Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identification of a Five-Immune Gene Model as an Independent Prognostic Factor in Hepatocellular Carcinoma

Haitao Chen
Zhongnan Hospital of Wuhan University
Jianchun Guo (spring0324@126.com)
Zhongnan Hospital of Wuhan University https://orcid.org/0000-0002-7793-5432

Primary research

Keywords: Immune gene, Prognosis, Risk model, Hepatocellular carcinoma

Posted Date: July 15th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-41766/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at BMC Cancer on March 16th, 2021. See the published version at https://doi.org/10.1186/s12885-021-08012-2.

Abstract

Background: Hepatocellular carcinoma (HCC) is a common cancer with a poor prognosis. We purposed to identify a prognostic risk model of HCC according to the differentially expressed (DE) immune genes.

Methods: The DE immune genes were identified based on 374 HCC and 50 adjacent normal samples from the Cancer Genome Atlas database. Univariate Cox analysis, Lasso regression, and multivariate Cox analysis were used to determine the immune genes used to construct the model based on the training group. The testing group and the entire group were applied for the validation of the model.

Results: A five-immune gene model comprising HSPA4, ISG20L2, NDRG1, EGF, and IL17D was identified. Based on the model, overall survival was significantly different between the high-risk and low-risk groups (P = 7.953e-06). The AUC for the model at 1- and 3-year was 0.849 and 0.74, respectively. The validating groups confirmed the reliability of the model. The risk score was identified as an independent prognostic factor and was closely related to the content of immune cells from HCC samples.

Conclusion: We identified a five-immune gene model, which could be treated as an independent prognostic factor of HCC.

Background

Hepatocellular carcinoma (HCC) is common cancer all over the world, being the third leading cause of cancer-associated death. Its global morbidity and mortality continue to increase, causing more than 600,000 deaths per year [1, 2]. Considering that the symptoms and signs of early HCC can hardly be noticed, the diagnosis of HCC patients is often lagging, and the prognosis is poor [3, 4]. Besides, due to the existence of individual specificity, patients with a similar pathological stage may still have vast differences in the overall survival (OS) [5]. Therefore, it is urgent and necessary to develop a molecular model that could accurately predict the prognosis of HCC.

In the last few years, immunotherapy has become an essential method for HCC patients [3]. A variety of strategies, including cancer vaccines, adoptive cellular therapy, and immune checkpoint blockade (ICB), have been explored. ICB has got promising outcomes in HCC [6, 7]. Besides, several immune checkpoint inhibitors, such as anti-PD-L1, anti-CTLA-4, and anti-PD-1 antibodies, have also displayed potential therapeutic effects on HCC [8, 9], and these immune checkpoints have been reported to participate in the induction and maintenance of immune tolerance in HCC [10, 11]. All these published papers above proved that immunity carries out a vital role in the progress of HCC.

Previous studies indicated that immune genes were related to the response and prognosis of HCC patients to immunotherapy [12]. Dong et al. demonstrated that high expression of STAT5A, STAT5B, and STAT6 was associated with better prognosis in HCC patients [13]. This article proposed to identify a trustworthy model of HCC on the basis of immune genes, and demonstrated its clinical utility for HCC patients.

Methods

Basic information

Clinical information and expression data and were obtained from The Cancer Genome Atlas (TCGA) database. Cancer-related transcription factors (TFs) and immune genes were available in the Cistrome database and the ImmPort database, respectively. Also, immune infiltrate data was available in Tumor Immune Estimation Resource.

Detection of the DE immune genes

We applied the Wilcoxon signed-rank test to identify DE genes and DE immune genes with R software. The cut-offs were false discovery rate (FDR) < 0.05 and Log_2(fold change [FC]) > 1. The DE genes and immune genes were presented in the volcano plot and heatmap using the gplots package and Pheatmap package.

Function enrichment analyses of the DE immune genes

We applied the Database for Annotation, Visualization, and Integrated Discovery (*DAVID*) (v6.8) for Gene Ontology (GO) analysis and ClusterProfiler package for the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. GO terms, comprising cellular component, biological process, and molecular function, were considered significantly enriched when Bonferroni correction < 0.001 and FDR < 0.05, and same for KEGG terms when p.adjust < 0.001 and GeneRatio > 0.05.

Construction of a DE immune genes-TFs network

We detected the DE TFs according to the DE genes and cancer-related TFs. Cor.test() was used to estimate the correlation between DE immune genes and DE TFs. Correlation coefficient > 0.6 and p < 0.001 was set as a threshold to get more robust interaction pairs between DE immune genes and DE TFs. The qualified interactions were imported to Cytoscape (v3.7.2) to construct a network.

Establishment of a prognostic risk model

A risk model was developed with the training group. We performed univariate analysis to detect the qualified immune genes. Then, we conduct Lasso regression to eliminate functionally similar genes. At last, multivariate analysis was applied to find out the ultimate immune genes. The risk score was calculated according to the following formula:

$$\text{The risk score} = \sum_{i=1}^{N}(\text{Exp} * \text{Coef})$$

N, Exp, and Coef represented gene number, expression level, and coefficient value, respectively. We divided all HCC patients into high-risk and low-risk groups due to the median of the risk score. A low-risk score shows good survival for HCC patients. Kaplan-Meier analysis was employed to compare the survival rate

between the two groups. The SurvivalROC package was applied to perform receiver operating characteristic (ROC) analysis. It is reported that an area under the ROC (AUC) > 0.60 was viewed as suitable for predictions, and an AUC > 0.75 was believed to have perfect predictive value [14]. Risk curves and the heatmap of risk genes were also utilized to assess this model.

Validation of the model

We conducted a survival analysis of the OS and performed the ROC analysis for the validation of the model in the testing group as well as the entire group.

The prognostic value of the risk model

All the clinical parameters, as well as the risk score, were analyzed with univariate and multivariate Cox analyses. Indicators were considered independent prognostic factors when $P < 0.05$ in both Cox analyses.

Applicability of the model

The correlation of the model (the risk score and expression level of the immune genes) with clinical parameters (age, gender, the histological grade, as well as the pathological stage), were analyzed. Patients were separated into two subgroups according to age (>=60 and <60 years old), gender (male and female), grade (G1&2 for well-differentiated and G3&4 for poor-differentiated), and stage (stage I&II and stage III&IV). Besides, we assessed the correlation between the risk score and the content of immune cells using the Pearson correlation coefficient test.

Results

Basic information

A total of 374 tumors and 50 normal samples were downloaded from the TCGA data portal. Three tumor samples lacking clinical data and 28 HCC samples with follow-up time less than 30 days were excluded. The clinical information of the 343 patients left with HCC was presented in Table 1. We randomly divided the 343 tumor samples into the training and testing groups (Table 2). The workflow of the current study was shown in Fig. 1.

Detection of the DE immune genes

5388 DE genes (325 down-regulated and 5063 up-regulated) and 325 DE immune genes (59 down-regulated and 266 up-regulated) were discovered, which were shown in the volcano map and heatmap (Fig. 2a-d). The top ten down-regulated and up-regulated genes were presented in Table S1.

GO and KEGG analyses of the DE immune genes

68 GO terms, comprising 14 molecular function terms, seven cellular component terms, and 47 biological process terms, were detected. Also, 65 enriched pathways were discovered. The top ten GO terms and pathway terms were depicted in Fig. 2e and Fig. 2f, respectively.

Construction of a DE immune genes-TFs network

318 tumor-related TFs were available in the Cistrome database. In total**,** 117 DE TFs (9 down-regulated and 108 up-regulated) were detected. The total down-regulated and first ten up-regulated DE TFs were showed in Table S2. The volcano map and heatmap of the DE immune genes were presented in Fig. 3a and Fig. 3b. Besides, 81 positive DE immune genes-TFs pairs were detected in the network (Fig. 3c), involving 19 DE up-regulated immune genes and 40 DE TFs.

Construction of a model

We firstly screened 39 immune genes with univariate Cox analysis (Fig. 4a). Subsequently, we got seven suitable prognostic immune genes using Lasso regression (Fig. 4b and 4c). Ultimately, we identified five immune genes, including HSPA4, ISG20L2, NDRG1, EGF, and IL17D, all of which were high hazard genes (Fig. 4d and 4e). The risk score = (0.0412 ×HSPA4 expression level) + (0.0932 × ISG20L2 expression level) + (0.0062 × NDRG1 expression level) + (0.3969 × EGF expression level) + (0.0746 × IL17D expression level).

All HCC patients were classified into a high-risk group (n = 71) and a low-risk group (n = 71) on the basis of the median risk score. OS was significantly different between the two risk groups (P = 7.953e-06) (Fig. 5a). The AUC for the model at 1 and 3 years of overall survival was 0.849 and 0.74, respectively (Fig. 5b). We sequenced the risk scores of HCC patients and examined their distribution in Fig. 5c. The survival status of HCC patients was displayed on the dot plot (Fig. 5d). The heatmap revealed the expression patterns of prognostic immune genes between two groups (Fig. 5e).

Validation of the model

The risk score of each HCC patient was determined in the testing and entire groups and subsequently classified into two subgroups. The survival curves were significantly different between the high-risk and low-risk subgroups in the two cohorts (all P < 0.05) (Fig. 6a and c). The AUCs in the testing group and the entire group at 3-year were 0.651 and 0.682, respectively (Fig. 6b and d).

The prognostic value of the model

In the entire group, the risk score and the pathological stage were identified to closely associate with OS (all P < 0.001) (Fig. 7a and b). These results revealed that the model, as well as the pathological stage,

were independent prognostic factors. Interestingly, further comparison demonstrated that the risk score could be more accurate in predicting OS at one and three years, compared with the pathological stage. The AUCs at one year for the pathological stage and the risk score were 0.674 and 0.804, respectively (Fig. 7c), and those at three years were 0.676 and 0.682, respectively (Fig. 7d).

Applicability of the model

In the entire group, the values of several factors (EGF, HSPA4, IL17D, ISG20L2, and the risk score) were positively related to the histological grade of HCC (all $P < 0.05$) (Fig. 8a–e). The expression level of IL17D was higher in females than in males ($P < 0.05$) (Fig. 8f). Besides, as NDRG expression increased, the value of the pathological stage decreased ($P < 0.05$) (Fig. 8g). Also, the risk score was closely associated with the content of all the immune cells in HCC samples (all $P < 0.05$) (Fig. 9a-f).

Discussion

HCC is a common tumor, which is highly malignant and recurrent. Currently, it is urgent to find a molecular model to achieve the early diagnosis and prognosis prediction due to the needs of an individualized and precise treatment. In this study, we identified an immune gene model, which might be treated as an independent prognostic factor. Also, the immune gene model was closely related to clinical factors and could reflect the tumor immune microenvironment of HCC. Besides, GO and KEGG analyses of the DE immune genes and the network between the DE immune genes and TFs were conducted, which could indicate the direction for the future basic research of HCC.

In this article, we constructed a five-DE immune gene model, comprising HSPA4, ISG20L2, NDRG1, EGF, and IL17D. All the DE genes were detected between HCC and normal tissues. Therefore, this model could simultaneously contribute to the early diagnosis of HCC. Besides, all five DE immune genes are expected to be new molecular targets for immunotherapy. HSPA4, also called Apg-2, is a member of the HSP110 family. It is expressed in many organs [15] and can be induced by various conditions, including oncogenic stress. Gotoh et al. [16] demonstrated that HSPA4 was overexpressed in HCC. Duzgun et al. revealed that the overexpression of HSPA4 was correlated with worse OS in head and neck squamous cell carcinoma and breast invasive carcinoma [17]. NDRG1 was demonstrated to be a biomarker for metastasis and indicated poor prognosis in HCC [18,19], which is in line with our study. Lu et al. [20] also found NDRG1 was up-regulated in HCC patients and could be used as a potential therapeutic target for HCC. ISG20L2, as a target of miR-139-3p, was also found related to HCC prognosis in a previous study [21].

In this study, the risk score was detected as an independent prognostic variable, which could even provide better predictions than the pathological stage. Also, we noticed that the risk score was closely correlated with the histological grade and the pathological stage. These results similarly confirm the dependability of the risk model. Previous reports have shown that immune infiltration is a vital factor affecting the efficacy as well as the prognosis of HCC [22,23]. Ma et al. [24] reported that $PD1^{Hi}$ $CD8^{+}$ T cells correlated

with poor clinical outcome in HCC. In this article, we discovered that the risk score was highly associated with the infiltration of all the immune cells. All the results above demonstrated the prognostic reliability of our model for HCC.

In this study, we analyzed the enriched DE immune genes and constructed a network based on the DE immune genes and TFs. Inflammatory response, immune response, and growth factor activity were significantly enriched GO terms. It is known that HCC often occurs in the context of chronic inflammatory liver disease [25]. The interaction between cancer cells and tumor microenvironment can maintain tumor proliferation, invasion, and metastasis by providing growth factors to generate immune response conducive to tumor progression [26]. The enriched KEGG terms included cytokine-cytokine receptor interaction and antigen processing. As an inflammatory pathway, cytokine-cytokine receptor interaction was highly associated with the progress of HCC [27]. Antigen processing and presentation were also found to be an enriched pathway in HCC in a previous study [28]. We also identified the DE cancer TFs and detected the DE immune gene-TF pairs with high correlation. The enriched functions and pathways, as well as highly related TFs, could provide ideas for future experiments.

The present article has some advantages. First of all, we built the model using multiple algorithms and verified it with two validating groups. Therefore, our risk model of HCC was truthful and trustworthy. Furthermore, the risk score could independently and perfectly predict the prognosis of HCC. At last, our risk model could partly reflect the tumor immune microenvironment of HCC.

The current study has its limitations. First, we established a prognostic risk model based on the public databases, and it was not confirmed in the clinical researches. In addition, the identified DE immune genes and TFs, as well as the enriched functions and pathways, require further research.

Conclusion

We identified a five-immune gene model, which could be treated as an independent prognostic factor of HCC.

Abbreviations

HCC
hepatocellular carcinoma; OS:overall survival; ICB:immune checkpoint blockade; DE:differentially expressed; TCGA:The Cancer Genome Atlas; TFs:transcription factors. FDR:false discovery rate; FC:fold change; *DAVID*:The Database for Annotation, Visualization, and Integrated Discovery; GO:gene ontology; KEGG:Kyoto Encyclopedia of Genes and Genomes. ROC:receiver operating characteristic; AUC:area under the ROC.

Declarations

Acknowledgments
None.

Authors' contributions
HTC and JCG conceived the study and performed the bioinformatics analyses. HTC and JCG downloaded and organized the clinical and gene expression data. HTC performed the statistical analyses. HTC and JCG wrote the manuscript. JCG critically revised the article for essential intellectual content and administrative support. All authors read and approved the final manuscript.

Funding
This work was not funded by any grant.

Availability of data and materials
All analyzed data are included in this published article and its supplementary information file. The original data are available upon reasonable request to the corresponding author.

Ethics approval and consent to participate
Not applicable.

Consent for publication
All the authors agree to the publication clause.

Competing interests
The authors declare that they have no competing interests.

References

1. Cronin KA, Lake AJ, Scott S, Sherman RL, Noone AM, Howlader N, et al. Annual Report to the Nation on the Status of Cancer, part I: National cancer statistics. Cancer. 2018;124:2785 – 800.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424.
3. Zongyi Y, Xiaowu L. Immunotherapy for hepatocellular carcinoma. Cancer Lett. 2019;470:8–17.
4. Zhang X, Li J, Ghoshal K, Fernandez S, Li L. Identification of a Subtype of Hepatocellular Carcinoma with Poor Prognosis Based on Expression of Genes within the Glucose Metabolic Pathway. Cancers (Basel). 2019;11.
5. Zhu R, Yang X, Guo W, Xu XJ, Zhu L. An eight-mRNA signature predicts the prognosis of patients with bladder urothelial carcinoma. PeerJ. 2019;7:e7836.
6. Siu LL, Ivy SP, Dixon EL, Gravell AE, Reeves SA, Rosner GL. Challenges and Opportunities in Adapting Clinical Trial Design for Immunotherapies. Clin Cancer Res. 2017;23:4950–8.
7. Tai D, Choo SP, Chew V. Rationale of Immunotherapy in Hepatocellular Carcinoma and Its Potential Biomarkers. Cancers (Basel). 2019;11.

8. Joerger M, Guller U, Bastian S, Driessen C, von Moos R. Prolonged tumor response associated with sequential immune checkpoint inhibitor combination treatment and regorafenib in a patient with advanced pretreated hepatocellular carcinoma. J Gastrointest Oncol. 2019;10:373–8.
9. Agdashian D, ElGindi M, Xie C, Sandhu M, Pratt D, Kleiner DE, et al. The effect of anti-CTLA4 treatment on peripheral and intra-tumoral T cells in patients with hepatocellular carcinoma. Cancer Immunol Immunother. 2019;68:599–608.
10. Liu F, Zeng G, Zhou S, He X, Sun N, Zhu X, et al. Blocking Tim-3 or/and PD-1 reverses dysfunction of tumor-infiltrating lymphocytes in HBV-related hepatocellular carcinoma. Bull Cancer. 2018;105:493–501.
11. Zhou G, Sprengers D, Boor PPC, Doukas M, Schutz H, Mancham S, et al. Antibodies Against Immune Checkpoint Molecules Restore Functions of Tumor-Infiltrating T Cells in Hepatocellular Carcinomas. Gastroenterology. 2017;153:1107-19.e10.
12. Li W, Wang H, Ma Z, Zhang J, Ou-Yang W, Qi Y, et al. Multi-omics Analysis of Microenvironment Characteristics and Immune Escape Mechanisms of Hepatocellular Carcinoma. Front Oncol. 2019;9:1019.
13. Dong Z, Chen Y, Yang C, Zhang M, Chen A, Yang J, et al. STAT gene family mRNA expression and prognostic value in hepatocellular carcinoma. Onco Targets Ther. 2019;12:7175–91.
14. Cho SH, Pak K, Jeong DC, Han ME, Oh SO, Kim YH. The AP2M1 gene expression is a promising biomarker for predicting survival of patients with hepatocellular carcinoma. J Cell Biochem. 2019;120:4140–6.
15. Kaneko Y, Kimura T, Kishishita M, Noda Y, Fujita J. Cloning of apg-2 encoding a novel member of heat shock protein 110 family. Gene. 1997;189:19–24.
16. Gotoh K, Nonoguchi K, Higashitsuji H, Kaneko Y, Sakurai T, Sumitomo Y, et al. Apg-2 has a chaperone-like activity similar to Hsp110 and is overexpressed in hepatocellular carcinomas. FEBS Lett. 2004;560:19–24.
17. Duzgun MB, Theofilatos K, Georgakilas AG, Pavlopoulou A. A Bioinformatic Approach for the Identification of Molecular Determinants of Resistance/Sensitivity to Cancer Thermotherapy. Oxid Med Cell Longev. 2019;2019:4606219.
18. Cheng J, Xie HY, Xu X, Wu J, Wei X, Su R, et al. NDRG1 as a biomarker for metastasis, recurrence and of poor prognosis in hepatocellular carcinoma. Cancer Lett. 2011;310:35–45.
19. Luo Q, Wang CQ, Yang LY, Gao XM, Sun HT, Zhang Y, et al. FOXQ1/NDRG1 axis exacerbates hepatocellular carcinoma initiation via enhancing crosstalk between fibroblasts and tumor cells. Cancer Lett. 2018;417:21–34.
20. Lu WJ, Chua MS, So SK. Suppressing N-Myc downstream regulated gene 1 reactivates senescence signaling and inhibits tumor growth in hepatocellular carcinoma. Carcinogenesis. 2014;35:915–22.
21. Zhu Y, Zhou C, He Q. High miR-139-3p expression predicts a better prognosis for hepatocellular carcinoma: a pooled analysis. J Int Med Res. 2019;47:383–90.

22. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science. 2016;351:1463–9.
23. Chen QF, Li W, Wu PH, Shen LJ, Huang ZL. Significance of tumor-infiltrating immunocytes for predicting prognosis of hepatitis B virus-related hepatocellular carcinoma. World J Gastroenterol. 2019;25:5266–82.
24. Ma J, Zheng B, Goswami S, Meng L, Zhang D, Cao C, et al. PD1(Hi) CD8(+) T cells correlate with exhausted signature and poor clinical outcome in hepatocellular carcinoma. J Immunother Cancer. 2019;7:331.
25. Sia D, Jiao Y, Martinez-Quetglas I, Kuchuk O, Villacorta-Martin C, Castro de Moura M, et al. Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features. Gastroenterology. 2017;153:812–26.
26. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–74.
27. Souza T, Jennen D, van Delft J, van Herwijnen M, Kyrtoupolos S, Kleinjans J. New insights into BaP-induced toxicity: role of major metabolites in transcriptomics and contribution to hepatocarcinogenesis. Arch Toxicol. 2016;90:1449–58.
28. Zhang C, Peng L, Zhang Y, Liu Z, Li W, Chen S, et al. The identification of key genes and pathways in hepatocellular carcinoma by bioinformatics analysis of high-throughput data. Med Oncol. 2017;34:101.

Figures

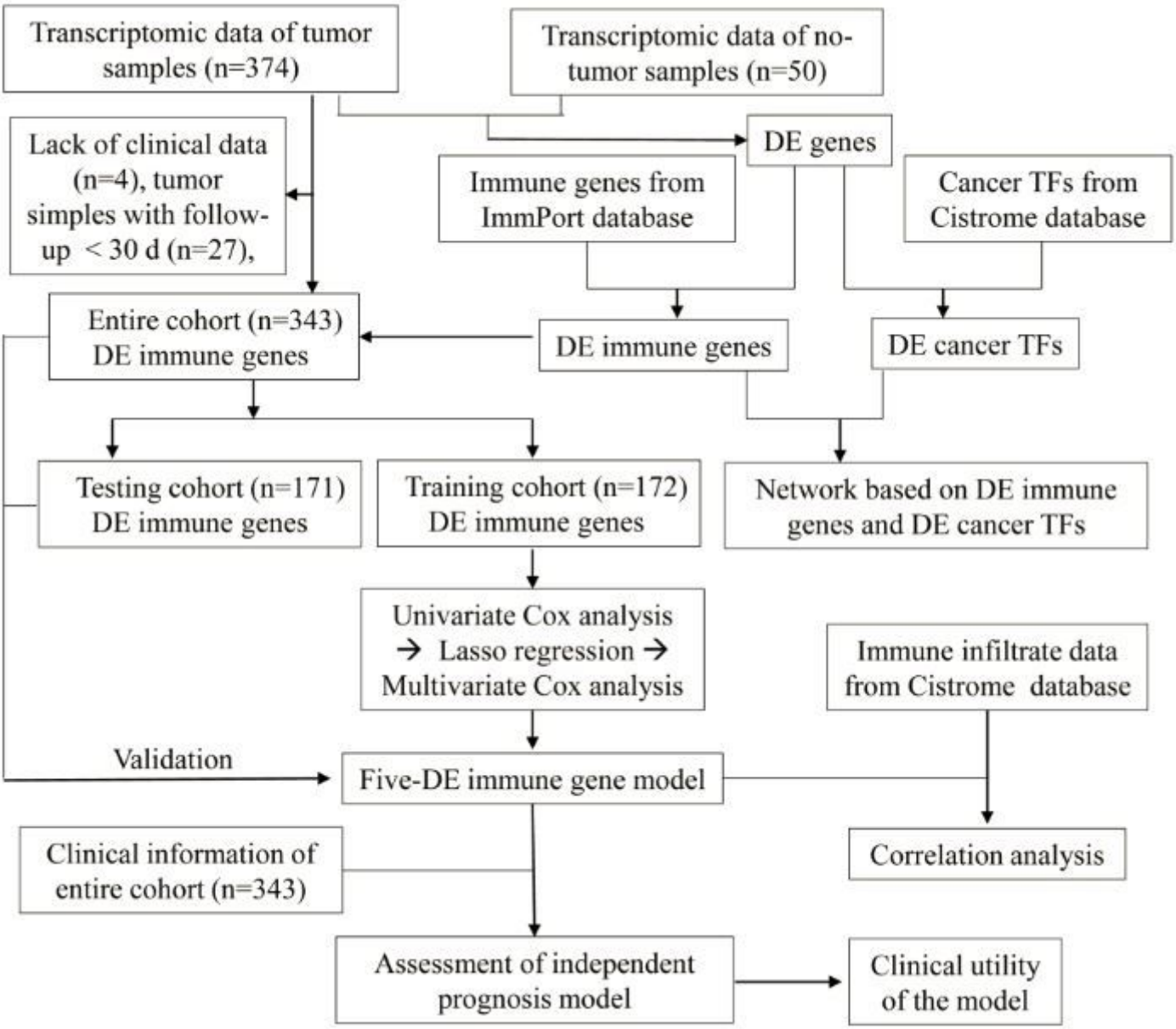

Figure 1

The workflow for the study. DE, differentially expressed; TFs, transcription factors.

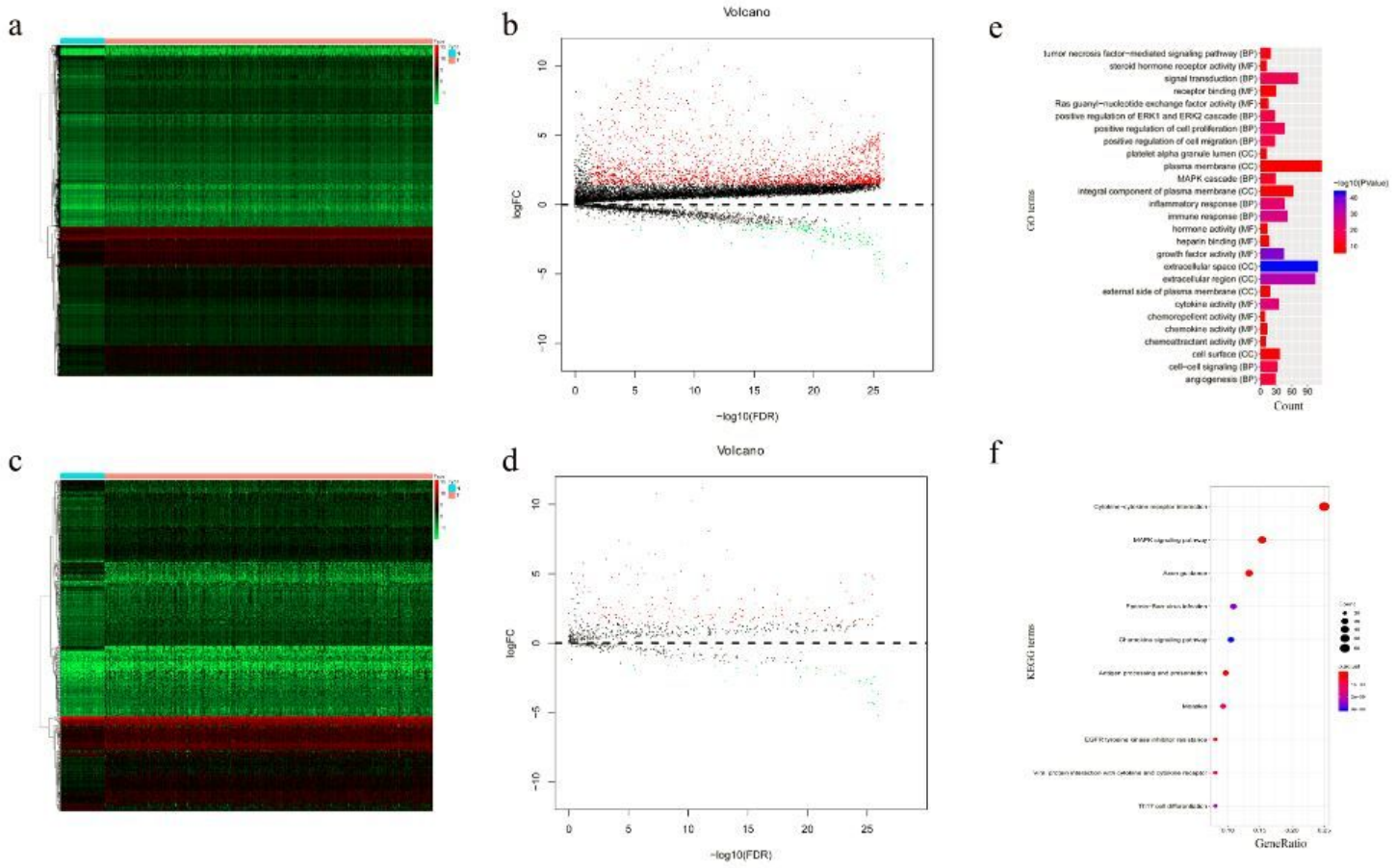

Figure 2

Detection of the differentially expressed (DE) immune genes. (a, b) The heatmap and the volcano plot showing the DE genes. (c, d) The heatmap and the volcano plot showing the DE immune genes. (e, f) Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis showing the DE immune genes.

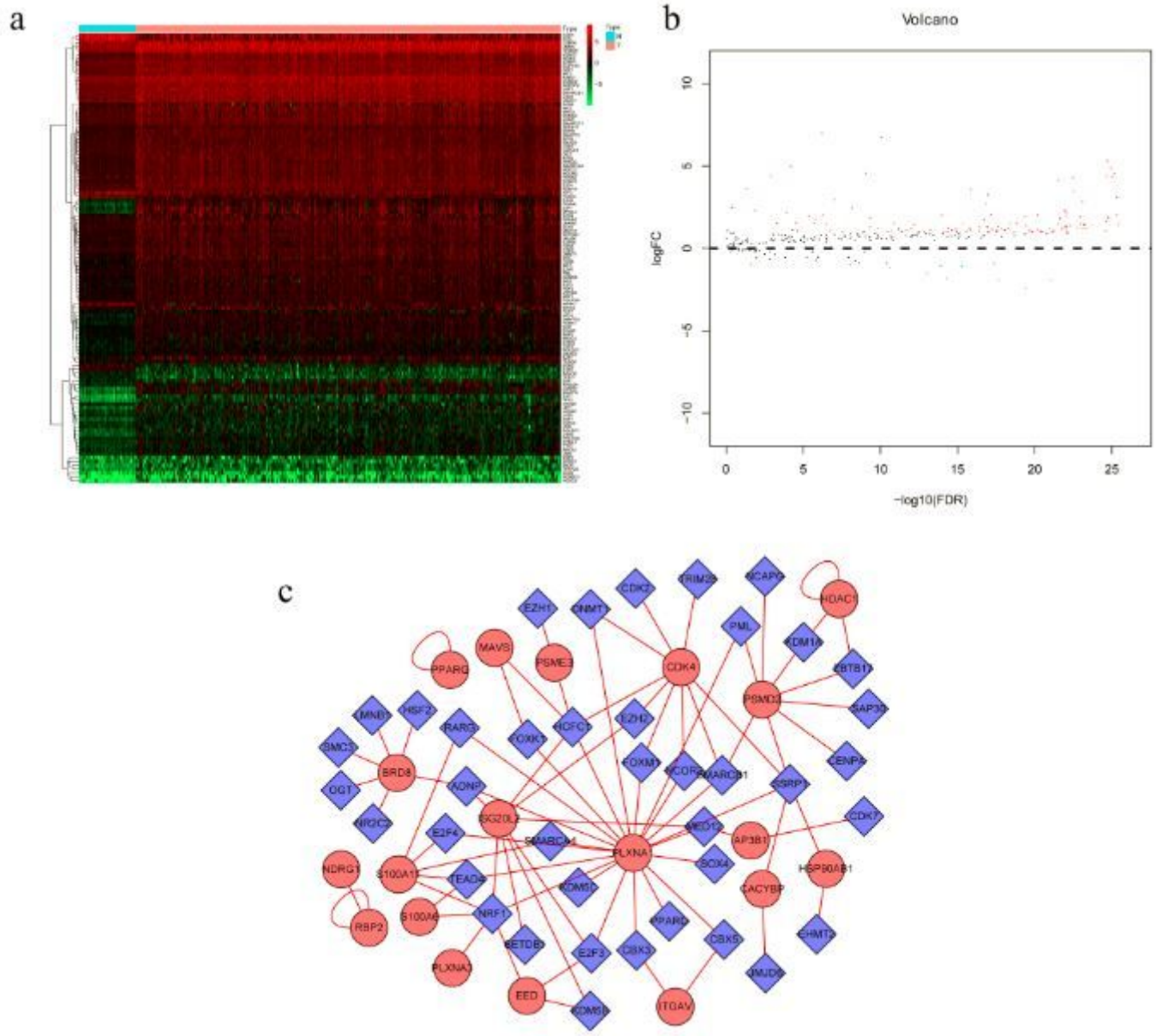

Figure 3

Construction of a differentially expressed (DE) immune genes-transcription factors (TFs) network. (a) The heatmap showing the DE TFs. (b) The volcano plot showing the DE TFs. (c) A network including the DE immune genes and their related TFs.

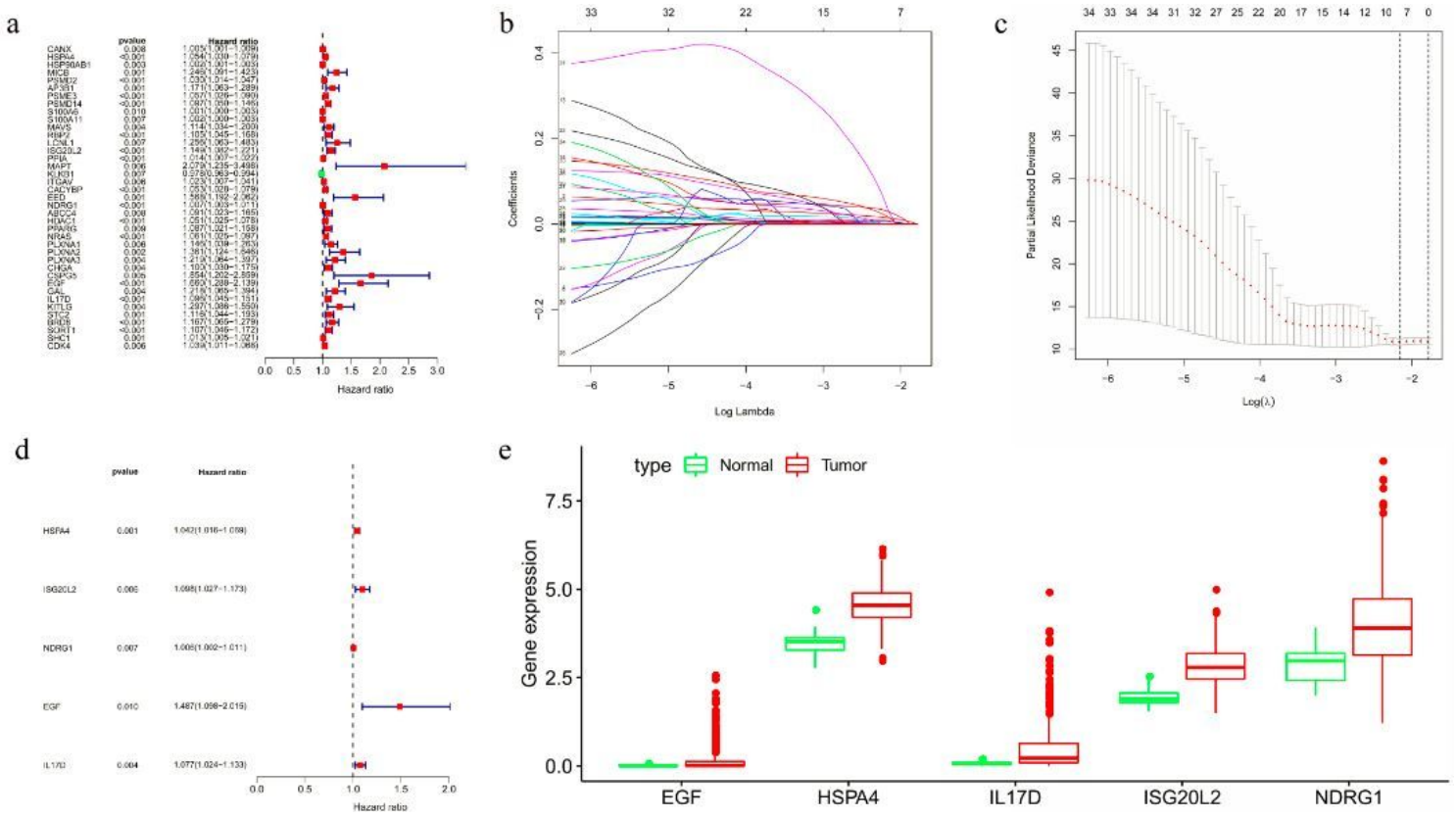

Figure 4

Identification of a model in the training group. (a) Screening of potential prognostic genes with univariate analysis. (b, c) Further analysis of possible prognostic genes with Lasso regression. (d) Identification of optimal immune genes via multivariate analysis. (e) The expression level of the optimal immune genes of the model.

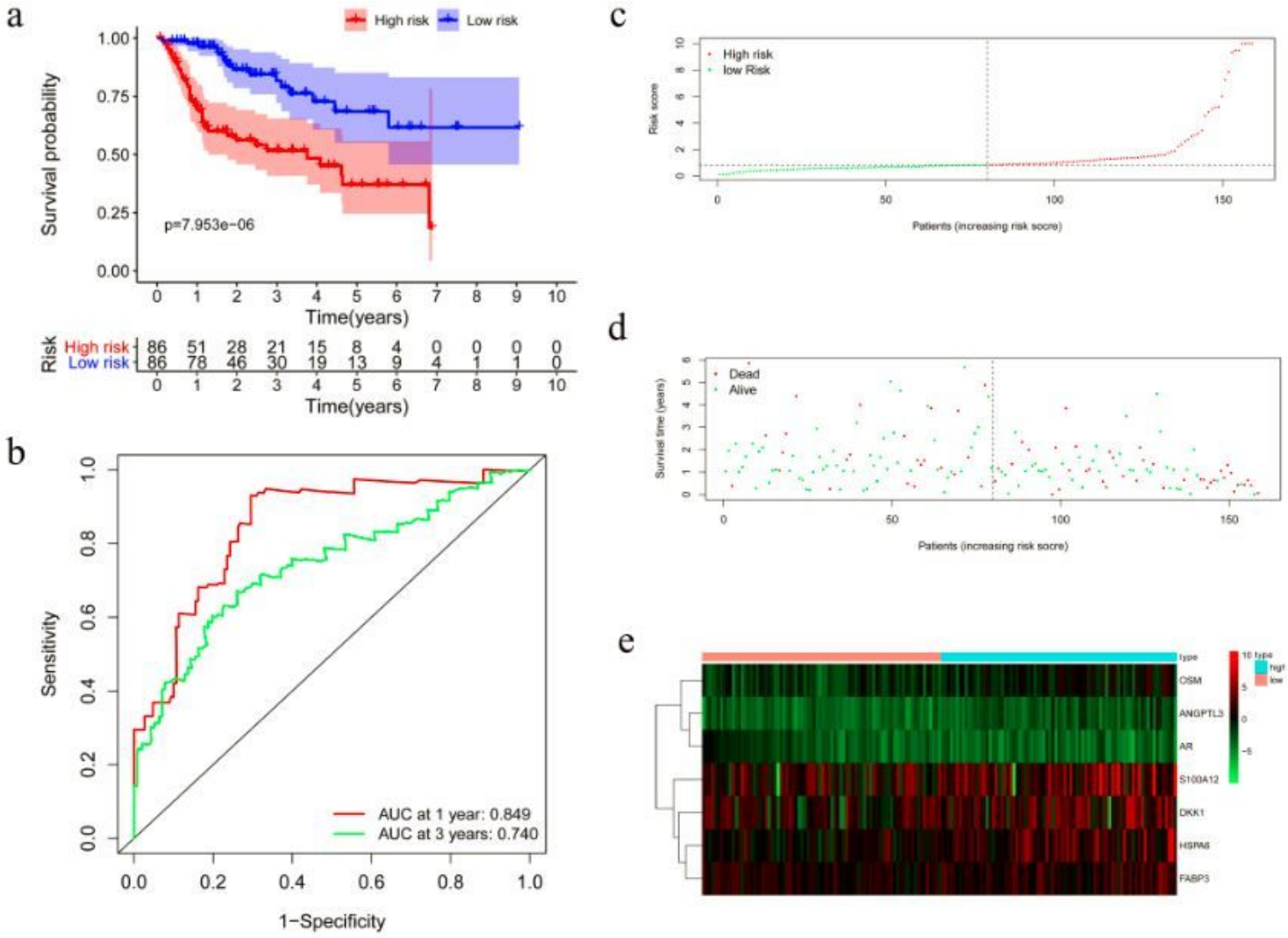

Figure 5

Model assessment in the training group. (a) Survival analysis between the high-risk and low-risk subgroups. (b) The time-dependent receiver operating characteristic (ROC) curve for the prognostic risk model at 1 and 3 years. (c) The risk score distribution of HCC patients. (d) Survival status scatter plots of hepatocellular carcinoma (HCC) patients. (e) Heatmap of the risk immune genes.

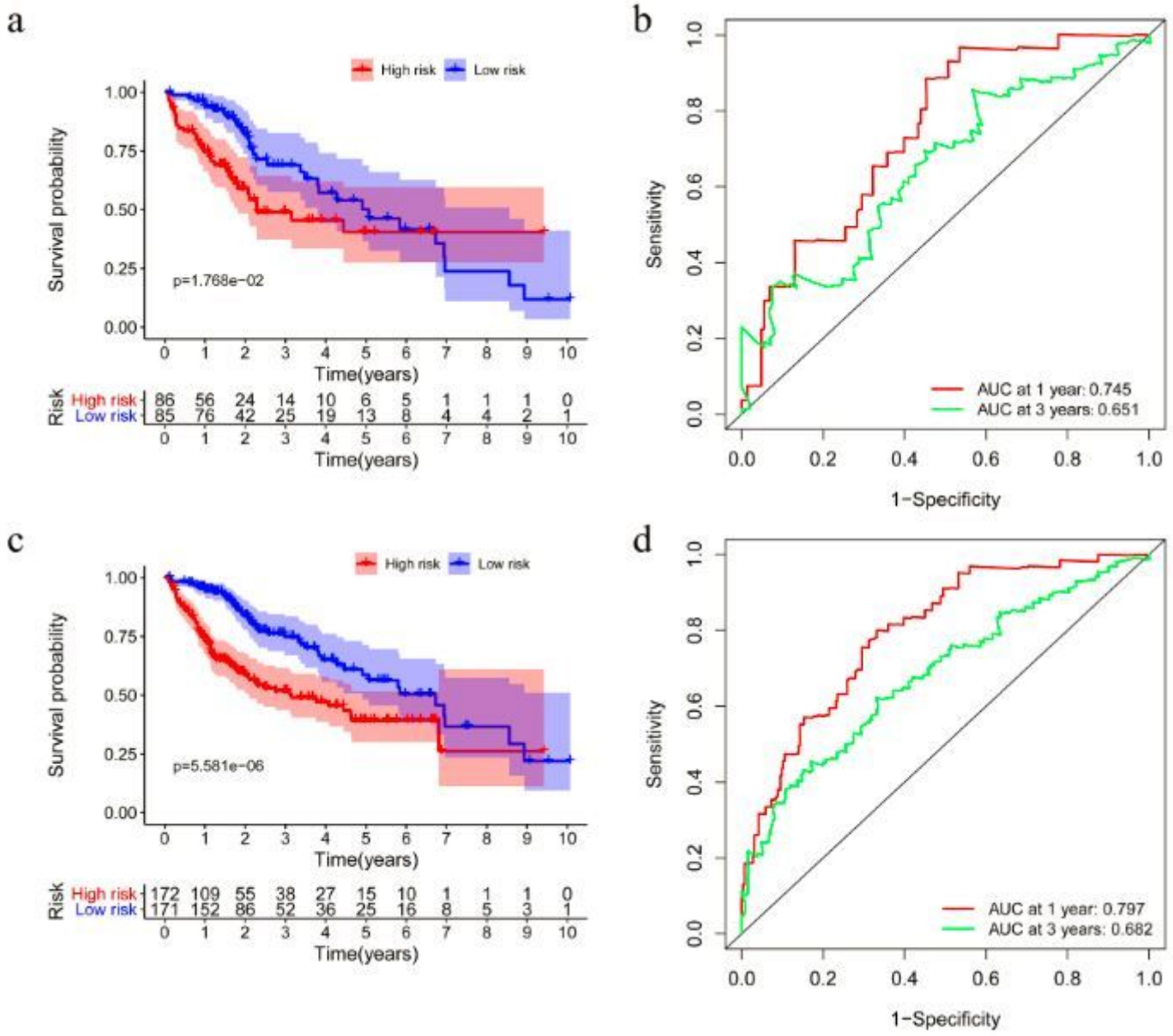

Figure 6

Validation of the prognostic model. (a) Overall survival (OS) of high-risk and low-risk patients in the testing group. (b) The time-dependent receiver operating characteristic (ROC) curve analysis of the testing group. (c) OS in the entire group. (d) The ROC curve analysis in the entire group.

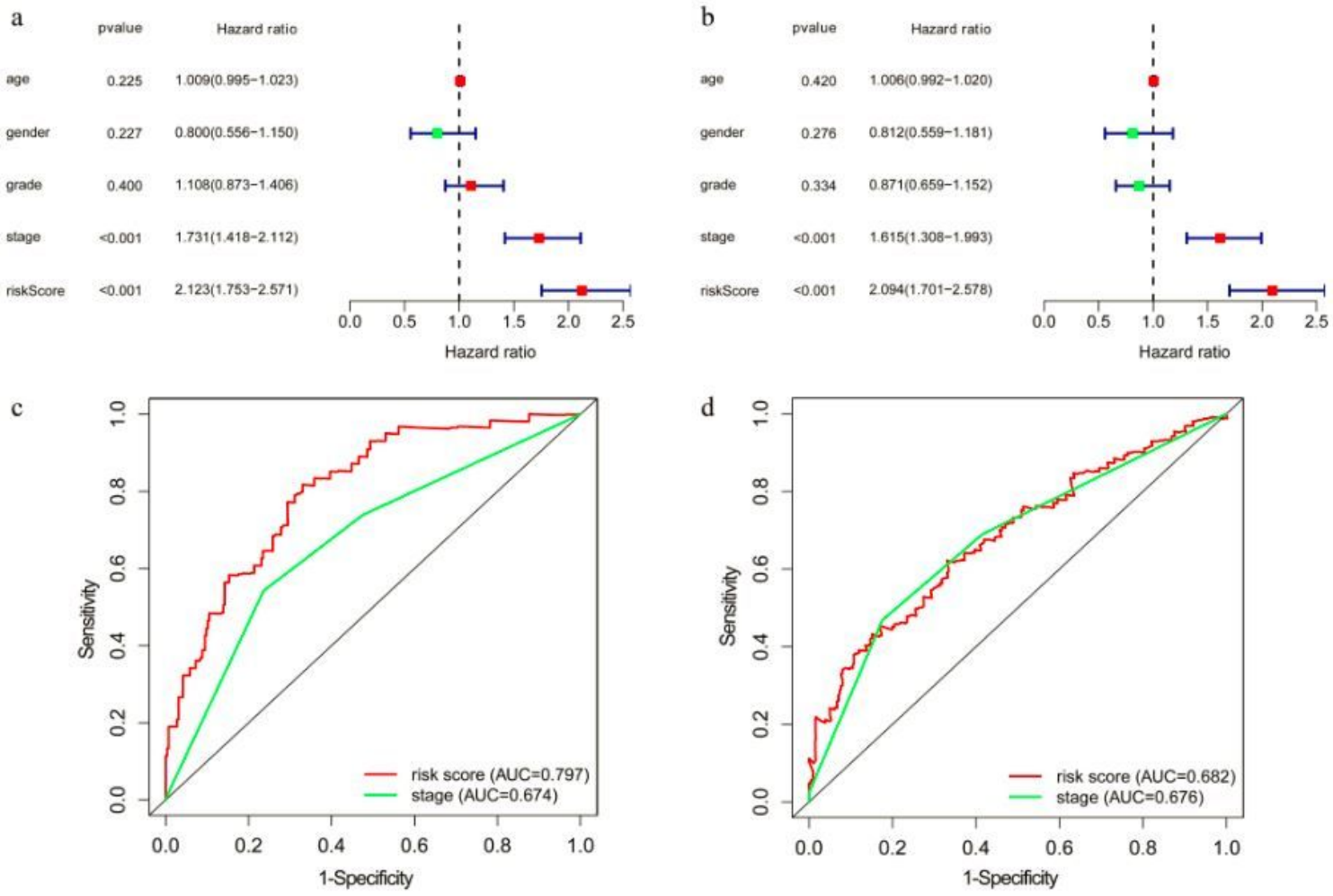

Figure 7

The prognostic value of the model. (a, b) Univariate and multivariate analyses of the entire group. (c, d) The time-dependent receiver operating characteristic (ROC) curve for two independent prognostic factors in the entire group at one year and three years.

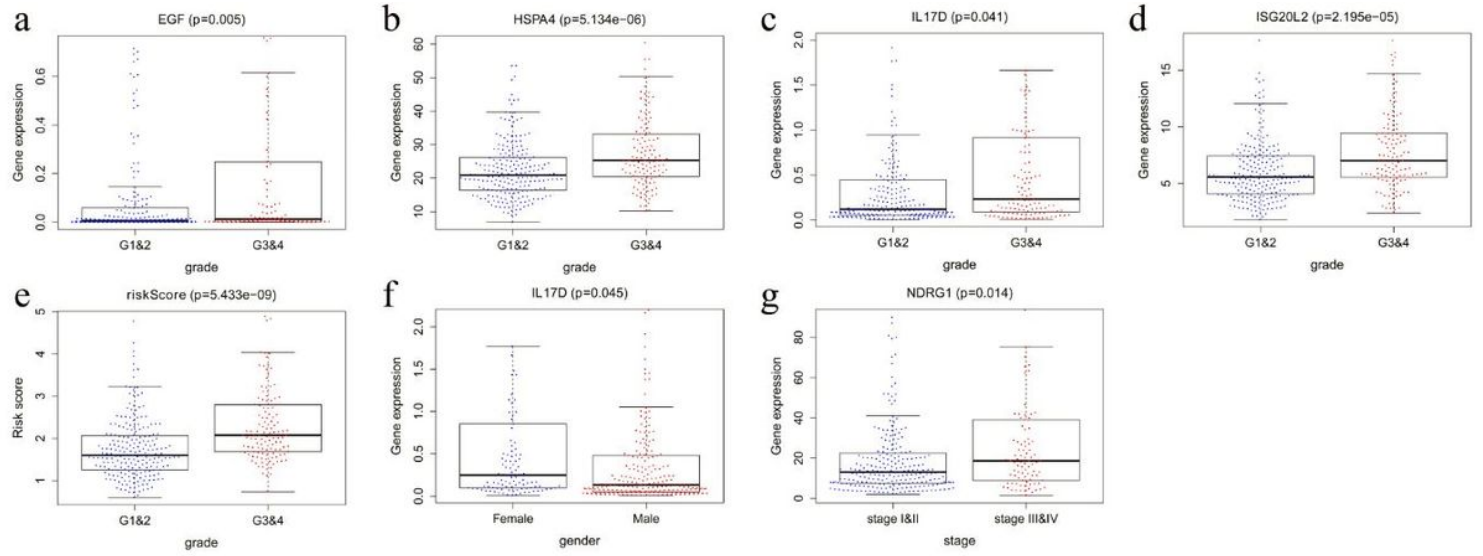

Figure 8

Clinical utility of the model in the entire group. (a-d) The correlation of EGF, HSPA4, IL17D, and ISG20L2 expression with the histological grade. (e) The relationship of the risk score with the histological grade. (f) The correlation of IL17D expression with gender. (g) The association of NDRG1 expression with the pathological stage.

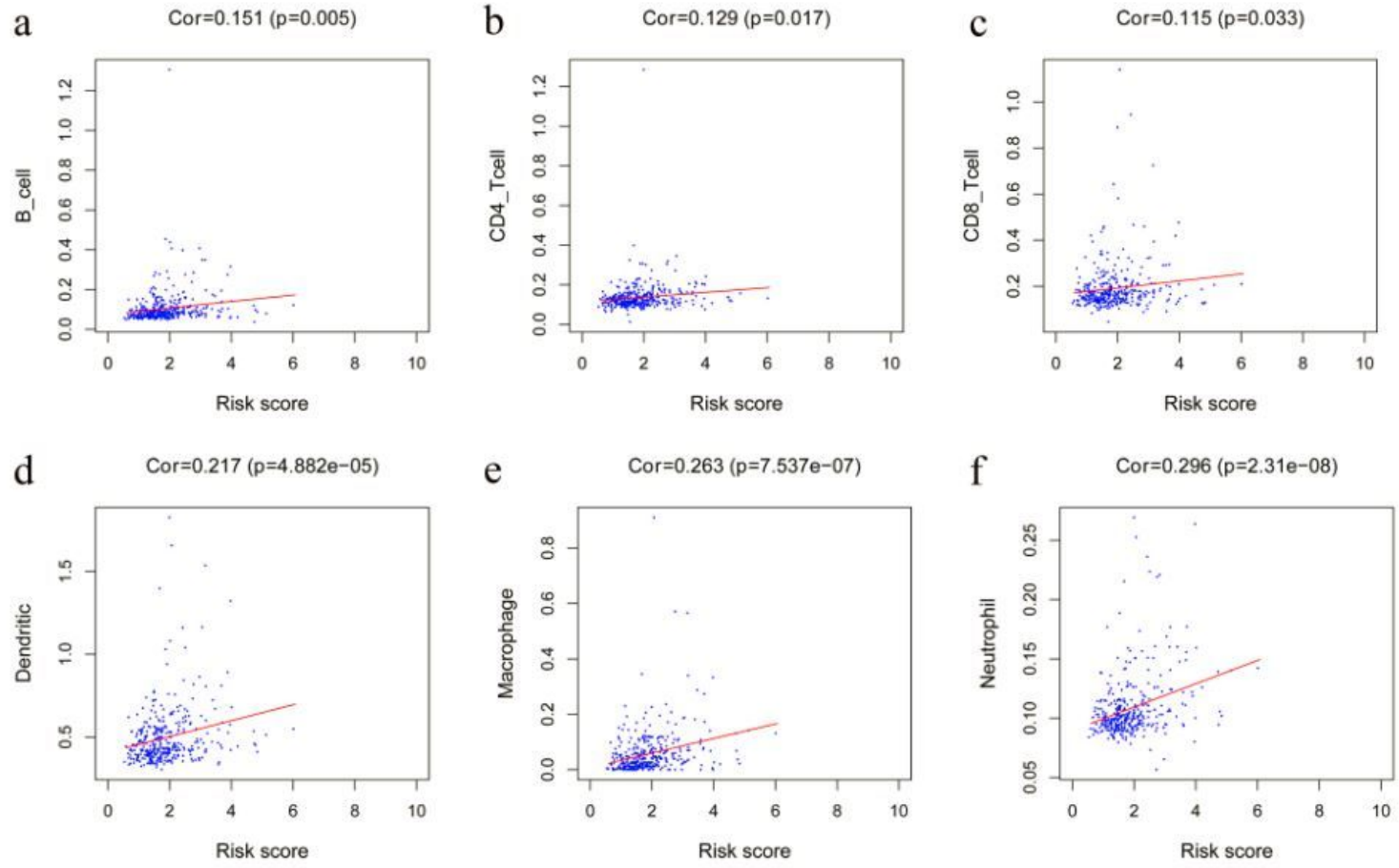

Figure 9

Correlation of the risk score with immune cell infiltration in the entire group.

Supplementary Files

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

- TableS2.docx
- TableS1.docx