Identification and Validation of Tumour Microenvironment-Based Immune Related Signature for Hepatocellular Carcioma: Immunotherapeutic Implications

Hong Yu
Peking University Shenzhen Hospital
Shao Wang
Peking University Shenzhen Hospital
Tao Zhou
Peking University Shenzhen Hospital
Jia Sun
ShenZhen Beike Biotechnology Research Institute
Tian Qi
Ascentawits Pharmaceuticals
Xian Cheng
Peking University Shenzhen Hospital
De Li
Peking University Shenzhen Hospital
Jun Chen
Peking University Shenzhen Hospital
Wei Zheng (zhengw2013@yeah.net)
Peking University Shenzhen Hospital

Research Article

Keywords: Immune-related gene signature, HCC, survival, WCGNA

Posted Date: January 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-142371/v1

Abstract

***Background**:* Even though treatment outcomes for hepatocellular carcinoma patients have significantly improved, prognostic clinical evaluation remains a substantial challenge due to the heterogeneity and complexity of cancer. Accumulating evidence has revealed that the tumor immune microenvironment is critical for progression and prognosis of hepatocellular carcinoma. A powerful predictive model could assist physicians to better monitor patient treatment outcomes and improve overall survival rates. Therefore, we introduced tumor immune-related genes into a model that could be used for patient risk classification.

***Results**:* First, the Single-sample gene set enrichment analysis (ssGSEA) and Weighted gene co-expression networks construction (WGCNA) methods were applied to identify highly associated immunity genes. Following this, a multi-immune-related gene-based signature determined by The least absolute shrinkage and selection operator (LASSO) Cox regression analysis was used to determine risk stratification. In addition, this predictive model was evaluated according to its performance as a prognostic model in the training and testing datasets. Furthermore, tumor mutation burden and biological enrichment analysis were applied to reveal the potential mechanisms through which the gene signature functions.

***Conclusion**:* In conclusion, our four-gene signature model may be clinically applied in hepatocellular carcinoma patients at high risk of mortality for personalized therapy.

Introduction

Hepatocellular carcinoma (HCC), which is the most common type of primary liver cancer and is characterized by a high degree of heterogeneity (Sia, et al., 2017), is the second leading cause of cancer-related deaths worldwide (Ho, et al., 2019).Beyond standard systemic therapies such as targeted therapy and chemotherapy, accumulating research demonstrates robust and durable responses from immune checkpoint inhibition (ICI) (Keenan, et al., 2019). Immunotherapy is becoming the new standard of care for advanced HCC worldwide (Waidmann, 2018), therefore, providing new therapeutic opportunities by modulating the tumor microenvironment immune response (Sim and Knox, 2018). Treatment options for patients with HCC are rapidly changing and increasing owing to the development of biological and clinical knowledge (Greten, et al., 2019). Even during the treatment process, patients face a high risk of recurrence due to lack of personalized treatment (Wan, et al., 2010). The efficacy and safety of immunotherapy should be evaluated during clinical treatment of malignancies in HCC (Li, et al., 2015). Therefore, it is necessary to develop an independent prognostic signature at the molecular level to determine hepatocellular carcinoma prognosis based on the host immune status (Chen, et al., 2011).

With the rapid development of big bioinformatics data, an approach that integrates gene signatures with classical clinical parameters may provide a great advantage in cancer prognosis (Xu, et al., 2017). Molecular classifications based on somatic mutation profiles and RNA expression profiles related to

patient prognosis will enhance precision medicine during immune responses (Nakagawa, et al., 2019). There are significant and consistent immune system alterations in hepatocellular carcinoma, and some prediction-related models have been established (Bhattacharya, et al., 2016; Bruix and Llovet, 2002). Owing to the availability of various public cohorts providing mRNA and protein expression data, we were able to investigate the prognostic roles of immune-related genes in hepatocellular carcinoma using The Cancer Genome Atlas (TCGA) datasets in this study (Tomczak, et al., 2015).

In the current study, we integrated TCGA Hepatocellular Carcinoma (LIHC) cohorts with transcriptional activity (424 cases total) to select immune-related genes for prognostic prediction (Zhu, et al., 2014). Additionally, we analyzed the affected genes to select for optimal immune-related genes for prognosis based on risk score stratification. Furthermore, an integrated analysis was performed to examine the accuracy of the model in predicting overall survival. Finally, we explored the role of four immune-related genes in different clinical characteristics and biological functions.

Results

Immuno-profiling identifies three subtypes in the TCGA dataset

To cluster hepatocellular carcinoma samples from the TCGA dataset, we first calculated the enrichment levels of the samples using ssGSEA scores from 29 immune gene signatures. Next, samples were clustered hierarchically based on ssGSEA scores, stratifying the samples into three clusters (Fig. 1A). The three clusters were defined as Immunity High (Immunity_H), Immunity Medium (Immunity_M), and Immunity Low (Immunity_L). In addition, our correlation analysis showed that tumor purity was significantly lower in the Immunity_H cluster and significantly higher in the Immunity_L cluster (Kruskal–Wallis test, Fig. 1B). To measure the level of immune cell infiltration, the purity of the tumor, and the stromal content scores in the three clusters, we used the R package “ESTIMATE”. As shown in Fig. 1C, the three subtypes were clustered and divided by lymphocyte_infiltration, tumor_purity, stromal_score, immune_score, and percent of lymphocyte infiltration.

Construction of weighted gene co-expression network

WGCNA analysis is widely applied to examine the Pearson correlation coefficient between gene expression levels (Wu, et al., 2019). Here, we used WCGNA analysis to construct a gene co-expression network in the immunity_H subgroup. To identify the immunity-related correlation gene modules, the mRNA modules in each subtype were analyzed. As presented in Fig. 2A, immunity_H was significantly correlated with the green module, MEpurple, MEsaddlebrown, MEorange, METan, and MEdarkorange. The genes in the green module were used for subsequent analysis. As presented in Fig. 2B, immunity was significantly correlated with the module genes. In addition, we detected the overall gene expression level in the green module in the three immunity subtypes. We found that the genes in the green module were significantly increased in the high immunity group (Fig. 2C).

Gene signature model construction and performance

To exclude genes with a high correlation in the signature model, hub genes obtained from the above selection were further analyzed using the LASSO-penalized Cox analysis (Fig. 3A). Next, we analyzed these nine genes using multivariate Cox regression analysis in the TCGA cohort and identified four genes (*IL18RAP, CSF3R, KIAA1429, PIK3R6*) for the construction of our signature model (Fig. 3B & table.1). Next, we constructed a prognostic model using the risk-score formula: risk score = (coefficients × gene expression level), for each patient. Patients were classified into high- and low-risk groups with median risk scores in the training dataset (Fig. 3C). The high-risk group demonstrated a higher mortality rate than the low-risk group in the training dataset (Fig. 3C) with a P = 3.919e-03 in the log-rank test (table.2). In addition, we found that the area under the curve (AUC) for overall survival (OS) was 0.749 in the time-dependent receiver operating characteristic (ROC) curve (Fig. 3C). Furthermore, the ranked risk scores of patients in the testing dataset showed that the high-risk group demonstrated higher mortality than the low-risk group (AUC = 0.740) (Fig. 3D).

Gene signature model performance evaluation

To evaluate the gene signature performance, we first analyzed the risk status of these two groups. As shown in Fig. 4A, the cut-off value could distinguish between survival statuses and the expression profiles of these four genes in the two groups from the training dataset were visualized in a heatmap. Furthermore, a nomogram was built by including the expression level of the signature genes and the overall survival rate of the patients from the training dataset (Fig. 4B). Calibration plots were used to assess the performance of the nomogram in predicting the 1-, 3-, and 5-year OS rates in the training dataset (Figure. 4C). The c-index of the prediction model was 0.803. Furthermore, we validated the performance of the gene signature in the testing dataset. We found that the risk score formula could also be used to classify the survival status (Fig. 4D & Table 3). Furthermore, nomogram analysis revealed that the gene signature and these four genes would also serve as risk factors to predict the overall survival rate of patients from the testing dataset (Fig. 4E). The c-index of the prediction model was 0.708 (Fig. 4F).

Immunity status in different groups

To further evaluate the immune status difference between the high- and low-risk groups, we compared the immunity-related scores in these two groups in both the training and validation datasets. We found that all four scores were significantly different between these two groups in the training dataset (Fig. 5A). Consistently, the scores were found to be significantly different in the validation dataset (Fig. 5B).

Immunity is positively associated with TMB

Since the TMB level is closely associated with the immunity status of patients, we further analyzed the TMB level of different immunity level groups. Correspondingly, we found that the TMB level was higher in the high-risk group than in the low-risk group in the training dataset (Fig. 6A). However, there was no significant difference between the high- and low-risk groups in the testing dataset (Fig. 6B).

Functional analysis of the four genes

Functional analyses were performed for the genes in the immunity_H module to clarify their biological significance. GO analysis was applied based on biological process (BP), cellular component (CC), and molecular function (MF). As displayed in Fig. 7, these results demonstrated that the immunity_H module was predominantly enriched in immune-related terms such as “cytokine-cytokine receptor interaction” and “natural killer-mediated cytotoxicity”.

Individual gene expression analysis

To further examine the expression profile of these four genes in liver cancer, we used the TCGA dataset. We first examined the different expression statuses in normal and tumor tissues. As shown in Fig. 8A, we found that *CSF3R* and *IL18RAP* increased in the tumor tissue, whereas *KIAA2429* and *PIK3R6* decreased in the tumor tissue. Furthermore, we found that the expression levels of these four genes were associated with the OS of the patients (Fig. 8B). In addition, we also analyzed the spatial expression of these four genes in the human protein altas database. We found that all four genes could be detected in normal tissue and cancer samples (Fig. 8C).

Discussion

Gene signatures are becoming a popular approach in cancer prognosis research, especially immune-based prognostic signatures, which are emerging as a novel hallmark of prognosis (Gaetano, et al., 2015). In this study, a novel and efficient four-gene immune prognostic signature was established using TCGA datasets. The signature showed a high correlation with OS in patients, highlighting the robustness of the gene signature. Furthermore, KEGG and GO enrichment analysis found that the enriched pathways were related to immune-related pathways, which is consistent with previous signatures of the HCC immune microenvironment. Our signature may therefore provide potential biomarkers for malignancy and survival predictions in HCC patients and provide guidance for targeted therapy and immunotherapy. Somatic missense mutations have been mostly studied with regard to their role in the generation of neoantigens, which can be used to identify potential cancer immunity drivers (Porta-Pardo and Godzik, 2016). Genomic studies were applied in investigating the landscape of molecular alterations in HCC; however, only ~ 25% of tumours harbor potentially targetable drivers. Our gene signature could also reflect the different TMB level of subgroups of the patients, which might be used to guiding the treatment options, including biomarker-driven treatments such as targeted therapy and immunotherapies(Llovet, et al., 2018).

Here, we constructed a signature model including genes such as *IL18RAP*, *CSF3R*, *KIAA1429*, and *PIK3R6*. These four genes are mostly associated with cancer development. KIAA1429, an m(6)A methyltransferase (Chen, et al., 2019), predicts OS for breast cancer (Liu, et al., 2019). In addition, KIAA1429 promotes migration and invasion of HCC by inhibiting ID2 (Cheng, et al., 2019). The KIAA1429-GATA3 regulatory model based on the m6A modification provided insights into epi-transcriptomic dysregulation in hepatocarcinogenesis and metastasis (Lan, et al., 2019). KIAA1429 could promote

breast cancer progression and was correlated with pathogenesis, thus possibly representing a promising therapeutic strategy in combination with CDK1 treatment (Qian, et al., 2019). KIAA1429 could also act as an independent predictive factor in HCC and correlated with survival and prognosis (Liu, et al., 2020). The granulocyte colony-stimulating factor (G-CSF or CSF3) and its receptor CSF3R would regulate granulopoiesis and hematopoietic stem cell mobilization (Zhang, et al., 2018). G-CSFR is also associated with the development of myelodysplasia/AML in patients with severe congenital neutropenia (Kunter, et al., 2011). G-CSFR also controls myeloid progenitor proliferation and differentiation to neutrophils (Klimiankou, et al., 2016). G-CSFR plays an important role in the production of neutrophil granulocytes (Dwivedi, et al., 2019). The IL-18 receptor accessory protein (IL-18RAP) was reported to be important in the development of esophageal cancer (Zhu, et al., 2016). A novel 2-gene signature, including IL18RAP, was generated for HCC prognosis (Tian, et al., 2020). However, the functional roles of these four genes in liver cancer remain unclear. Our findings implicate these genes as prognostic markers for liver cancer and could reveal their potential functions in liver cancer progression.

In summary, we identified four novel immune-related genes using TCGA dataset analysis that can be used to determine liver cancer prognosis. Our signature model may reveal relevant dysregulated immune genes in liver cancer patients and predict responses to immunotherapy, and potentially to combined therapy. However, further validation of our gene signature in other independent cohorts and in functional experiments would support this conclusion.

Conclusions

We proposed and independently validated three reproducible immune molecular subtypes of liver cancer, which may provide insights for personalized immunotherapy in patients.

Materials And Methods

Data collection

mRNA expression data and clinical parameters were downloaded from the target database. In the dataset, we included 424 samples with clinical information and RNA sequencing (RNA-seq) data. Fragments per kilobase million (FPKM) data were collected. Ethical approval was not necessary for our study, because all data were retrieved from a public database. Perl (ActivePerl 5.28) was used for data extraction and code was available upon request. The dataset was further split into the training dataset:validation dataset with a 3:1 ratio.

Single-sample gene set enrichment analysis (ssGSEA)

Immune features in each sample were evaluated according to 29 immune Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway categories using ssGSEA analysis using the R package, GSVA [15, 16]. Briefly, the mRNA expression levels of genes in each sample were measured and ranked to generate enrichment scores.

Sub-Clustering

Based on ssGSEA scores, we performed hierarchical clustering of the samples using the “clusterProfile” R package. Immune cell infiltration levels, tumor purity, and stromal content were calculated using ESTIMATE for each sample. To calculate the correlation of the three clusters with immune cell infiltration levels, we used the “pHeatmap” package to cluster samples.

Weighted gene co-expression networks construction (WGCNA) and module selection

WGCNA was used to identify gene modules that were significantly associated with the immunity_H group. Gene modules that were specifically amplified in the three immunity subtypes were identified.

Prognostic values of gene signature analysis

Patients with clinical features including survival data, age, tumor stage, and sex were used for subsequent analysis. The LASSO Cox regression analysis in the “Glmnet” R package was used to detect immune-related genes and construct a prognostic gene signature. $P < 0.001$ was considered statistically significant. Multivariate Cox regression analysis of the selected genes was conducted using the forward stepwise procedure.

Building the prognostic immune-related gene signature

The R packages “survival” and “survminer” were applied to calculate the survival rate, hazard ratios (HR), and 95% confidence intervals (CI), and plot the Kaplan–Meier survival curve. The following risk score formula was used: risk score = mRNA expression level of *IL18RAP* * -2.736 + mRNA expression level of *CSF3R* * 0.123 + mRNA expression level of *KIAA1429* * 0.093 + mRNA expression level of *PIK3R6* * 0.777. The R packages “survival” and “survminer” were also used to find the optimal cut-off to stratify patients. In addition, the “survivalROC” R package was applied to investigate the prognostic value of the gene signature. $P < 0.05$ was considered statistically significant.

Predictive nomogram generation

A nomogram was generated using independent prognostic measures with the “RMS” R package. A calibration plot was applied to examine the calibration value of the nomogram.

Functional enrichment analysis

Functional enrichment analysis was performed using the “clusterProfile” R package. Genes enriched in Gene Ontology (GO) and KEGG pathway categories were identified. The GO terms and KEGG pathways with $P < 0.05$ and enrichment scores > 2 were considered statistically significant and were plotted using the R package ggplot2.

Tumor mutation burden (TMB) analysis

TMB was estimated as 10 kb in length with the Perl software. Code was available upon request.

Statistical analysis

Statistical analysis was performed using GraphPad Prism V8.0.2 and R software V. 4.0.0. $P < 0.05$ was considered statistically significant, unless otherwise specified. *$P < 0.01$, **$P < 0.05$, ***$P < 0.005$, ****$P < 0.001$

List Of Abbreviations

Hepatocellular carcinoma (HCC)

immune checkpoint inhibition (ICI)

The Cancer Genome Atlas (TCGA)

Liver Hepatocellular Carcinoma (LIHC)

the single-sample gene-set enrichment analysis (ssGSEA)

receiver operating characteristic (ROC) curve

cellular component (CC)

molecular function (MF)

Gene Ontology (GO)

Kyoto encyclopedia of genes and genomes (KEGG)

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that there is no conflict of interests.

Funding

This work was supported by grants from the Sanming Project of Medicine in Shenzhen (Grant No. SZSM201612071), the Cell Technology Center and Transformation Base, Innovation Center of Guangdong-Hong Kong-Macao Greater Bay Area, Ministry of Science and Technology of China (Grant No. YCZYPT [2018]03-1), Shenzhen Key Discipline of Stem Cell Clinical Research, the Key Medical Disciplines Construction Funding in Shenzhen (Grant No. SZXK078)

AUTHOR CONTRIBUTIONS

J.C. and W.Z. conceived and designed the study. J.C., W.Z., H.Y. and S.W. collected the data, analyzed the data and wrote the manuscript. T.Z., J.S., T.Q., X.C., and D.J. performed the experiments. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

Not applicable

References

1. Bhattacharya, S., *et al.* ImmPort: disseminating data to the public for the future of immunology. *Immunologic Research* 2016;58(2–3):234–239.
2. Bruix, J. and Llovet, J.M. Prognostic prediction and treatment strategy in hepatocellular carcinoma. 2002;35(3):519–524.
3. Chen, G.M., *et al.* Prediction of survival in patients with liver cancer using artificial neural networks and classification and regression trees. *Taiwan Journal of Public Health* 2011;30(5):481–493.
4. Chen, X.Y., Zhang, J. and Zhu, J.S. The role of m(6)A RNA methylation in human cancer. *Mol Cancer* 2019;18(1):103.
5. Cheng, X., *et al.* KIAA1429 regulates the migration and invasion of hepatocellular carcinoma by altering m6A modification of ID2 mRNA. *Onco Targets Ther* 2019;12:3421–3428.
6. Dwivedi, P., *et al.* Phospho serine and threonine analysis of normal and mutated granulocyte colony stimulating factor receptors. *Sci Data* 2019;6(1):21.
7. Gaetano, B., *et al.* The Immune System in Hepatocellular Carcinoma and Potential New Immunotherapeutic Strategies. *Biomed Research International* 2015;2015:1–12.
8. Greten, T.F., *et al.* Targeted and Immune-Based Therapies for Hepatocellular Carcinoma. *Gastroenterology* 2019;156(2):510–524.
9. Ho, D.W., *et al.* Single-cell transcriptomics reveals the landscape of intra-tumoral heterogeneity and stemness-related subpopulations in liver cancer. *Cancer letters* 2019;459:176–185.

10. Keenan, B.P., Fong, L. and Kelley, R.K. Immunotherapy in hepatocellular carcinoma: the complex interface between inflammation, fibrosis, and the immune response. *Journal for immunotherapy of cancer* 2019;7(1):267.
11. Klimiankou, M., *et al.* Role of CSF3R mutations in the pathomechanism of congenital neutropenia and secondary acute myeloid leukemia. *Ann N Y Acad Sci* 2016;1370(1):119–125.
12. Kunter, G., Woloszynek, J.R. and Link, D.C. A truncation mutant of Csf3r cooperates with PML-RARalpha to induce acute myeloid leukemia in mice. *Exp Hematol* 2011;39(12):1136–1143.
13. Lan, T., *et al.* KIAA1429 contributes to liver cancer progression through N6-methyladenosine-dependent post-transcriptional modification of GATA3. *Mol Cancer* 2019;18(1):186.
14. Li, S., Yang, F. and Ren, X. Immunotherapy for hepatocellular carcinoma. *Drug Discov Ther* 2015;9(5):363–371.
15. Liu, J., *et al.* The Cancer Genome Atlas (TCGA) based m(6)A methylation-related genes predict prognosis in hepatocellular carcinoma. *Bioengineered* 2020;11(1):759–768.
16. Liu, L., *et al.* N6-methyladenosine-related Genomic Targets are Altered in Breast Cancer Tissue and Associated with Poor Survival. *Journal of Cancer* 2019;10(22):5447–5459.
17. Llovet, J.M., *et al.* Molecular therapies and precision medicine for hepatocellular carcinoma. *Nature reviews. Clinical oncology* 2018;15(10):599–616.
18. Nakagawa, H., Fujita, M. and Fujimoto, A. Genome sequencing analysis of liver cancer for precision medicine. *Seminars in cancer biology* 2019;55:120–127.
19. Porta-Pardo, E. and Godzik, A. Mutation Drivers of Immunological Responses to Cancer. *Cancer Immunol Res* 2016;4(9):789–798.
20. Qian, J.Y., *et al.* KIAA1429 acts as an oncogenic factor in breast cancer by regulating CDK1 in an N6-methyladenosine-independent manner. *Oncogene* 2019;38(33):6123–6141.
21. Sia, D., *et al.* Liver Cancer Cell of Origin, Molecular Class, and Effects on Patient Prognosis. *Gastroenterology* 2017;152(4):745–761.
22. Sim, H.W. and Knox, J. Hepatocellular carcinoma in the era of immunotherapy. *Curr Probl Cancer* 2018;42(1):40–48.
23. Tian, Z., *et al.* Bioinformatics Analysis of Prognostic Tumor Microenvironment-Related Genes in the Tumor Microenvironment of Hepatocellular Carcinoma. *Medical science monitor: international medical journal of experimental and clinical research* 2020;26:e922159.
24. Tomczak, K., Czerwińska, P. and Wiznerowicz, M. Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary Oncology* 2015;19(1A):A68-77.
25. Waidmann, O. Recent developments with immunotherapy for hepatocellular carcinoma. *Expert Opin Biol Ther* 2018;18(8):905–910.
26. Wan, Y.W., *et al.* Hybrid Models Identified a 12-Gene Signature for Lung Cancer Prognosis and Chemoresponse Prediction. *PloS one* 2010;5(8):e12222.

27. Wu, C., *et al.* Immunomodulatory Effects and Induction of Apoptosis by Different Molecular Weight Chitosan Oligosaccharides in Head Kidney Macrophages From Blunt Snout Bream (Megalobrama amblycephala). *Front Immunol* 2019;10:869.
28. Xu, G., *et al.* A 15-gene signature for prediction of colon cancer recurrence and prognosis based on SVM. *Gene* 2017;604:33–40.
29. Zhang, H., *et al.* Gain-of-function mutations in granulocyte colony-stimulating factor receptor (CSF3R) reveal distinct mechanisms of CSF3R activation. *J Biol Chem* 2018;293(19):7387–7396.
30. Zhu, J., *et al.* Association of the interleukin-18 receptor 1 and interleukin-18 receptor accessory protein polymorphisms with the risk of esophageal cancer. *Biomed Rep* 2016;4(2):227–235.
31. Zhu, Y., Qiu, P. and Ji, Y. TCGA-Assembler: open-source software for retrieving and processing TCGA data. *Nature Methods* 2014;11(6):599–600.

Tables

Table. 1 The survival analysis by cox regression model in the validation group. Coef: Coefficient; HR: Hazard ratio; CI: confidence interval.

Coef

ID	*coef*	*HR*	*HR.95% CI*	*pvalue*
IL18RAP	-2.736	0.065	0.014-0.310	**0.001**
CSF3R	0.123	1.131	1.042-1.227	**0.003**
KIAA1429	0.093	1.098	0.972-1.240	0.134
PIK3R6	0.777	2.174	1.024-4.616	**0.043**

Table. 2 The survival analysis by univariate and multivariate cox regression model in the training group. HR: Hazard ratio; CI: confidence interval.

	Univariate analysis			Multivariate analysis		
ID	*HR*	*HR.95% CI*	*pvalue*	*HR*	*HR.95% CI*	*pvalue*
Age	1.01	0.99-1.03	0.352	1.01	0.99-1.04	0.259
Gender	1.16	0.65-2.08	0.607	1.53	0.83-2.83	0.172
Grade	0.81	0.54-1.21	0.305	0.67	0.43-1.03	0.069
Stage	1.71	1.29-2.28	**<0.001**	1.73	1.3-2.31	**<0.001**
RiskScore	1.26	1.16-1.36	**<0.001**	1.34	1.22-1.47	**<0.001**

Table. 3 The survival analysis by univariate and multivariate cox regression model in the validation group. HR: Hazard ratio; CI: confidence interval.

	Univariate analysis			Multivariate analysis		
ID	*HR*	*HR.95% CI*	*pvalue*	*HR*	*HR.95% CI*	*pvalue*
Age	1.01	0.99-1.03	0.247	1.01	0.99-1.03	0.336
Gender	0.59	0.34-1.01	0.053	0.81	0.45-1.45	0.482
Grade	1.40	0.98-1.99	0.063	1.33	0.9-1.97	0.148
Stage	1.64	1.21-2.23	**0.002**	1.58	1.15-2.19	**0.005**
RiskScore	1.61	1.31-1.99	**<0.001**	1.35	1.07-1.7	**0.012**

Figures

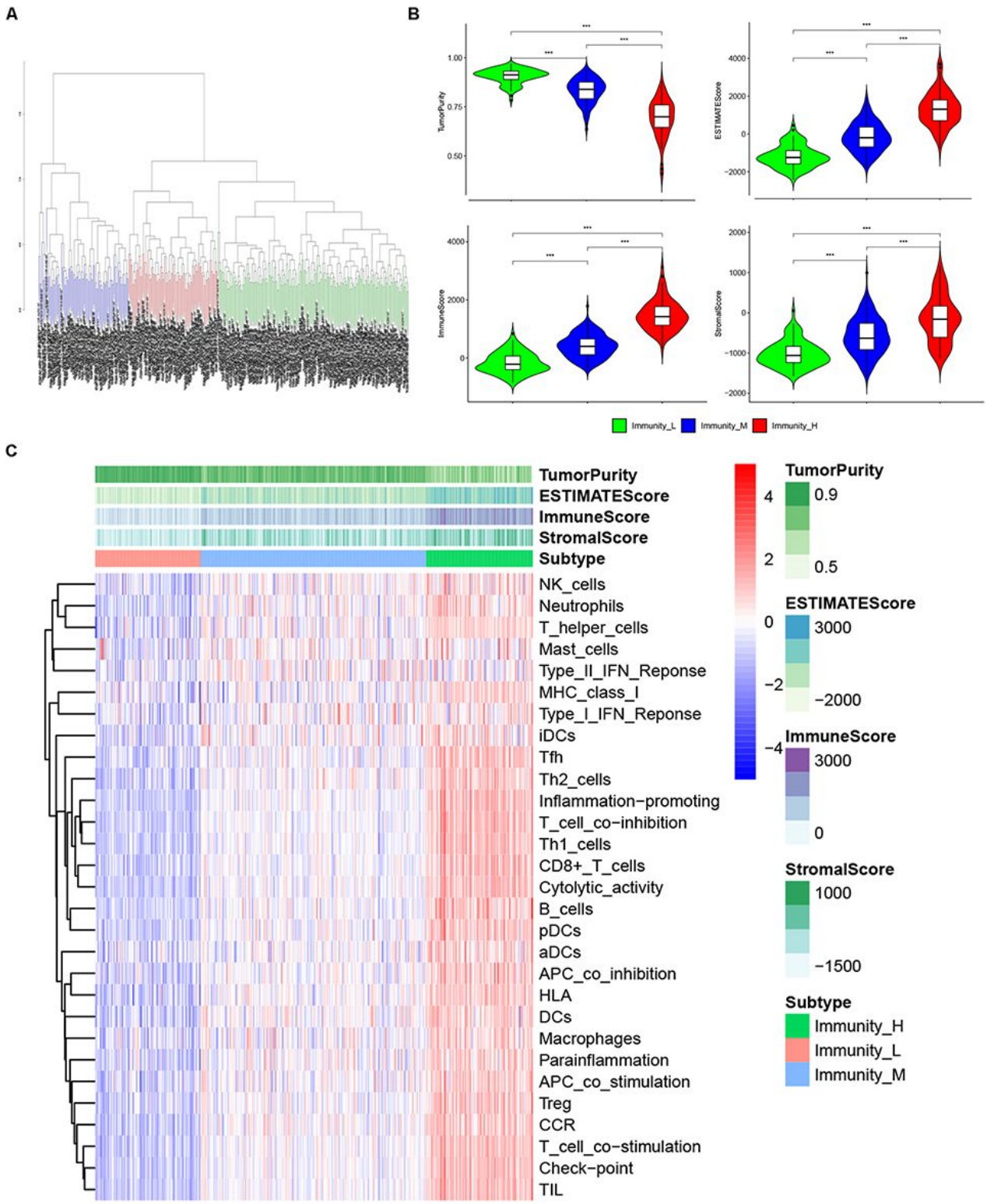

Figure 1

Hierarchical clustering of the datasets yields three immunity subtypes. (A) Dendrogram showing hierarchical clustering of the samples into three subtypes (red, blue and green). (B) Violin plot showing differences in the tumor_purity level between Immunity High (Immunity_H, red), Immunity Medium (Immunity_M, blue), and Immunity Low (Immunity_L, green) groups . ***P < 0.001, one-way ANOVA test.

(C) Heatmap showing Tumor_purity, Stromal_score, Immune_score Lymphocyte_infiltration and percent of lymphocyte infiltration in the three subtypes.

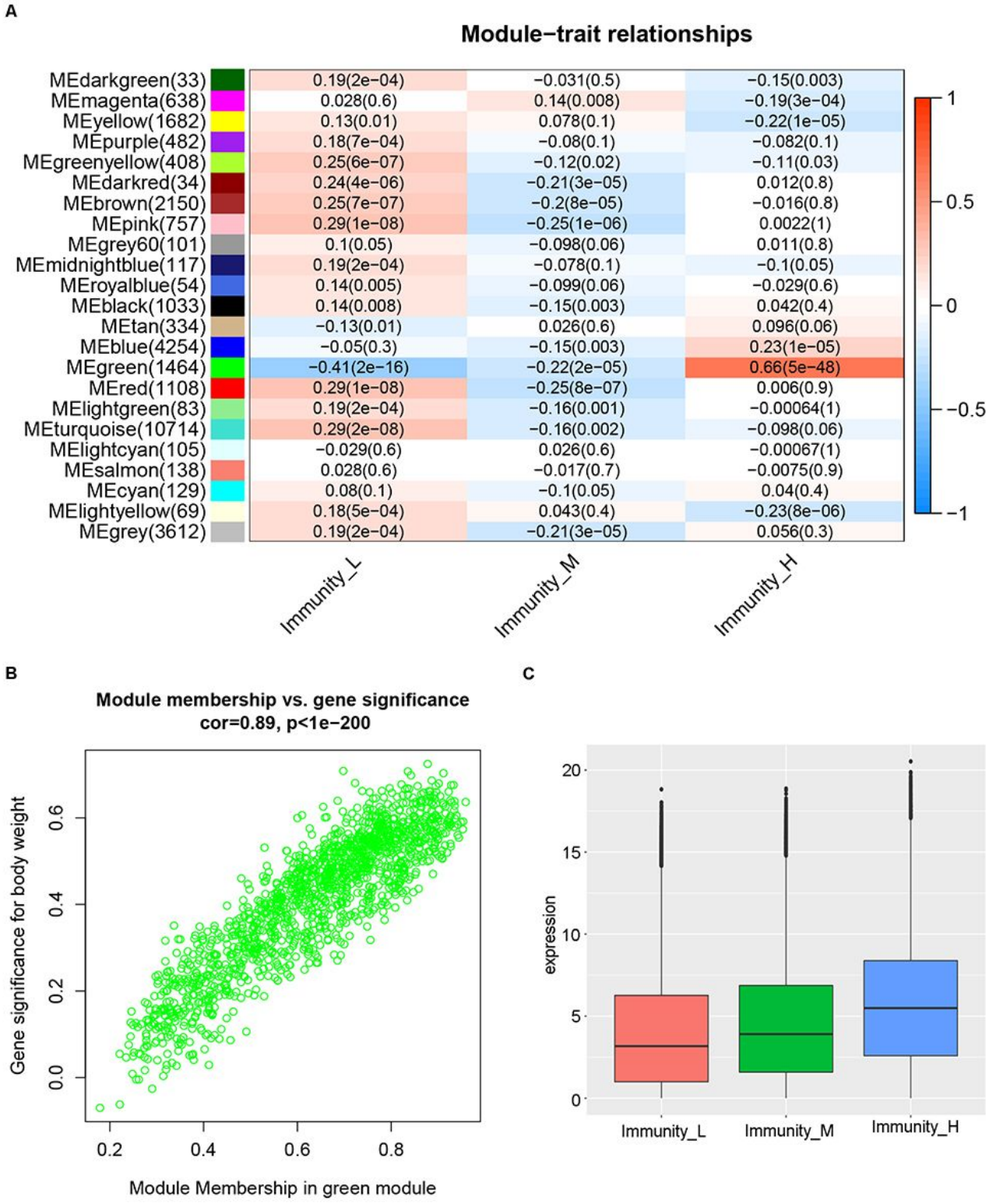

Figure 2

WGCNA revealed immunity-related modules (A) Heatmap of the correlation between modules and immunity subtypes. (B) Scatterplot of the significant genes in the Immunity_H subtype versus module membership. Gene significance (GS) - module membership (MM) exhibits a very significant correlation.

Pearson's coefficient and P-values are indicated above each plot. (C) Box-plot of the expression profile of MEgreen genes in the different Immunity level subgroups. **P < 0.05, one-way ANOVA test.

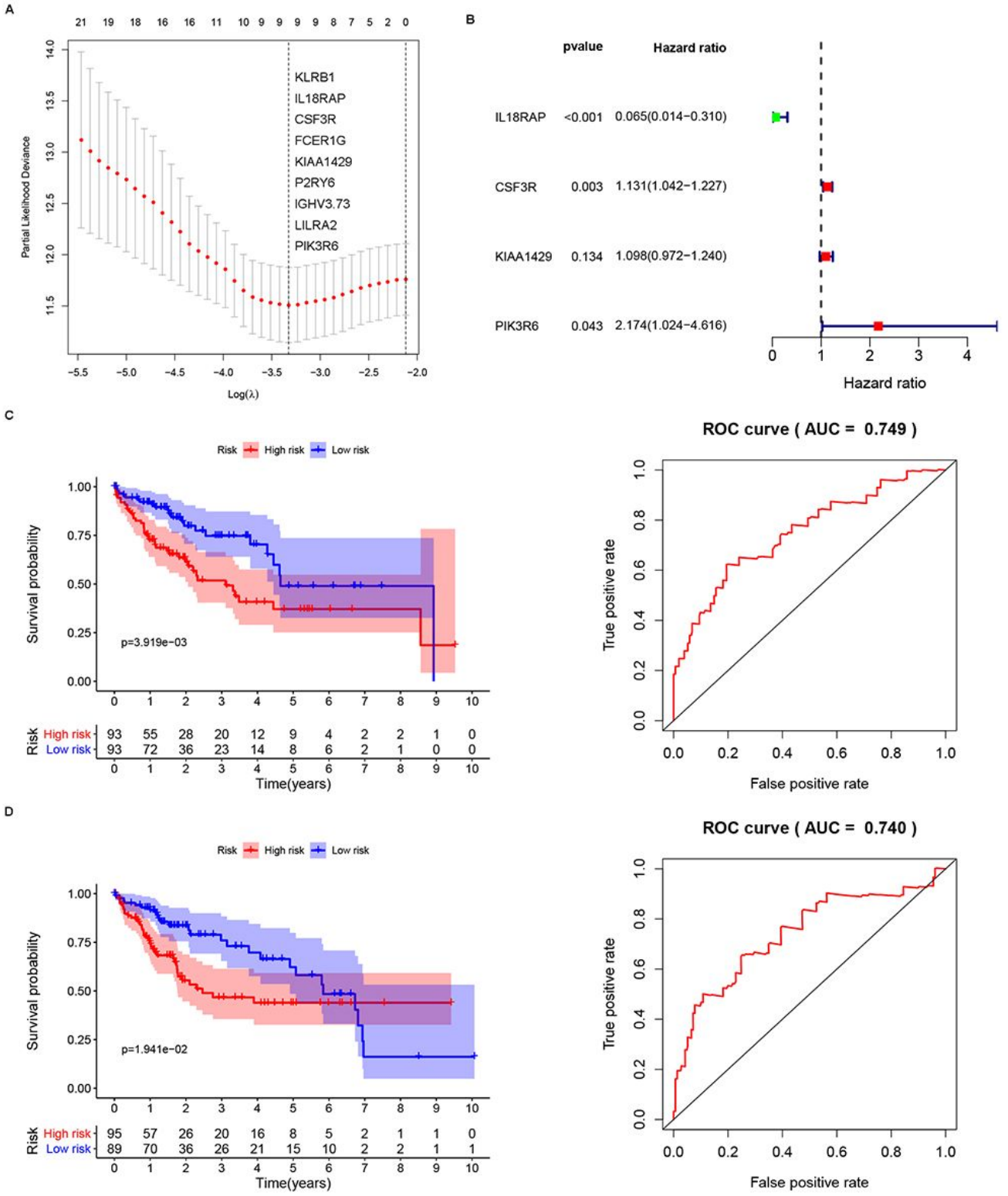

Figure 3

Construction of the integrated gene signature and distribution of risk scores based on gene signature classification. (A) LASSO coefficient profiles of the differentially expressed genes. Nine genes were selected for subsequent analysis. (B) Forrest plot showing the four differentially expressed genes

according to multivariate Cox regression analysis. (C) Kaplan-Meier (K-M) survival plot of the survival rate between the high- and low-risk groups in the training dataset. P-value was calculated using the log-rank test. ROC curve of the gene signatures used to determine the accuracy of the predication model in the training dataset. (D) K-M plot of the survival rate in high- and low-risk groups in the testing dataset. P-value was calculated using the log-rank test. ROC curve of the gene signatures used to determine the accuracy of the prediction model in the testing dataset.

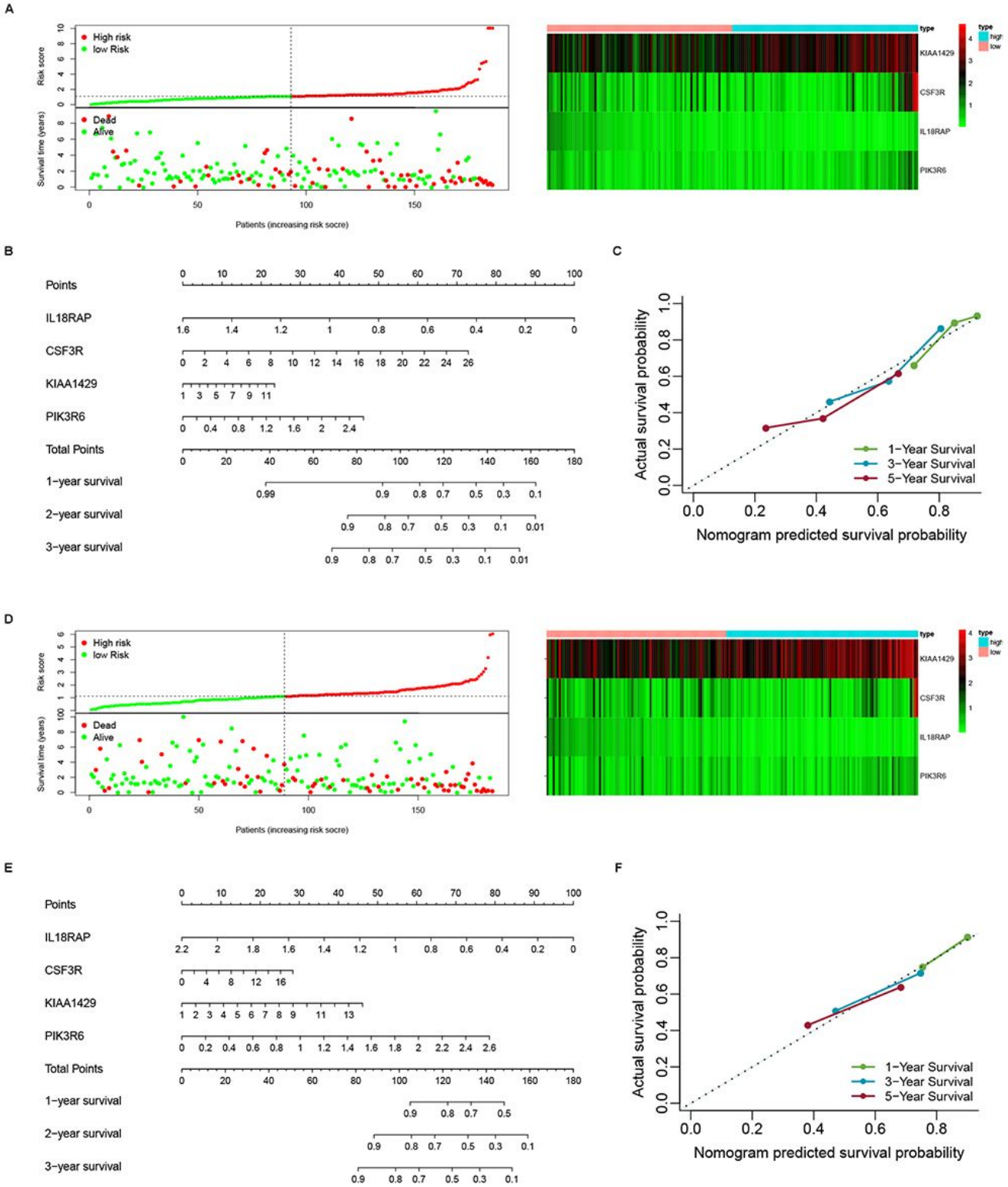

Figure 4

Analysis of the clinicopathological features in the two groups. (A) The K-M survival curve, ROC curves, and distribution of risk scores based on gene signature classification. (A) Prognostic gene signature analysis in the training dataset: dot plot showing risk score distribution; vital status of the patient is indicated and color-coded; the immune-related gene expression in the prognostic model is shown in the heatmap. (B) Nomograms predicting the 1-, 2-, and 3-year survival probability in liver cancer patients in the training dataset. Total points were obtained by adding the points of individual covariates based on the point scale. (C) Calibration plot for the nomogram for 1-, 2-, and 3-year survival groups in the training dataset. The dashed line indicates the ideal reference line, whereas dashes represent nomogram-predicted probabilities. (D) Prognostic gene signature model analysis in the testing dataset: dot plot showing risk score distribution; the vital status of the patient is color-coded; heatmap shows the immune-related gene expression in the prognostic model. (E) Nomograms predicting the 1-, 2-, and 3-year survival probability in liver cancer patients in the testing dataset. Total points were obtained by adding the points of individual covariates based on the point scale. (F) Calibration plot for the nomogram for 1-, 2-, and 3-year survival groups in the testing dataset. The dashed line indicates the ideal reference line, whereas dashes represent nomogram-predicted probabilities.

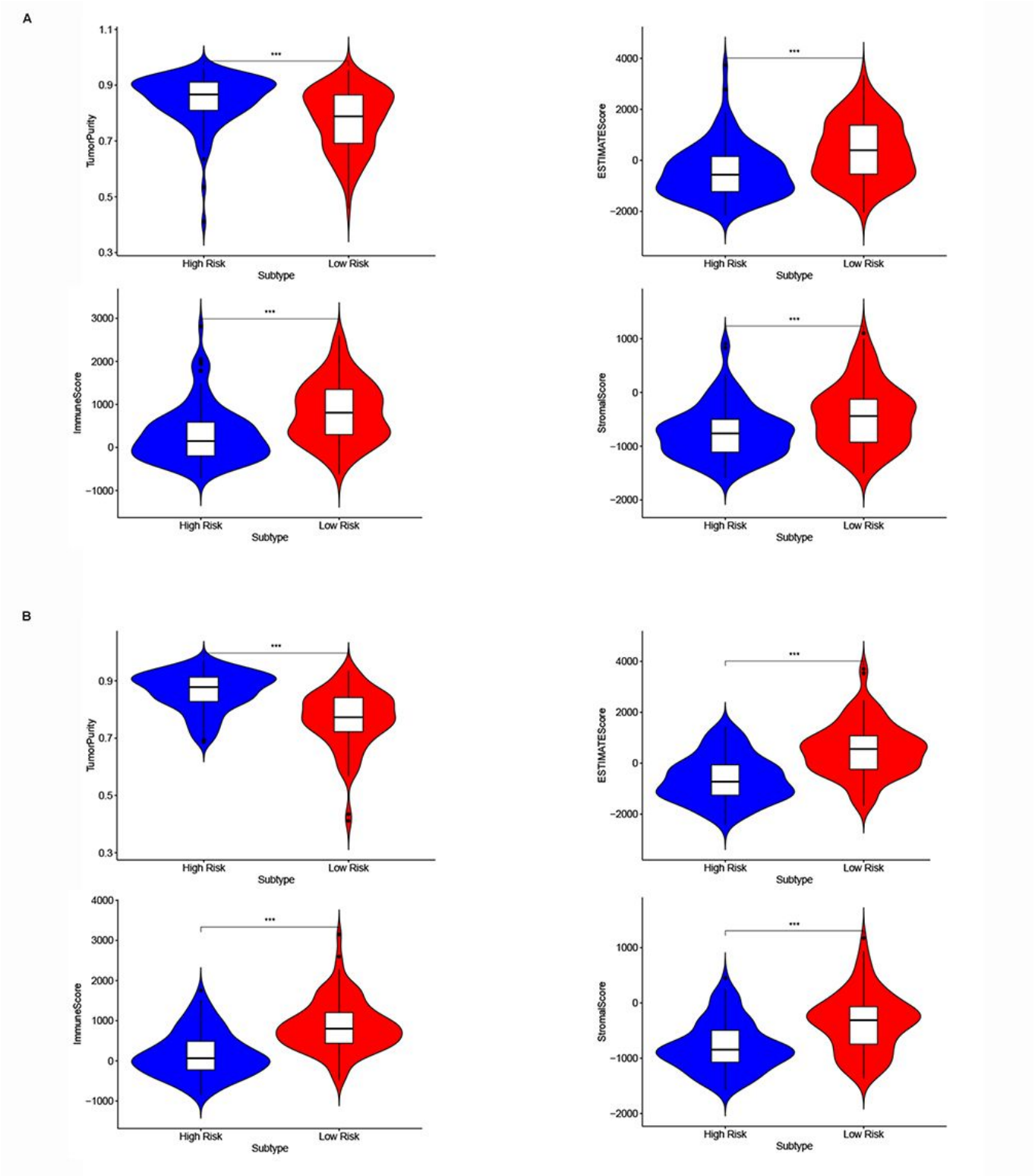

Figure 5

Analysis of tumor purity, immune scores, ESTIMATE score, and stromal scores in the two groups. (A) Tumor purity, immune scores, ESTIMATE score, and stromal scores in the two groups in shown in violin plots in the training dataset. (B) The violin plots display tumor purity, immune scores, ESTIMATE score, and stromal scores in the two groups in the setting dataset, based on student t test.

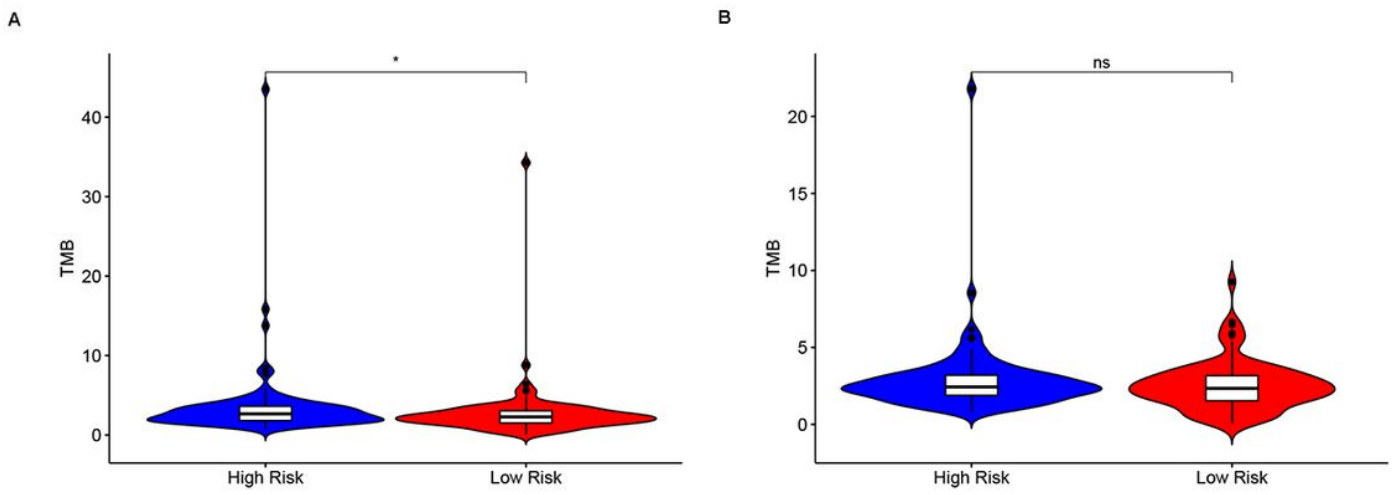

Figure 6

Analysis of the tumor burden in the two groups. (A) Violin plots display TMB scores in the two groups in the training dataset. (B) The violin plots show tumor purity, immune scores, ESTIMATE scores, and stromal scores comparison in the two groups in the testing dataset, based on student t test.

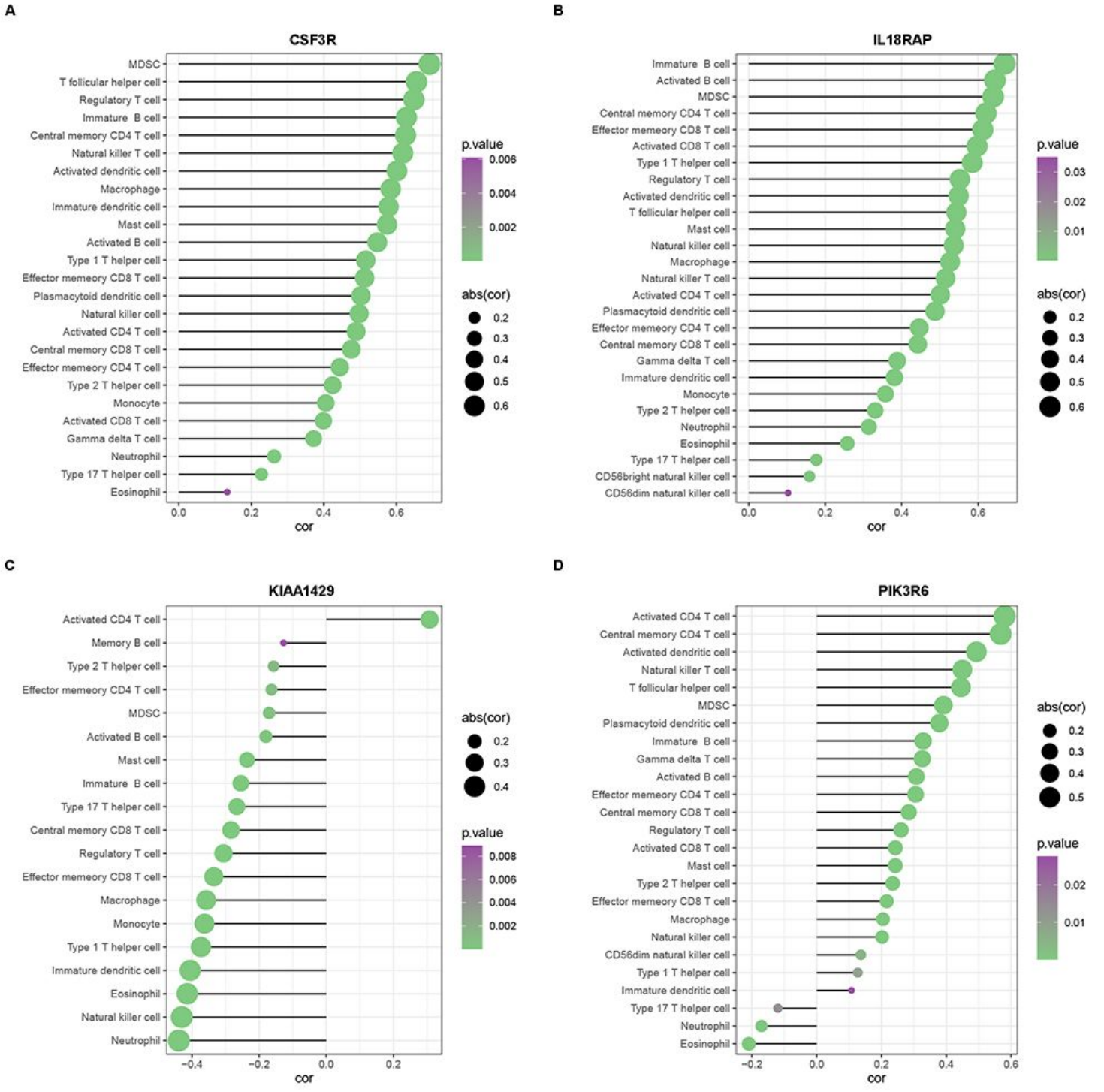

Figure 7

Biological pathway enrichment based on the genes in the model. (A) Bubble plot displaying the GO enrichment of the top 10 upregulated and downregulated pathways. (B) Bubble plot displaying the KEGG enrichment of the top 10 upregulated and downregulated pathways. Risk_H represents the high-risk group (red bubble). Risk_L represents the low-risk group (blue bubble). Bubble size represents -log10(p value). Gene ontology enrichment analysis for biological process: BP, cellular component: CC, molecular

function: MF. The size of the bubble indicates the number of genes, and the color indicates the P-value. KEGG: Kyoto Encyclopedia of Genes and Genomes.

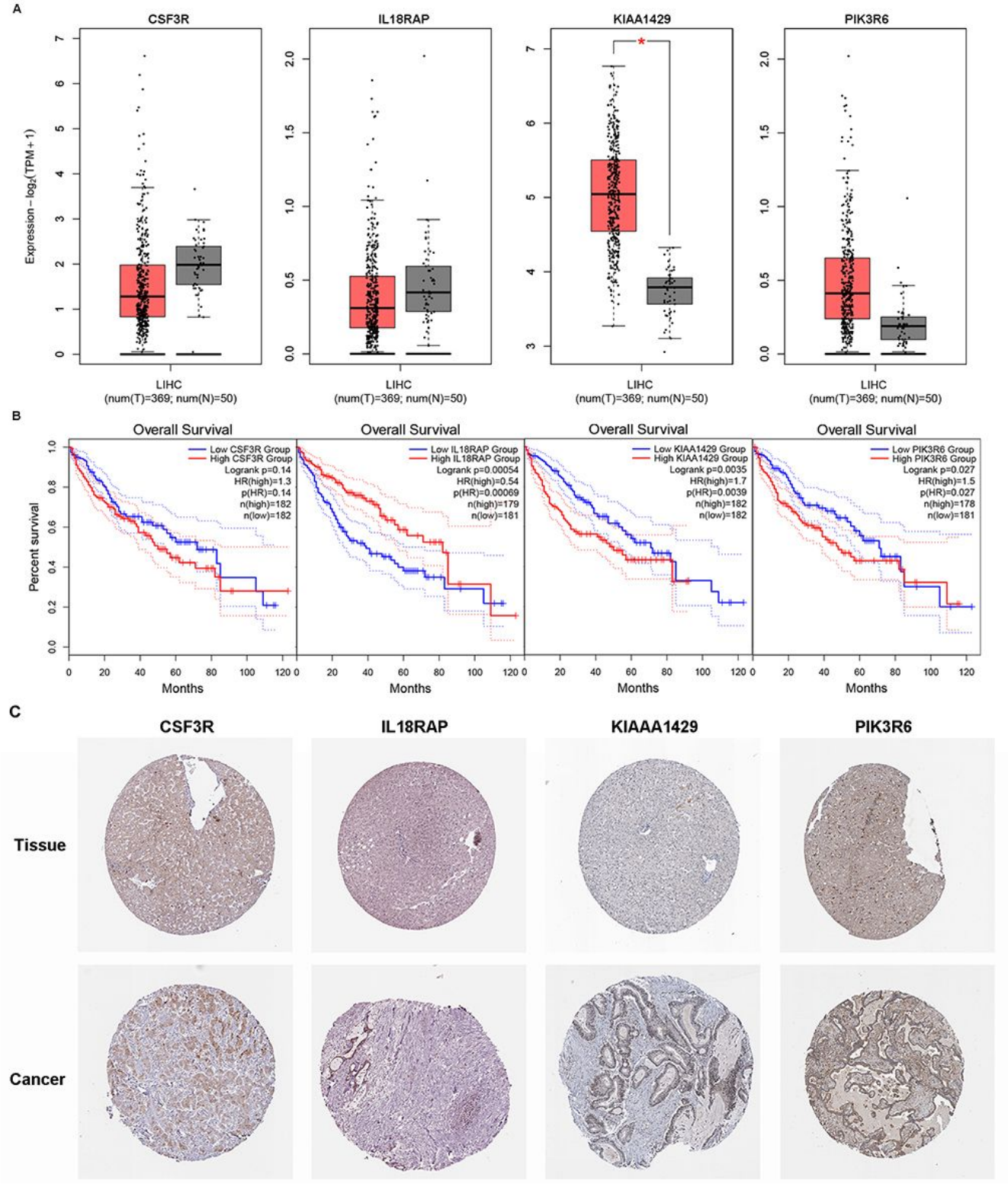

Figure 8

Gene function analysis of the four genes in the model. (A) Gene expression level analysis is shown using Box-plot. (B) Kaplan-Meier survival plot of the survival rate between the high- and low-expression groups

in the testing dataset. (C) IHC demonstrates the spatial expression of the four genes in the tumor and tissue groups from the human protein atlas database.