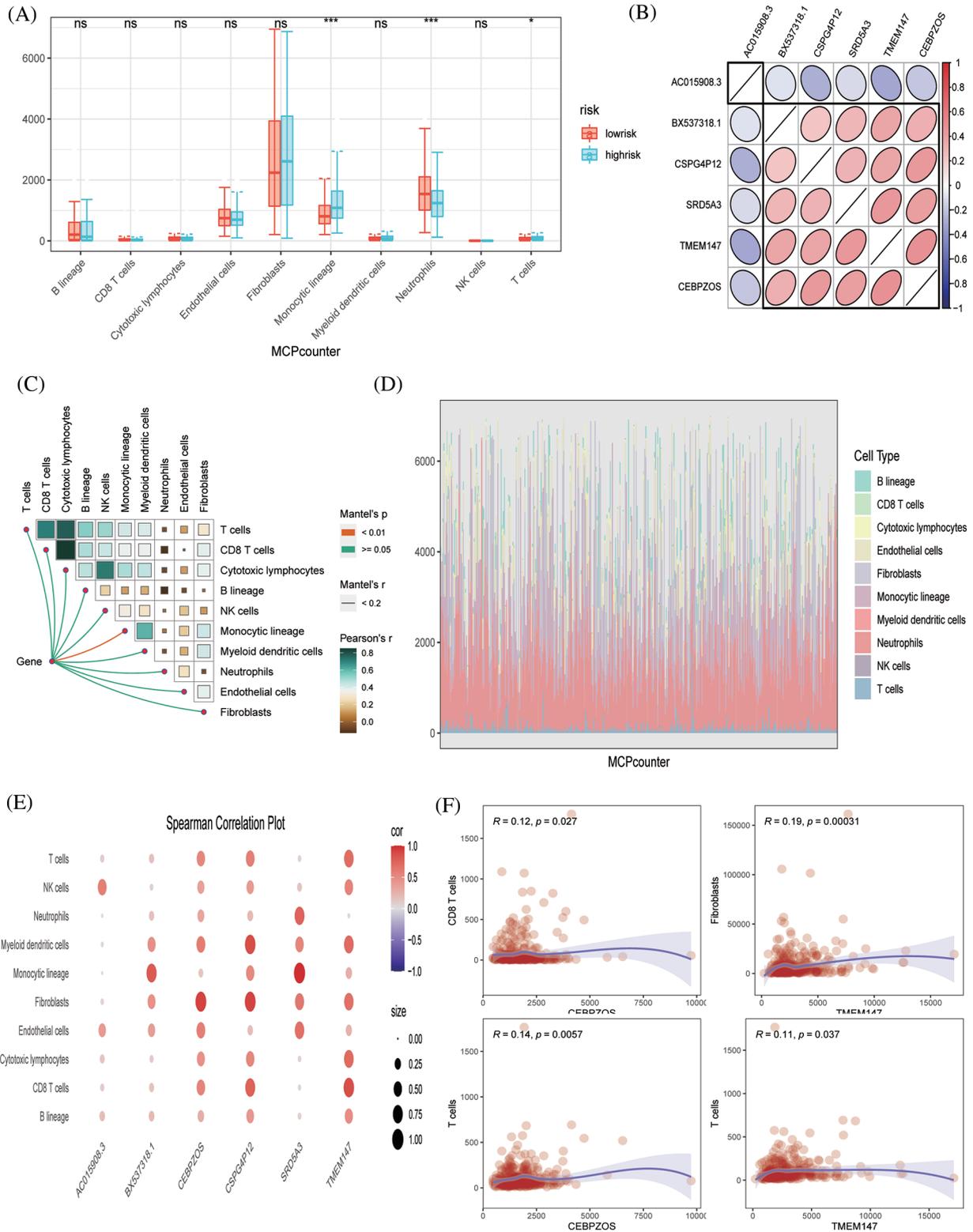
Fig. 9C. Differences in the area under the curve (AUC) values were similarly generated for different numbers of features in the same model (Fig. 9D). The differences in model prediction performance due to methodological differences between steps in the pipeline constructed through multiple normalization, feature selection, and



**FIGURE 4.** Construction of predictive model (A, B) Six genes of prognostic significance were selected from 143 genes based on LASSO regression to construct predictive models. (C) Risk values for patients with different clinical parameters, including gender, stage T, age, stage. (D) K-M survival analysis curves of TCGA-LIHC cohort. (E) Survival time and survival status distribution between high- and low-risk groups of the prediction model in the TCGA-LIHC cohort. (F, G) K-M survival analysis in training and validating cohorts.

dimensionality reduction methods to filter the optimal model are shown in Figs. 9E and 9F. The machine learning model obtained by the following method was selected to achieve the best AUC value and was therefore considered to be the best model for pathomics model

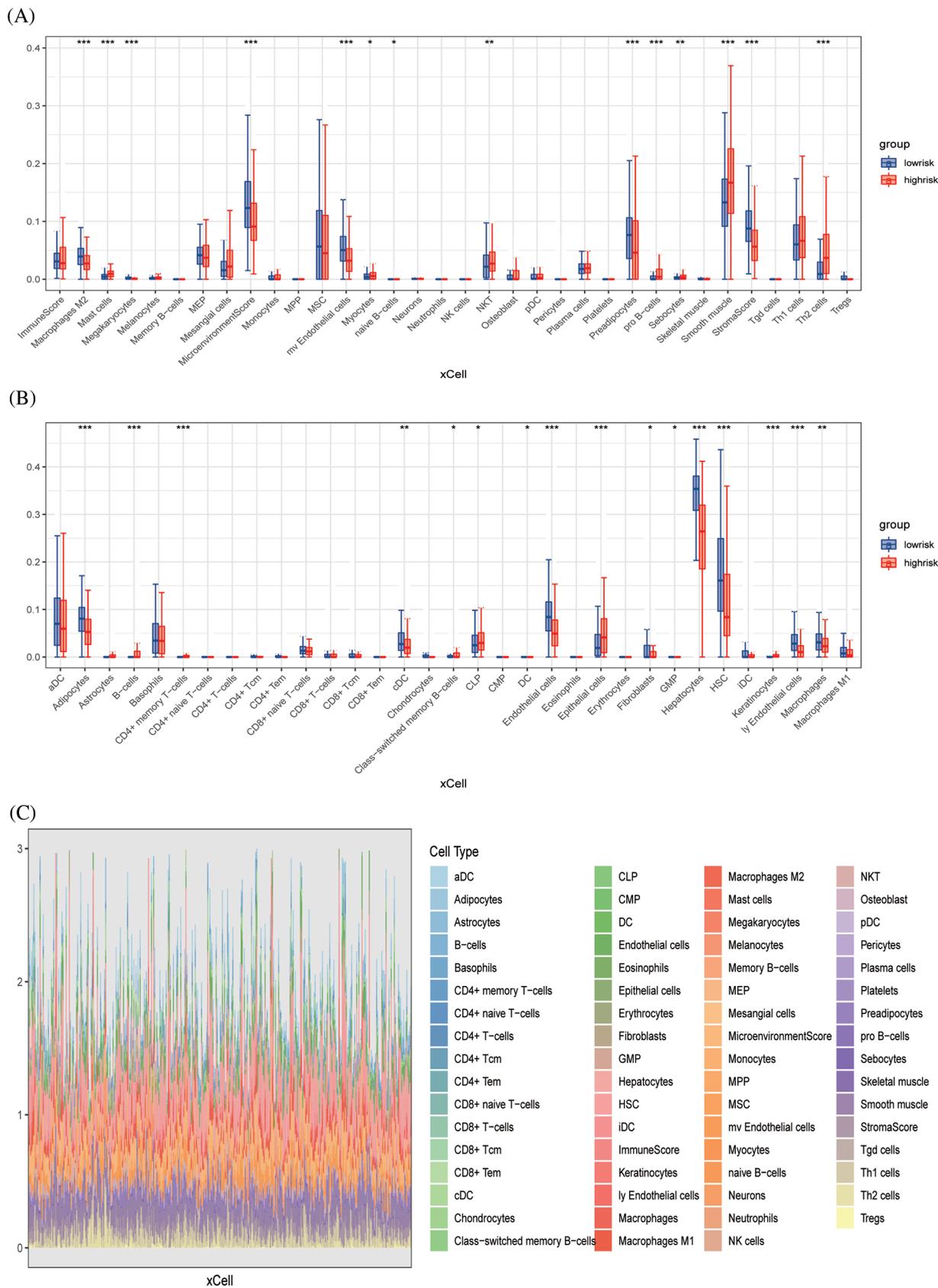
construction and optimal pipeline formation. The feature vectors of the transformed feature matrix were independent of each other. Before building the model, we used Relief to select the features. Relief selects a subset and recursively finds the relative features according



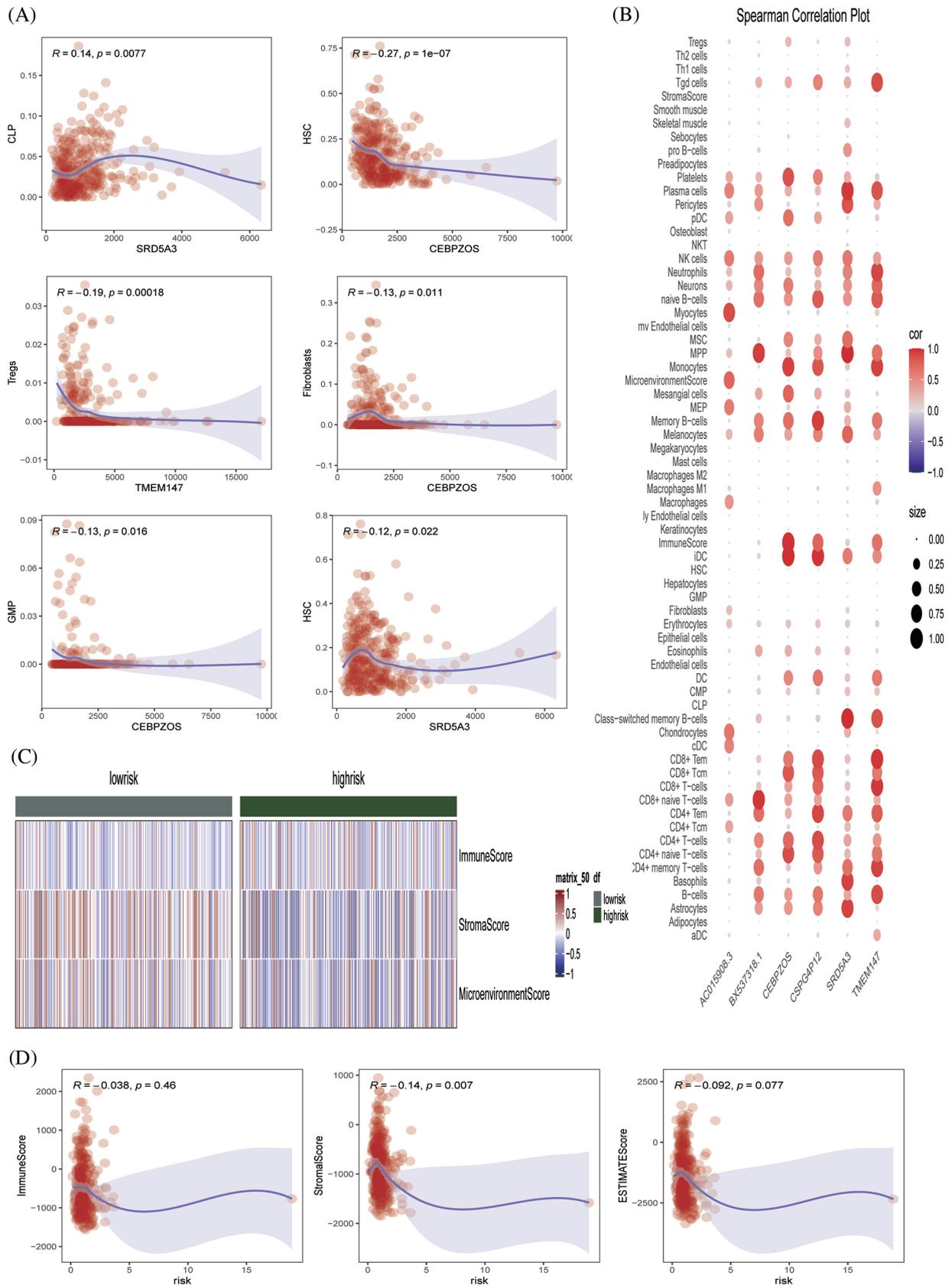
**FIGURE 5.** Immune infiltration evaluation (A) MCPcount algorithm constructed a box line plot of the difference in infiltration levels of several immune cells between the low-risk and high-risk groups in the predictive model. (B) Spearman correlation plots displaying the relationship among six prognostic genes. (C) Correlation of genes and immunocytes. (D) Stacked plots of the distribution of 10 major immune cells in the low-risk and high-risk groups. (E) Spearman correlation plots further visualized and clarified the relationship between the six genes and the specific cell types of immune infiltration. (F) The scatterplot showing the relationship between specific genes and immune cells.

to the label. Logistic regression with the LASSO constraint was used as the classifier. Logistic regression with LASSO constraint is a linear classifier based on logistic, and the hyper-parameters were set according to the model performance on the validation dataset. The AUC and

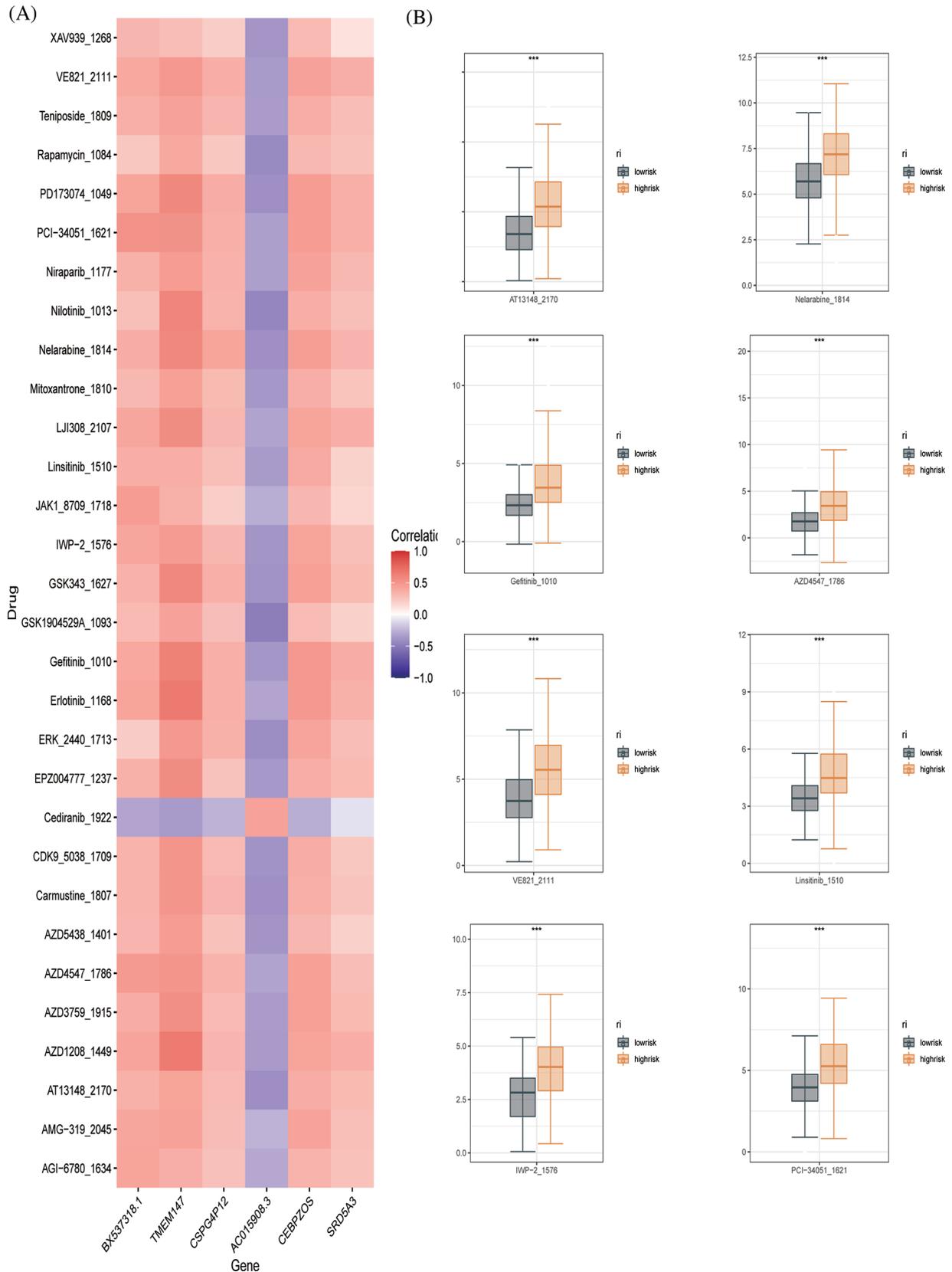
accuracy of the model were 0.852 and 0.867, respectively, for the test dataset (Fig. 9G). In the H-L test (Fig. 9H), we also demonstrated the excellent goodness-of-fit of the constructed LIHC pathomics model, demonstrating its clinical applicability.



**FIGURE 6.** Immune infiltration validation (A, B) Boxplots showing differences in immune cell infiltration between high- and low-risk groups. (C) Stacked plots of the distribution of 50 major immune cells in the high- and low-risk groups.



**FIGURE 7.** Immune infiltration validation (A) Scatterplot showing the relationship between specific genes and immune cells. (B) Spearman correlation plots further visualized and clarified the relationship between the six genes and the specific cell types of immune infiltration. (C) Heat map showing the immune scores of patients in the high- and low-risk groups. (D) Scatterplot showing the relationship between score and ImmuneScore, StromaScore, and ESTIMATE score.

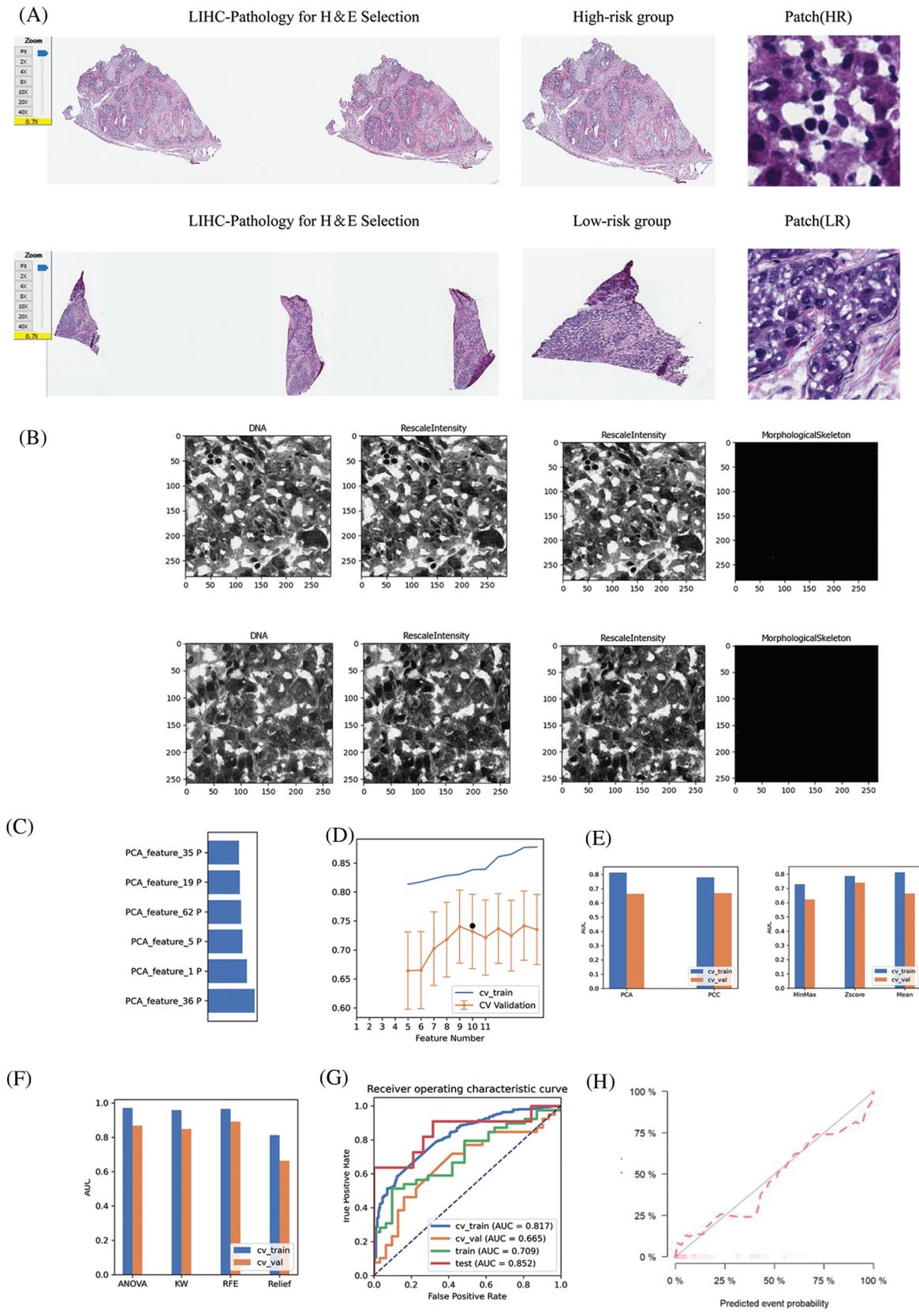


**FIGURE 8.** Drug sensitivity evaluation (A) Heat map displaying the correlation between six genes and therapeutic sensitivity of multiple chemotherapeutic drugs. (B) Box line plots showing significant therapeutic sensitivity between high- and low-risk subgroups for PD173074, PCI-34051, IWP-2, Linsitinib, VE821 AZD4547, Gefitinib, Nelarabine, and AT13148.

**Discussion**

Immunotherapy has a wide application prospect for cancer, but recent studies and clinical guidance have shown that its

efficacy in liver cancer is much lower than that in other cancers. From the perspective of multi-genomics, we identified prognostic markers of HCC based on immune signaling pathways and explored the direct or indirect effect



**FIGURE 9.** Pathomic marker mining and risk prediction model construction for LIHC. (A) HE-stained sections corresponding to patients with LIHC in the high- and low-risk groups screened in the pathology section repository; specific pathological images and patches outlined in the region of interest; (B) DNA, RescaleIntensity intensity analysis maps obtained from pathomics analysis. (C) Types of pathomic features constituting the pathomic risk model and their respective weights in the final model. (D) Linear analysis plot of the effect of the number of pathomic features on the AUC values of the final model. (E) Bar graph showing the effect of the selected dimensionality reduction method and data normalization method on the predictive efficacy of the pathomic risk model. (F) Analysis of the effect of different pathomic feature selection methods on the prediction of the pathomic risk model accuracy. (G) Histogram of the impact analysis of the constructed pathomics models on prediction ROC curves (training, validation, CV-training, and CV-validation groups, respectively). (H) Goodness-of-fit test curves for pathomics model.

of prognostic markers of HCC and immunotherapy to provide more reference for the potential application of effective immunotherapy.

Significant progress has also been made in immunotherapy for liver cancer. Studies have reported that nivolumab against PD-1, a fully human immunoglobulin G4 monoclonal anti-PD-1 antibody, was first tested for safety and efficacy in a phase 1/2 study of more than 250 patients with advanced HCC [41]. The expanded cohort confirmed a good safety profile, with 19% of grades 3 and 4 treatment-related adverse events consisting primarily of transient laboratory abnormalities. Nivolumab was subsequently granted accelerated approval by the US Food and Drug Administration in September 2017 for the treatment of patients with advanced HCC previously treated with sorafenib [42,43]. Several other monotherapy immune checkpoint inhibitors targeting the PD-1/PD-L1 axis, including atezolizumab (NCT04157985), durvalumab (NCT03847428), toripalimab (NCT03949231), and tislelizumab (NCT03412773), also showed an immune response against liver cancer [10,44–48]. Cell-based immunotherapies, such as tumor-infiltrating lymphocytes or cytokine-induced killer cells (CIKs) for adoptive cell transfer, have potent antitumor effects with little cytotoxicity to normal cells. The efficacy and safety of activated CIKs in adjuvant therapy were evaluated in 230 patients with HCC, who underwent surgical resection, radiofrequency ablation (RFA), or percutaneous ethanol injection [49]. When there is a high endogenous tumor antigen load in the primary tumor, immunotherapy in the neoadjuvant setting can enhance T-cell priming and potentially eliminate micrometastases, which may be the source of postoperative tumor recurrence [47,50–52].

In this study, we first performed differential gene analysis based on immune signaling pathways in 368 patients with HCC from the TCGA-LIHC cohort. By consensus cluster recognition, we divided patients into three clusters and further analyzed the expression of differential genes. GO and KEGG functional annotations were used to demonstrate that differential genes shared specific functional processes with immune signaling pathways. Six genes with prognostic significance, including *BX537318.1*, *TMEM147*, *CSPG4P12*, *AC015908.3*, *CEBPZOS* and *SRD5A3*, were selected from 143 genes according to random forest algorithm to construct a prognostic model, and the whole set was cut into 8:2—training set and validation set cohorts. The high- and low-risk groups were distinguished by the median risk score, and the two groups had significant prognostic differences. The ssGSEA and xCell algorithms were also used to calculate the immune infiltration score, which was visualized by a stacking correlation heat map and scatter plot, respectively.

Studies have shown that *TMEM147* interacts with the lamin B receptor in the endoplasmic reticulum, affecting the expression level and localization of the receptor [53], and *TMEM147* can also promote the proliferation of prostate cancer [54]. Studies have suggested that *CEBPZOS* is a prognostic marker of liver cancer related to energy metabolism [55]. Loss or overexpression of *SRD5A3* is closely related to breast cancer and loss of glycosylation

function [56–58]. However, studies of these genes in liver cancer are relatively rare. We demonstrated that *SRD5A3* plays a critical role in HCC progression.

In recent years, more and more researchers have established cancer prognostic models based on novel cell death mechanisms, but they are missing in clinical applications [59]. Linking pathomics and genomics is a whole new direction, and this approach links the performance of models from theory to practice and is essential for performance validation before clinical application [27,60]. Recently, an increasing number of researchers have linked pathomics and genomics. In one study, researchers revealed the landscape of m6A methylation modification patterns in bladder cancer by radiogenomics mapping by linking pathomics and genomics [61]. Another study validated this approach using glioma and clear cell renal cell carcinoma datasets from The Cancer Genome Atlas (TCGA), which contains paired whole-slide images, genotype, and transcriptome data with survival and histological grading markers, providing strong help for pathomics and genomics linkages [27]. In our study, we combined the model with clinical findings by obtaining pathological patterns corresponding to patients in the high and low model groups by TCGA. For the future clinical treatment, we provide a reliable route, based on our model to diagnose the patient's risk, through clinical biopsy to determine the patient's disease progression, and finally based on the risk score, and finally choose the corresponding drug for treatment.

In summary, we constructed a stable prognostic marker that can predict the overall survival of patients with liver cancer by analyzing HCC-related data in the TCGA database. In addition, we used a combination of bioinformatics and pathomics analyses to elucidate the complex cellular and phenotypic heterogeneity in the HCC ecosystem; the information obtained from these studies is critical for the development of successful therapies. The complex timing of HCC with diverse immune cell subsets and extensive tumor-immune cell crosstalk also emphasizes the need for combination therapeutic approaches targeting these components. As a next step, we will combine experimental and clinical data to further clarify the clinical feasibility of this prognostic marker. Immunotherapy has great potential in the early stages of the disease, and several trials investigating the effectiveness of immune-based approaches in neoadjuvant and adjuvant settings are ongoing. Finally, the development of biomarkers that can effectively predict the response to immunotherapy is essential for identifying optimal therapeutic targets and selecting the appropriate treatment for each patient.

Our study still has some limitations. Our study is based on data from the TCGA platform linking genomic and pathological group outcomes in patients, which leads to challenging validation of our model. Although we have an internal validation set and cross-validation, it is better if there are external validations that can satisfy both genomic and pathological groups. In addition, our model also has its certain advantages. Compared with the previous conventional model, the model combined with clinicopathologic pattern has more application conditions,

and is no longer a blank talk for predicting the prognosis of patients. It can find the differences between patients from the pathological point of view.

Conclusively, our study shows that the prognostic assessment model of HCC based on the immune signaling pathway is feasible and provides a reference value for potential immunotherapy for HCC treatment.

**Acknowledgement:** We sincerely appreciate all the members who participated in data collection and analysis.

**Funding Statement:** The authors received no specific funding for this study.

**Author Contributions:** Yang-Li was responsible for literature research, data screening, analysis and writing. Li-Jing was responsible for the whole article.

**Data Availability Statement:** All datasets generated for this study are included in the public database.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. et al. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6), 394–424. DOI 10.3322/caac.21492.
- Dulskas, A., Kavaliauskas, P., Zagminas, K., Jancoriene, L., Smailyte, G. (2021). Trends in incidence and mortality of primary liver cancer in lithuania 1998-2015. *International Journal of Environmental Research and Public Health*, 18(3), 1191. DOI 10.3390/ijerph18031191.
- Burton, A., Tataru, D., Driver, R. J., Bird, T. G., Huws, D. et al. (2021). Primary liver cancer in the UK: Incidence, incidence-based mortality, and survival by subtype, sex, and nation. *JHEP Reports*, 3(2), 100232. DOI 10.1016/j.jhepr.2021.100232
- Labib, P. L., Goodchild, G., Pereira, S. P. (2019). Molecular pathogenesis of cholangiocarcinoma. *BMC Cancer*, 19(1), 185. DOI 10.1186/s12885-019-5391-0.
- Razumilava, N. (2014). Gores GJ: Cholangiocarcinoma. *The Lancet*, 383(9935), 2168–2179. DOI 10.1016/S0140-6736(13)61903-0.
- Goyal, L., Kongpetch, S., Crolley, V. E., Bridgewater, J. (2021). Targeting FGFR inhibition in cholangiocarcinoma. *Cancer Treatment Reviews*, 95, 102170. DOI 10.1016/j.ctrv.2021.102170.
- Rizvi, S., Khan, S. A., Hallemeier, C. L., Kelley, R. K., Gores, G. J. (2018). Cholangiocarcinoma-evolving concepts and therapeutic strategies. *Nature Reviews Clinical Oncology*, 15(2), 95–111. DOI 10.1038/nrclinonc.2017.157.
- Li, X., Ramadori, P., Pfister, D., Seehawer, M., Zender, L. et al. (2021). The immunological and metabolic landscape in primary and metastatic liver cancer. *Nature Reviews Cancer*, 21(9), 541–557. DOI 10.1038/s41568-021-00383-9.
- González-González, M., Gutiérrez, M. L., Sayagués, J. M., Muñoz-Bellvis, L., Orfao, A. (2021). Genomic profiling of sporadic liver metastatic colorectal cancer. *Seminars in Cancer Biology*, 71, 98–108. DOI 10.1016/j.semcancer.2020.05.013.
- El-Khoueiry, A. B., Sangro, B., Yau, T., Crocenzi, T. S., Kudo, M. et al. (2017). Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): An open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *The Lancet*, 389(10088), 2492–2502. DOI 10.1016/S0140-6736(17)31046-2.
- Ye, F., Liang, Y., Hu, J., Hu, Y., Liu, Y. et al. (2021). DNA methylation modification map to predict tumor molecular subtypes and efficacy of immunotherapy in bladder cancer. *Frontiers in Cell and Developmental Biology*, 9, 760369. DOI 10.3389/fcell.2021.760369.
- Zhang, Y., Zhang, Z. (2020). The history and advances in cancer immunotherapy: Understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. *Cellular & Molecular Immunology*, 17(8), 807–821. DOI 10.1038/s41423-020-0488-6.
- Rizvi, S., Wang, J., El-Khoueiry, A. B. (2021). Liver cancer immunity. *Hepatology*, 73(Suppl 1), 86–103. DOI 10.1002/hep.31416.
- Yang, Y. (2015). Cancer immunotherapy: Harnessing the immune system to battle cancer. *The Journal of Clinical Investigation*, 125(9), 3335–3337. DOI 10.1172/JCI83871.
- Gajewski, T. F., Schreiber, H., Fu, Y. X. (2013). Innate and adaptive immune cells in the tumor microenvironment. *Nature Immunology*, 14(10), 1014–1022. DOI 10.1038/ni.2703.
- Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F. et al. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nature Medicine*, 24(5), 541–550. DOI 10.1038/s41591-018-0014-x.
- Han, Y., Liu, D., Li, L. (2020). PD-1/PD-L1 pathway: Current researches in cancer. *American Journal of Cancer Research*, 10(3), 727–742.
- Ai, L., Xu, A., Xu, J. (2020). Roles of PD-1/PD-L1 pathway: Signaling, cancer, and beyond. *Advances in Experimental Medicine and Biology*, 1248, 33–59. DOI 10.1007/978-981-15-3266-5\_3.
- Ghosh, C., Luong, G., Sun, Y. (2021). A snapshot of the PD-1/PD-L1 pathway. *Journal of Cancer*, 12(9), 2735–2746.
- Rowshanravan, B., Halliday, N., Sansom, D. M. (2018). CTLA-4: A moving target in immunotherapy. *Blood*, 131(1), 58–67. DOI 10.1182/blood-2017-06-741033.
- Yuan, Q., Ren, J., Li, L., Li, S., Xiang, K. et al. (2021). Development and validation of a novel N6-methyladenosine (m6A)-related multi-long non-coding RNA (lncRNA) prognostic signature in pancreatic adenocarcinoma. *Bioengineered*, 12(1), 2432–2448. DOI 10.1080/21655979.2021.1933868.
- Topalian, S. L., Drake, C. G., Pardoll, D. M. (2015). Immune checkpoint blockade: A common denominator approach to cancer therapy. *Cancer Cell*, 27(4), 450–461. DOI 10.1016/j.ccell.2015.03.001.
- Yuan, Q., Ren, J., Chen, X., Dong, Y., Shang, D. (2022). Contributions and prognostic performances of m7G RNA regulators in pancreatic adenocarcinoma. *Chinese Medical Journal*, 135(17), 2101–2103. DOI 10.1097/CM9.0000000000002179.
- Zimmermann, P., Curtis, N. (2019). Factors that influence the immune response to vaccination. *Clinical Microbiology Reviews*, 32(2), e00084. DOI 10.1128/CMR.00084-18.
- Baxter, D. (2014). Active and passive immunization for cancer. *Human Vaccines & Immunotherapeutics*, 10(7), 2123–2129. DOI 10.4161/hv.29604.

26. Bruix, J., Reig, M., Sherman, M. (2016). Evidence-based diagnosis, staging, and treatment of patients with hepatocellular carcinoma. *Gastroenterology*, 150(4), 835–853. DOI 10.1053/j.gastro.2015.12.041.
27. Chen, R. J., Lu, M. Y., Wang, J., Williamson, D. F., Rodig, S. J. et al. (2020). Pathomic fusion: An integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Transactions on Medical Imaging*, 41(4), 757–770. DOI 10.1109/TMI.2020.3021387.
28. Wen, P. Y., Kesari, S. (2008). Malignant gliomas in adults. *New England Journal of Medicine*, 359(5), 492–507. DOI 10.1056/NEJMra0708126.
29. Nayak, L., Lee, E. Q., Wen, P. Y. (2012). Epidemiology of brain metastases. *Current Oncology Reports*, 14(1), 48–54. DOI 10.1016/j.nec.2020.06.001.
30. Puchalski, R. B., Shah, N., Miller, J., Dalley, R., Nomura, S. R. et al. (2018). An anatomic transcriptional atlas of human glioblastoma. *Science*, 360(6389), 660–663. DOI 10.1126/science.aaf2666.
31. Jackson, H. W., Fischer, J. R., Zanotelli, V. R., Ali, H. R., Mechera, R. et al. (2020). The single-cell pathology landscape of breast cancer. *Nature*, 578(7796), 615–620. DOI 10.1038/s41586-019-1876-x.
32. Schapiro, D., Jackson, H. W., Raghuraman, S., Fischer, J. R., Zanotelli, V. R. et al. (2017). HistoCAT: Analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nature Methods*, 14(9), 873–876. DOI 10.1038/nmeth.4391.
33. Abdelmoula, W. M., Balluff, B., Englert, S., Dijkstra, J., Reinders, M. J. et al. (2016). Data-driven identification of prognostic tumor subpopulations using spatially mapped t-SNE of mass spectrometry imaging data. *Proceedings of the National Academy of Sciences of the United States of America*, 113(43), 12244–12249. DOI 10.1073/pnas.1510227113.
34. Tomczak, K., Czerwińska, P., Wiznerowicz, M. (2015). Review the cancer genome atlas (TCGA): An immeasurable source of knowledge. *Contemporary Oncology/Współczesna Onkologia*, 2015(1), 68–77. DOI 10.5114/wo.2014.47136.
35. Yi, M., Nissley, D. V., McCormick, F. (2020). Stephens RM: SsGSEA score-based Ras dependency indexes derived from gene expression data reveal potential Ras addiction mechanisms with possible clinical implications. *Scientific Reports*, 10(1), 1–16. DOI 10.1038/s41598-020-66986-8.
36. Wilkerson, M. D., Hayes, D. N. (2010). ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics*, 26(12), 1572–1573. DOI 10.1093/bioinformatics/btq170.
37. Consortium, G. O. (2004). The gene ontology (GO) database and informatics resource. *Nucleic Acids Research*, 32(suppl\_1), D258–D261. DOI 10.1093/nar/gkh036.
38. Subramanian, A., Kuehn, H., Gould, J., Tamayo, P., Mesirov, J. P. (2007). GSEA-P: A desktop application for Gene Set Enrichment Analysis. *Bioinformatics*, 23(23), 3251–3253. DOI 10.1093/bioinformatics/btm369.
39. McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B. A. et al. (2018). CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biology*, 16(7), e2005970. DOI 10.1371/journal.pbio.2005970.
40. Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H. et al. (2006). CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, 7(10), 1–11. DOI 10.1186/gb-2006-7-10-r100.
41. Finkelmeier, F., Waidmann, Q., Trojan, J. (2020). Nivolumab for the treatment of hepatocellular carcinoma. *Expert Opinion on Biological Therapy*, 20(7), 687–693. DOI 10.1080/14737140.2018.1535315.
42. Finkelmeier, F., Waidmann, O., Trojan, J. (2018). Nivolumab for the treatment of hepatocellular carcinoma. *Expert Review of Anticancer Therapy*, 18(12), 1169–1175. DOI 10.1080/14737140.2018.1535315.
43. Koppolu, V., Rekha Vasigala, V. K. (2018). Checkpoint immunotherapy by nivolumab for treatment of metastatic melanoma. *Journal of Cancer Research and Therapeutics*, 14(6), 1167–1175. DOI 10.4103/jcrt.JCRT\_1290\_16.
44. Zhu, A. X., Finn, R. S., Edeline, J., Cattan, S., Ogasawara, S. et al. (2018). Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): A non-randomised, open-label phase 2 trial. *The Lancet Oncology*, 19(7), 940–952. DOI 10.1016/S1470-2045(18)30351-6.
45. Finn, R. S., Qin, S., Ikeda, M., Galle, P. R., Ducreux, M. et al. (2020). Atezolizumab plus bevacizumab in unresectable hepatocellular carcinoma. *The New England Journal of Medicine*, 382(20), 1894–1905. DOI 10.1056/NEJMoa1915745.
46. Paz-Ares, L., Dvorkin, M., Chen, Y., Reinmuth, N., Hotta, K. et al. (2019). Durvalumab plus platinum-etoposide versus platinum-etoposide in first-line treatment of extensive-stage small-cell lung cancer (CASPIAN): A randomised, controlled, open-label, phase 3 trial. *The Lancet*, 394(10212), 1929–1939. DOI 10.1016/S0140-6736(19)32222-6.
47. Shi, L., Shang, X., Nie, K., Lin, Z., Zheng, M. et al. (2021). Identification of potential crucial genes associated with the pathogenesis and prognosis of liver hepatocellular carcinoma. *Journal of Clinical Pathology*, 74(8), 504–512. DOI 10.1136/jclinpath-2020-206979.
48. Mai, H. Q., Chen, Q. Y., Chen, D., Hu, C., Yang, K. et al. (2021). Toripalimab or placebo plus chemotherapy as first-line treatment in advanced nasopharyngeal carcinoma: A multicenter randomized phase 3 trial. *Nature Medicine*, 27(9), 1536–1543. DOI 10.1038/s41591-021-01444-0.
49. Lee, J. H., Lee, J. H., Lim, Y. S., Yeon, J. E., Song, T. J. et al. (2015). Adjuvant immunotherapy with autologous cytokine-induced killer cells for hepatocellular carcinoma. *Gastroenterology*, 148(7), 1383–1391.e1386. DOI 10.1053/j.gastro.2015.02.055.
50. Topalian, S. L., Taube, J. M., Pardoll, D. M. (2020). Neoadjuvant checkpoint blockade for cancer immunotherapy. *Science*, 367(6477), eaax0182. DOI 10.1126/science.eaax0182.
51. Miranda, A., Hamilton, P. T., Zhang, A. W., Pattnaik, S., Becht, E. et al. (2019). Cancer stemness, intratumoral heterogeneity, and immune response across cancers. *Proceedings of the National Academy of Sciences of the United States of America*, 116(18), 9020–9029. DOI 10.1073/pnas.1818210116.
52. Brown, M. P., Ebert, L. M., Gargett, T. (2019). Clinical chimeric antigen receptor-T cell therapy: A new and promising treatment modality for glioblastoma. *Clinical & Translational Immunology*, 8(5), e1050. DOI 10.1002/cti2.1050.
53. Christodoulou, A., Maimaris, G., Makrigiorgi, A., Charidemou, E., Lüchtenborg, C. et al. (2020). TMEM147 interacts with lamin B receptor, regulates its localization and levels, and affects cholesterol homeostasis. *Journal of Cell Science*, 133(16). DOI 10.1242/jcs.245357.
54. Wu, T., Han, N., Zhao, C., Huang, X., Su, P. et al. (2022). LncRNA TMEM147-AS1/miR-133b/ZNF587 axis regulates Warburg effect and promotes prostatic carcinoma invasion and proliferation. *The Journal of Gene Medicine*, e3453. DOI 10.1002/jgm.3453..

55. Xu, B., Peng, Z., An, Y., Yan, G., Yao, X. et al. (2022). Identification of energy metabolism-related gene signatures from scRNA-Seq data to predict the prognosis of liver cancer patients. *Frontiers in Cell and Developmental Biology*, 10, 858336. DOI 10.3389/fcell.2022.858336.
56. Jaeken, J., Lefeber, D. J., Matthijs, G. (2020). SRD5A3 defective congenital disorder of glycosylation: Clinical utility gene card. *The European Journal of Human Genetics*, 28(9), 1297–1300. DOI 10.1038/s41431-020-0647-3.
57. Zhang, Y. P., Na, W. T., Dai, X. Q., Li, R. F., Wang, J. X. et al. (2021). Over-expression of SRD5A3 and its prognostic significance in breast cancer. *World Journal of Surgical Oncology*, 19(1), 260. DOI 10.1186/s12957-021-02377-1.
58. Ye, F., Liang, Y., Cheng, Z., Liu, Y., Hu, J. et al. (2022). Immunological characteristics of alternative splicing profiles related to prognosis in bladder cancer. *Frontiers in Immunology*, 13, 911902. DOI 10.3389/fimmu.2022.911902.
59. Yan, C., Niu, Y., Ma, L., Tian, L., Ma, J. (2022). System analysis based on the cuproptosis-related genes identifies LIPT1 as a novel therapy target for liver hepatocellular carcinoma. *Journal of Translational Medicine*, 20(1), 1–18. DOI 10.1186/s12967-022-03630-1.
60. Lu, C., Shiradkar, R., Liu, Z. (2021). Integrating pathomics with radiomics and genomics for cancer prognosis: A brief review. *Chinese Journal of Cancer Research*, 33(5), 563–573. DOI 10.21147/j.issn.1000-9604.2021.05.03.
61. Ye, F., Hu, Y., Gao, J., Liang, Y., Liu, Y. et al. (2021). Radiogenomics map reveals the landscape of m6A methylation modification pattern in bladder cancer. *Frontiers in Immunology*, 4350. DOI 10.3389/fimmu.2021.722642.