A novel RNA binding protein-associated prognostic model to predict overall survival in hepatocellular carcinoma patients

Ye Liu
Wuhan University Zhongnan Hospital
Zhixiang Qin
Wuhan University Zhongnan Hospital
Hai Yang
Wuhan University Zhongnan Hospital
Yang Gu
Wuhan University Zhongnan Hospital
Kun Li (quentin2016@sina.com)
Wuhan University Zhongnan Hospital https://orcid.org/0000-0001-5469-6240

Primary research

Keywords: RNA binding proteins, Hepatocellular carcinoma, overall survival, prognostic model

Posted Date: July 2nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-37845/v1

Version of Record: A version of this preprint was published at Medicine on July 23rd, 2021. See the published version at https://doi.org/10.1097/MD.0000000000026491.

Abstract

Background Hepatocellular carcinoma (HCC) represents one of the deadliest malignancies worldwide. Despite significant advances in diagnosis and treatment, the mortality rate from HCC persists at a substantial level. This research strives to establish a prognostic model based on the RNA binding proteins (RBPs) that can predict HCC patients' OS. Methods There was an RNA-seq data set derived from the Cancer Genome Atlas (TCGA) databank which was included in our research as well as a Microarray data set (GSE14520). The differentially expressed RBPs between HCC and normal tissues were investigated in TCGA dataset. Subsequently, the TCGA data set was randomly split into a training and a testing cohort. The prognostic model of the training cohort was developed by applying univariate Cox regression and lasso Cox regression analyses and multivariate Cox regression analysis. In order to evaluate the prognostic value of the model, a comprehensive survival assessment was conducted. Results A total of 133 differentially expressed RBPs were identified. Five RBPs (RPL10L, EZH2, PPARGC1A, ZNF239, IFIT1) were used to construct the model. The model accurately predicted the prognosis of liver cancer patients in both the TCGA cohort and the GSE14520 validation cohort. HCC patients could be assigned into a high-risk group and a low-risk group by this model, and the overall survival of these two groups was significantly different. Furthermore, the risk scores obtained by our model were highly correlated with immune cell infiltration. . Conclusions Five RBPs-related prognostic models were constructed and validated to predict OS reliably in HCC individuals. It helps to identify patients at high risk of mortality with the risk prediction score, which optimizes personalized therapeutic decision-making.

Background

The fourth most prevalent cause of cancer-related mortality is liver cancer, which has the sixth highest incidence in the world. [1]. Approximately 90% of hepatic carcinomas originated from hepatocellular carcinoma (HCC) [2, 3]. It has been established that a wide range of risk factors for tumor development in HCC, such as nonalcoholic fatty liver, hepatitis B and C virus, alcohol consumption and diabetes mellitus [4]. The 5-year survival rate of hepatic carcinomas is 18%, which is the next most deadly malignancy following the pancreatic carcinoma. Patients from Asian countries, such as China, are in a worse situation, with a reported 5-year survival rate as low as approximately 12% [5]. It is quite frequent for HCC to recur after hepatectomy, with a recurrence rate of up to 70% at 5 years [6]. Despite great efforts to improve the early detection rate and develop new treatment strategies, patients with liver cancer still face poor prognosis, especially in patients with advanced metastatic HCC. At present, the accuracy of histopathological diagnosis in clinical prognosis prediction is insufficient, which limits the treatment of HCC. Therefore, it is critical to develop a high-precision molecular prediction tool in the future clinical practice.

It is generally acknowledged and accepted that RNA-binding proteins (RBPs) act to bind RNA via one or more spherical RNA-binding domains and modify the destiny or function of the bound RNAs [7]. RBPs can act with diverse kinds of RNA, such as mRNAs, tRNAs, snoRNAs, snRNAs and ncRNAs. It is assumed that roughly 50% of RBPs contribute directly or indirectly to the post-transcriptional modulation in gene

expression [8]. RBPs can form different ribonucleoprotein complexes to regulate RNA splicing, localization, stability, translation, polyadenylation, and degradation [9]. Given the critical role of RBPs in post-transcriptional modulation, it is unsurprising that RBPs are highly associated with a variety of biological functions and diseases [10]. By modulating the mRNAs of many oncogenes, growth factors, and cell cycle modulators, RBPs affect the expression patterns of cancer-related genes, as is well established. [11]. Liang et al. reported that overexpression of IFITM3 was associated with poor outcome in HCC cases, and inhibition of IFITM3 could restrain HCC cell proliferation, migrations, invasion and apoptosis [12]. Dong et al. found that RBM3 overexpression indicated that HCC patients had short relapse free survival and poor overall survival [13].

While several studies have investigated the relationship between RBPs and outcomes in patients with HCC, most have focused on the impact of a single gene on prognosis. Consequently, we developed a reliable model of prognostic risk to determine the outcome of patients with HCC. Moreover, we examined the association between risk factors (risk scores and risk genes) derived from the model and clinical characteristics.

Methods

Data source

The expression profile data of RNA in HCC patients and the respective clinical data had been derived from the Cancer Genome Atlas (TCGA) databank (https://portal.gdc.cancer.gov/repository). It consisted of 374 cases of HCC tissue and 50 cases of normal tissue. Another dataset used for validation, GSE14520, was obtained from the Gene Expression Omnibus (GEO) databank (https:/ /www.ncbi.nlm.nih.gov/geo/). Immune infiltration data was derived from the Cistrome project (http://www.cistrome.org/), including six types of tumor-infiltrating immune cells (B cells , CD4+ and CD8+ T cells, neutrophils , dendritic cells and macrophages). [14, 15]. 1542 RBPs summarized by predecessors were included in our study[16].

Identification of differentially expressed RBPs

MRNAs in RNA expression profile were annotated by human gene annotation files (GRCh38.99), which had been obtained from the Ensembl (http://asia.ensembl.org/index.html). The expression of mRNAs were analyzed by the edgeR package, the expression of mRNAs were normalized by method of Trimmed Mean of M[17]. Distinguish the differentially expressed mRNAs based on the criteria of |log 2 fold change| > 1 and FDR < 0.05. Then, RBPs in differentially expressed mRNAs were selected for further study. The heat map and the volcano map were drawn with pheatmap package and ggplot2 package.

Functional enrichment analysis and Protein-protein interaction (PPI) network construction of differentially expressed RBPs

With the Kyodo Encyclopedia of Genes and Genomes (KEGG) pathway assays and GO enrichment analysis, we investigated the possible biological functions of these differentially expressed RBPs.

Functional enrichment analysis results were obtained by clusterProfiler package and org.Hs.eg.db package. The enrichment results must meet the requirement that both p value was less than 0.05 to be considered statistically significant. The differential expressed RBPs were uploaded to the String database (https://string-db.org/) to obtain their interactions, and their PPI networks were subsequently visualized with Cytoscape software. [18, 19].

Establishment of the prognostic risk model

After removing RBPs not in GSE14520, a total of 81 differentially expressed RBPs remained for further study. Patients with a total survival time of less than one month and incomplete clinical information in the whole TCGA cohort were excluded. It was randomly divided into a training and a testing cohort for the entire TCGA cohort. Initially, univariate Cox regression analysis was performed to select the prognostic-associated RBPs, and $P < 0.05$ was considered to have significant difference. Then, Lasso regression was employed to further select prognostic-associated RBPs and delete prognostic-associated RBPs that correlated highly with one another[20]. Finally, Applications of multivariate Cox regression were conducted to develop a model of prognostic risk. Based on the regression coefficient from multivariate Cox regression analysis and mRNA expression level, the prognostic risk model was shown as risk score = ($Coefficient_{RBP1}$ × expression of RBP1) + ($Coefficient_{RBP2}$ × expression of RBP2) +⋯+ ($Coefficient_{RBPn}$ × expression of RBPn). The optimal cut-off values for risk scores were identified using the Survminer R package to classify patients into high- and low-risk catalogues. Outcome model accuracy was assessed by Kaplan-Meier (K-M) survival curves and the approach of the time-dependent receiver operating characteristic curve (ROC) analysis. The prognosis risk model was further evaluated employing the distribution of risk scores, the scatter plot of survival status and the expression heat map. Besides, the model of prognostic risk was varified in the testing, TCGA, GSE14520 cohorts.

Independent prognostic role of the prognostic risk model

Uni- and multi-variate Cox regression assays had been adopted to investigated whether the prognostic risk model might be independent of other clinical data (age, sex, histological grade, and pathological stage and risk score) for HCC patients. With the clinical characteristics as the independent variable and OS as the dependent variable, the ratio of hazard (HR) and 95 per cent confidence interval were measured. P value less than 0.05 was considered to have significant difference.

Building and validating a predictive nomogram

The present research constructed a nomogram for predicting the 1-year, 3-year, and 5-year OS probability of HCC patients utilizing independent prognostic factors that were selected by multivariate Cox regression assays. The accuracy of the nomogram was validated by comparing the prediction probability of the nomogram with the actual observation probability through the calibration curve. The better coincidence with the reference line indicated the higher accuracy of nomogram prediction. It had also been used to evaluate the predictive accuracy of the nomogram with ROC curves.

Statistical analysis

R software (version 3.6.1) and Perl (version 5.26.3) were used to analyze the RNA expression spectrum and respective clinical information data of HCC patients. The rank correlation between the risk score and level of immune infiltration was assessed with the Pearson's correlation coefficient test, the independent t-test was utilized to evaluate the differences between the variables. Statistical significance was identified by $p<0.05$.

Results

Screening of differentially expressed RBPs in HCC patients

RNA expression profiling of 424 samples (374 neoplasms and 50 normal tissues) for hepatocellular carcinoma patients were obtained from TCGA database. As shown in the heat map and volcano map (Additional file 1: Figure S1 A, B), 4657 differentially expressed mRNAs (3611 up-regulated and 1046 down-regulated) could be located in HCC tissues versus normal tissues. We then took the intersection of 1542 RBPs and 4657 differentially expressed mRNAs and got 133 differentially expressed RBPs, including 111 differentially expressed RBPs that were upregulated and 22 differentially expressed RBPS that were downregulated (Additional file 1: Figure S1 Figure1 C, D).

Functional enrichment analysis and protein-protein interaction (PPI) network establishment of differentially expressed RBPs

In order to reveal possible biological functions of differentially expressed RBPs, we performed GO term and KEGG pathway analysis. There were 247 pathways considering to be significantly enriched, including 152 biological process (BP) terms, 49 cellular component (CC) terms, 46 molecular function (MF) terms. The most significant BP, CC, MF were regulation of mRNA metabolic process, cytoplasmic ribonucleoprotein granule, catalytic activity and acting on RNA, respectively (Figure 1 A). Meanwhile, KEGG pathway analysis identified six significantly enriched pathways: mRNA surveillance pathway, RNA degradation, Spliceosome, RNA transport, Ribosome and Ribosome biogenesis in eukaryotes (Figure 1 B). To further investigate the potential interactions between differentially expressed RBPS, we developed a PPI network based on data from the STRING databank utilizing software Cytoscape (Figure 1 C). Figure1 D showed the top ten hub genes of the PPI network. The top ten hub genes were PABPC1, PIWIL1, ELAVL2, GSPT2, LIN28A, SNRPE, BOP1, DDX39A, DDX39B, DDX4 (Additional file 3: Table S1).

Identify the prognostic RBPs included in the risk model in the training cohort

Patients with an overall survival time of under 1 month and incomplete clinical data in the entire TCGA cohort were excluded, resulting in the inclusion of 319 patients in model construction. The entire TCGA cohort was split into a testing cohort (n = 159) and a training cohort (n = 160) randomly. Considering the validation in GSE14520, we included 81 differentially expressed RBPS in both training cohort and GSE14520 into the study. For the purpose of identifying the prognostic relevance of RBPs, approach of

univariate Cox regression was conducted to evaluate the expression of these RBPs in the training cohort, yielding 23 RBPs that were associated with prognosis (Figure 2 A). Lasso regression was employed to delete prognostic-associated RBPs that correlated highly with one another and identified 10 candidate prognostic-associated RBPs (Figure 2 B). Subsequently, all the candidate RBPs were measured by the approach of multivariate Cox regression assay. In the end, five RBPs were identified to establish the prognostic risk model. The five RBPs were ribosomal protein L10-like (RPL10L), enhancer of zeste homolog 2 (EZH2), peroxisome proliferator-activated receptor gamma, coactivator 1 alpha (PPARGC1A), zinc finger protein 239 (ZNF239) and interferon-induced protein with tetratricopeptide repeats 1 (IFIT1; Figure 2 C).

Establishment of the model for the prognostic risk in the training cohort

With the purpose of investigating the relevance of these five prospective RBPs in predicting the outcomes of patients with HCC, these five RBPs have been adopted to develop a prognostic model of risk. The risk score was determined by the following method: Risk score = (0.1400 × RPL10L expression) + (0.4536 ×EZH2 expression) + (-0.1195 × PPARGC1A expression) + (0.1537 ×ZNF239 expression) + (-0.2530 × IFIT1 expression). The optimal cut-off value of -0.142 for the risk score was determined utilizing the Survminer R package. Patients in the training group were divided to high-risk (n = 41) and low-risk (n = 119) groups in accordance with cut-off values. In accordance of Kaplan-Meier analysis, the high-risk group had remarkable poorer OS than the low-risk group ($p<0.001$; Figure 3 A). The prognostic value of the five RBPs characteristics was further assessed with time-dependent receiver operating characteristic curve (ROC) analysis. For 1-year survival, 3-year survival, and 5-year survival, the area under the ROC curve (AUC) was 0.763, 0.763 and 0.731, respectively. (Figure 3 B). The risk score analysis for the prognostic risk model in the training cohort was depicted in figure 3 C.

Validation of the prognostic risk model

With the aim of evaluating on whether the five RBPs prognostic risk models had parallel predictive value in cohorts of other HCC patients, the same formula was used to determine risk score for the testing cohort, the TCGA full cohort and the GSE14520 cohort separately. Individuals in the testing and the entire TCGA cohort were sorted to high- and low-risk groups in accordance with the optimal cut-off value for the training cohort. The best cutoff value of 1.111 in the GSE14520 cohort was obtained through the Survminer R package as the training cohort. Our finding was that in the testing cohort, TCGA cohort as well as GSE14520 cohort, patients in the high-risk group were in more adverse OS in comparison to the low-risk group. The 1-, 3-, and 5-year AUC of the testing cohort, TCGA cohort and GSE14520 cohort were 0.74, 0.731, 0.75 and 0.751, 0.747, 0.742 and 0.668, 0.679, 0.637, respectively (Figure 4). The distribution of risk scores, the scatter plot of survival status and the expression heat map for the three validation cohorts were shown in additional file 2: Figure S2. Prognosis for the high risk group has been shown to be unfavorable in comparison to the low risk group, which parallels the results of the training set. There was higher level of EZH2, RPL10L, ZNF239 expression in the high-risk category compared to the low-risk category, while PPARGC1A, IFIT1 level was reduced compared to the low-risk category. Taking all into

consideration, these findings suggested that our model of prognostic risk could reliably predict the OS of HCC patients.

Independent prognostic role of the prognostic risk model in the whole TCGA cohort

We evaluated the prognostic value of different clinical variables in the TCGA cohort of HCC patients by uni- and multi-variate Cox regression assay. Both uni- and multivariate Cox regression assays indicated that risk score and pathological stage were independent prognostic role. Among the parameters of age, gender, histological grade and pathological stage and risk score, 1-, 3-, and 5-year AUC of risk score were the largest and the hazard rate of the risk score was also the largest (Figure 5), which indicated that risk score predicted OS at 1,3 and 5 years more accurately than other clinical characteristics.

Building a predictive nomogram in the whole TCGA cohort

A nomogram created using two independent prognostic factors including pathological stage and risk score was used to predict OS at 1, 3 and 5 years in the whole TCGA cohort (Figure 6 A). ROC curve was used to evaluate the prediction accuracy of nomogram. The area under the ROC curve for 1-,3-, and 5-year was 0.792, 0.786, 0.763 (Figure 6 B), which manifested that the nomogram had very good prediction accuracy. Calibration curve analysis demonstrated that the 1-, 3-, and 5-year survival rates that were predicted by the nomogram corresponded well with the observed survival rates. (Figure 6 C, D, E).

Clinical utility of the prognostic risk model in the whole TCGA cohort

For the purposes of assessing the clinical utility of the predictive model, the correlation between risk factors (risk score and risk genes) derived from the model and the clinical properties of the entire TCGA cohort was assessed. As shown in Figure 7, compared with low histological grades, high histological grades had higher values of EZH2, ZNF239, and risk score, and lower values of IFIT1 and PPARGC1A (all $p<0.05$). Compared with low pathological stage, high pathological stage had higher values of EZH2 and risk score, and lower values of IFIT1 and PPARGC1A (all $p<0.05$). The value of IFIT1 for male was higher than that for female, and the value of ZNF239 was lower than female. The expression of RPL10L in elderly patients was greater than that of young patients. These findings demonstrated that these RBP genes were intimately associated with the development of HCC.

Tumor-infiltrating immune cells played a key function in the genesis and progress of tumors[21]. Therefore, the potential association between risk score and level of immune infiltration in the whole TCGA set was analyzed. As shown in Figure 8, risk score had a positive correlation with neutrophil, B cell, CD4/CD8+ T cells, dendritic, macrophage and (all $p<0.05$). These findings revealed that the score obtained by our model were highly correlated with immune cell infiltration.

Discussion

Various reports have established that RBPs aberrantly express themselves in diverse human diseases, including those of human malignancies. [12, 13]. Many RBPs have been defined as key molecules in the

occurrence and development of cancer. Abnormal expression of RBPs has also been strongly associated with the outcome of cancer victims. Therefore, RBPs may be essential for the progression and outcome of HCC.

In our study, a total of 133 differentially expressed RBPs were obtained by comparing liver cancer tissues with normal tissues based on the data from TCGA-LIHC. To investigate the possible biological functions of differentially expressed RBPs, GO term and KEGG pathway analyses were performed. Moreover, to further research the underlying interaction of differentially expressed RBPs, we constructed the PPI network and identified ten hub RBPs. These results will lay a foundation for the future study of the mechanism of HCC occurrence and progression. Next, 23 prognostic-associated RBPs in the training cohort were obtained by applying univariate Cox regression assay. Lasso- and multi-variate Cox regression analysis were used to find vital RBPs (RPL10L, EZH2, PPARGC1A, ZNF239, IFIT1) from the prognostic-associated RBPs. Subsequently, a prognostic model of RNA-binding protein-related was developed in accordance with the regression coefficients from multivariate Cox regression analysis and the mRNA expression status of the five RBPs. The K-M survival curves and time-dependent ROC curve analyses revealed that this model possessed good diagnostic ability and could screen out the patients with poor prognosis. Besides, we verified the stability and reliability of the model in the testing cohort, whole TCGA-LIHC cohort and GSE14520 cohort. It was shown by uni-and multivariate Cox regression assays that our model, as well as pathological stage, could serve as independent predictors of prognosis in HCC patients. Research has shown that pathological stage could independently predict the prognosis of liver cancer patients [22], which is consistent with our results. Further research found that our model predicted OS at 1-, 3-, and 5 years in liver cancer patients more accurately than other clinical parameters. We also found that risk factors (risk score and risk gene) obtained in the model were closely related to the progress of HCC, which, on the other hand, confirmed that our model had good prediction performance. These findings suggest that our prognostic risk model could reliably predict the OS of HCC patients.

In the current study, among these five RBPs, EZH2, PPARGC1A and IFIT1 have been reported essential to the progression and prognosis of cancer. Song et al. demonstrated increased levels of H3K27me3 by overexpression of EZH2 and silenced the Wnt signaling inhibitor expression, leading to initiation of Wnt/β-Catenin signaling and subsequent induction of cell proliferation and tumor development. [23]. Kido et al. conformed that PPARGC1A was beneficial to the survival of individuals with liver malignancies and had a negative correlation with the expression of the testis specific protein Y [24]. Zhang et al. identified elevated expression of IFIT1 as a positive outcome indicator of progression-free survival as well as the period of overall survival in glioblastoma. [25]. In our study, we found that PPARGC1A and IFIT1 were favorable factors for prognosis of HCC, while EZH2 was a risk factor for prognosis of HCC, which indicated that they had the prospect of becoming a new molecular target for liver cancer treatment.

Numerous researches and clinical studies have confirmed the significance of immune infiltration in solid neoplasms. [26]. Wei et al. reported that sperm-associated antigen 5 (SPAG5) was a marker of adverse outcome, and its expression was positively associated with the infiltration of neutrophils, CD8+ T cells, B cells , dendritic and macrophages. [27]. For the sake of evaluating whether our model had an impact on

immune cell infiltration in HCC, We performed an analysis of the correlation between risk score and exposure to immune infiltration levels. We found that risk score had a positive correlation with B cell, CD4+ T cell, CD8+ T cell, dendritic, macrophage and neutrophil, which also indicated that our model had good predictive performance.

However, we should admit some limitations of our research. Firstly, the current research results rely on gene mining methods and a prospective cohort study is required to validate the results. Additionally, the potential mechanisms of how the RBPs-associated gene affects the progression of liver cancer needs further study.

Conclusions

We developed and validated five RBPs-related models for prognosis to reliably predict OS in HCC patients. Scoring of risk prediction facilitates the screening of patients at high risk of mortality, thereby optimizing decision-making for individualized treatment.

Abbreviations

AUC: area under the curve; BP: biological process; CC: cellular component; EZH2: enhancer of zeste homolog 2; GEO: gene expression omnibus; HCC: hepatocellular carcinoma; HR: ratio of hazard; IFIT1: interferon-induced protein with tetratricopeptide repeats 1; KEGG: Kyodo Encyclopedia of Genes and Genomes; K-M: Kaplan-Meier; MF: molecular function; OS: overall survival; PPARGC1A: peroxisome proliferator-activated receptor gamma, coactivator 1 alpha; PPI: Protein-protein interaction; RPL10L: ribosomal protein L10-like; ROC: receiver operating characteristic; SPAG5: sperm-associated antigen 5; TCGA: The Cancer Genome Atlas; ZNF239: zinc finger protein 239;

Declarations

Authors' contributions

Ye Liu: wrote the manuscript and data analysis. Zhixiang Qin: statistical analysis. Hai Yang: data acquisition. Yang Gu: made important contributions to data analysis and manuscript revision. Kun Li: conceptualization and project administration. All authors contributed substantially to its revision and each author approved the final manuscript.

Acknowledgements

Not applicable

Conflicts of interests

All authors declare that they have no conflict of interest.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The data set of current research analysis can be obtained from the open database.

Funding

Not applicable.

References

1. Villanueva A: **Hepatocellular Carcinoma**. *N Engl J Med* 2019, **380**(15):1450-1462.
2. Llovet JM, Zucman-Rossi J, Pikarsky E, Sangro B, Schwartz M, Sherman M, Gores G: **Hepatocellular carcinoma**. *Nature reviews Disease primers* 2016, **2**:16018.
3. Long J, Bai Y, Yang X, Lin J, Yang X, Wang D, He L, Zheng Y, Zhao H: **Construction and comprehensive analysis of a ceRNA network to reveal potential prognostic biomarkers for hepatocellular carcinoma**. *Cancer cell international* 2019, **19**:90.
4. Sun XY, Yu SZ, Zhang HP, Li J, Guo WZ, Zhang SJ: **A signature of 33 immune-related gene pairs predicts clinical outcome in hepatocellular carcinoma**. *Cancer medicine* 2020, **9**(8):2868-2878.
5. Zheng R, Qu C, Zhang S, Zeng H, Sun K, Gu X, Xia C, Yang Z, Li H, Wei W *et al*: **Liver cancer incidence and mortality in China: Temporal trends and projections to 2030**. *Chinese journal of cancer research = Chung-kuo yen cheng yen chiu* 2018, **30**(6):571-579.
6. Craig AJ, von Felden J, Garcia-Lezana T, Sarcognato S, Villanueva A: **Tumour evolution in hepatocellular carcinoma**. *Nature reviews Gastroenterology & hepatology* 2020, **17**(3):139-152.
7. Hentze MW, Castello A, Schwarzl T, Preiss T: **A brave new world of RNA-binding proteins**. *Nature reviews Molecular cell biology* 2018, **19**(5):327-341.
8. Coppin L, Leclerc J, Vincent A, Porchet N, Pigny P: **Messenger RNA Life-Cycle in Cancer Cells: Emerging Role of Conventional and Non-Conventional RNA-Binding Proteins?***International journal of molecular sciences* 2018, **19**(3).
9. Pereira B, Billaud M, Almeida R: **RNA-Binding Proteins in Cancer: Old Players and New Actors**. *Trends in cancer* 2017, **3**(7):506-528.
10. Kechavarzi B, Janga SC: **Dissecting the expression landscape of RNA-binding proteins in human cancers**. *Genome biology* 2014, **15**(1):R14.

11. Bisogno LS, Keene JD: **RNA regulons in cancer and inflammation**. *Current opinion in genetics & development* 2018, **48**:97-103.
12. Liang Y, Li E, Min J, Gong C, Gao J, Ai J, Liao W, Wu L: **miR-29a suppresses the growth and metastasis of hepatocellular carcinoma through IFITM3**. *Oncology reports* 2018, **40**(6):3261-3272.
13. Dong W, Dai ZH, Liu FC, Guo XG, Ge CM, Ding J, Liu H, Yang F: **The RNA-binding protein RBM3 promotes cell proliferation in hepatocellular carcinoma by regulating circular RNA SCD-circRNA 2 production**. *EBioMedicine* 2019, **45**:155-167.
14. Zheng R, Wan C, Mei S, Qin Q, Wu Q, Sun H, Chen CH, Brown M, Zhang X, Meyer CA *et al*: **Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis**. *Nucleic acids research* 2019, **47**(D1):D729-d735.
15. Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y, Shin H, Wong SS, Ma J, Lei Y *et al*: **Cistrome: an integrative platform for transcriptional regulation studies**. *Genome biology* 2011, **12**(8):R83.
16. Gerstberger S, Hafner M, Tuschl T: **A census of human RNA-binding proteins**. *Nat Rev Genet* 2014, **15**(12):829-845.
17. Robinson MD, Oshlack A: **A scaling normalization method for differential expression analysis of RNA-seq data**. *Genome biology* 2010, **11**(3):R25.
18. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. *Genome research* 2003, **13**(11):2498-2504.
19. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P *et al*: **STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets**. *Nucleic acids research* 2019, **47**(D1):D607-d613.
20. Cheung-Lee WL, Link AJ: **Genome mining for lasso peptides: past, present, and future**. *Journal of industrial microbiology & biotechnology* 2019, **46**(9-10):1371-1379.
21. She Y, Kong X, Ge Y, Yin P, Liu Z, Chen J, Gao F, Fang S: **Immune-related gene signature for predicting the prognosis of head and neck squamous cell carcinoma**. *Cancer cell international* 2020, **20**:22.
22. Li W, Lu J, Ma Z, Zhao J, Liu J: **An Integrated Model Based on a Six-Gene Signature Predicts Overall Survival in Patients With Hepatocellular Carcinoma**. *Frontiers in Genetics* 2020, **10**.
23. Song H, Yu Z, Sun X, Feng J, Yu Q, Khan H, Zhu X, Huang L, Li M, Mok MTS *et al*: **Androgen receptor drives hepatocellular carcinogenesis by activating enhancer of zeste homolog 2-mediated Wnt/β-catenin signaling**. *EBioMedicine* 2018, **35**:155-166.
24. Kido T, Lau YC: **The Y-linked proto-oncogene TSPY contributes to poor prognosis of the male hepatocellular carcinoma patients by promoting the pro-oncogenic and suppressing the anti-oncogenic gene expression**. *Cell & bioscience* 2019, **9**:22.
25. Zhang JF, Chen Y, Lin GS, Zhang JD, Tang WL, Huang JH, Chen JS, Wang XF, Lin ZX: **High IFIT1 expression predicts improved clinical outcome, and IFIT1 along with MGMT more accurately predicts prognosis in newly diagnosed glioblastoma**. *Human pathology* 2016, **52**:136-144.

26. Shao S, Cao H, Wang Z, Zhou D, Wu C, Wang S, Xia D, Zhang D: **CHD4/NuRD complex regulates complement gene expression and correlates with CD8 T cell infiltration in human hepatocellular carcinoma**. *Clinical epigenetics* 2020, **12**(1):31.

27. Chen W, Chen X, Li S, Ren B: **Expression, immune infiltration and clinical significance of SPAG5 in hepatocellular carcinoma: A gene expression-based study**. *The journal of gene medicine* 2020, **22**(4):e3155.

Figures

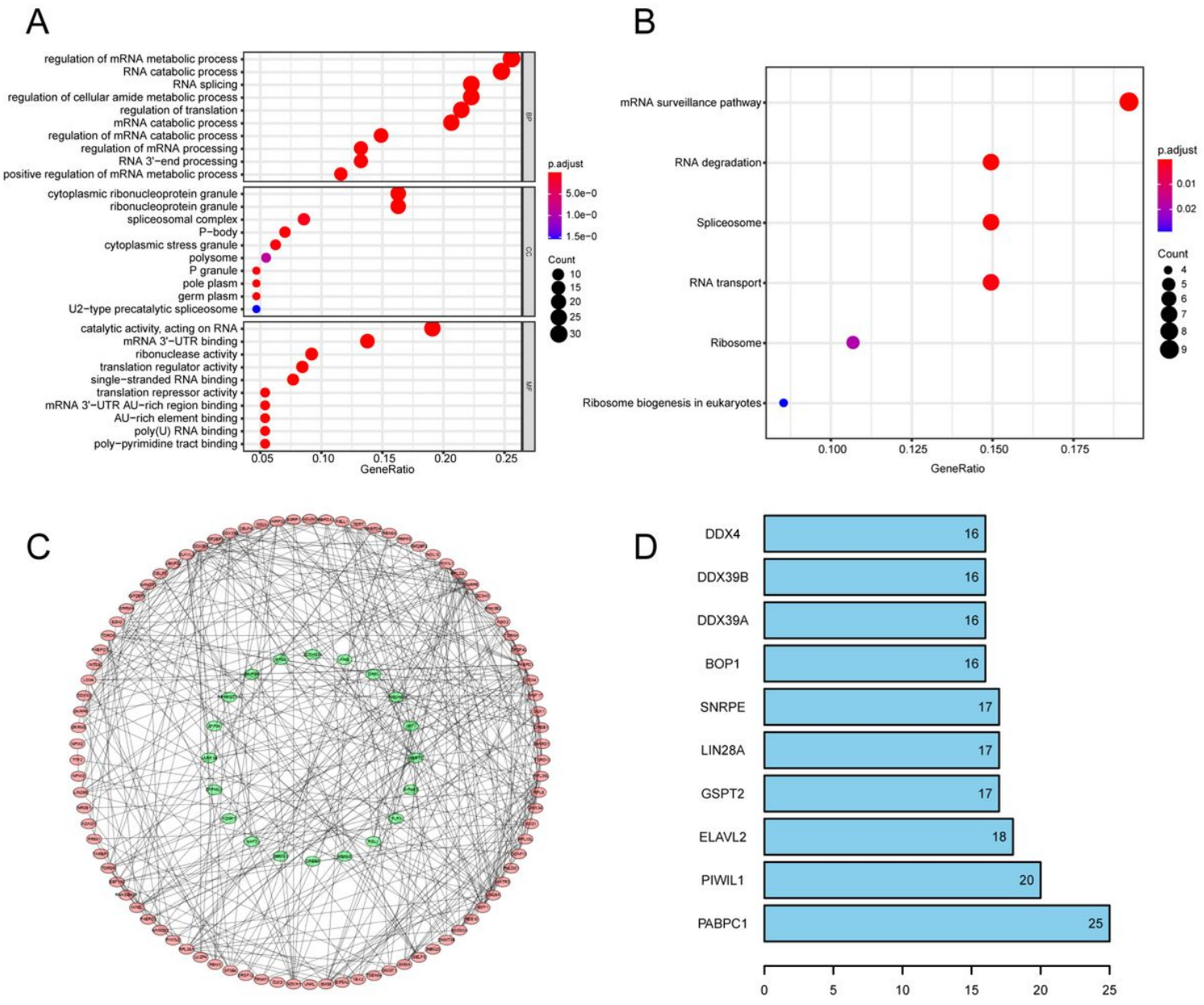

Figure 1

Differentially expressed RBPs for functional enrichment profiling and protein-protein interaction (PPI) network formation. GO and KEGG analyses of differentially expressed RBPs (A, B). PPI network of

differentially expressed RBPs (C). The top ten hub RBPs (D).

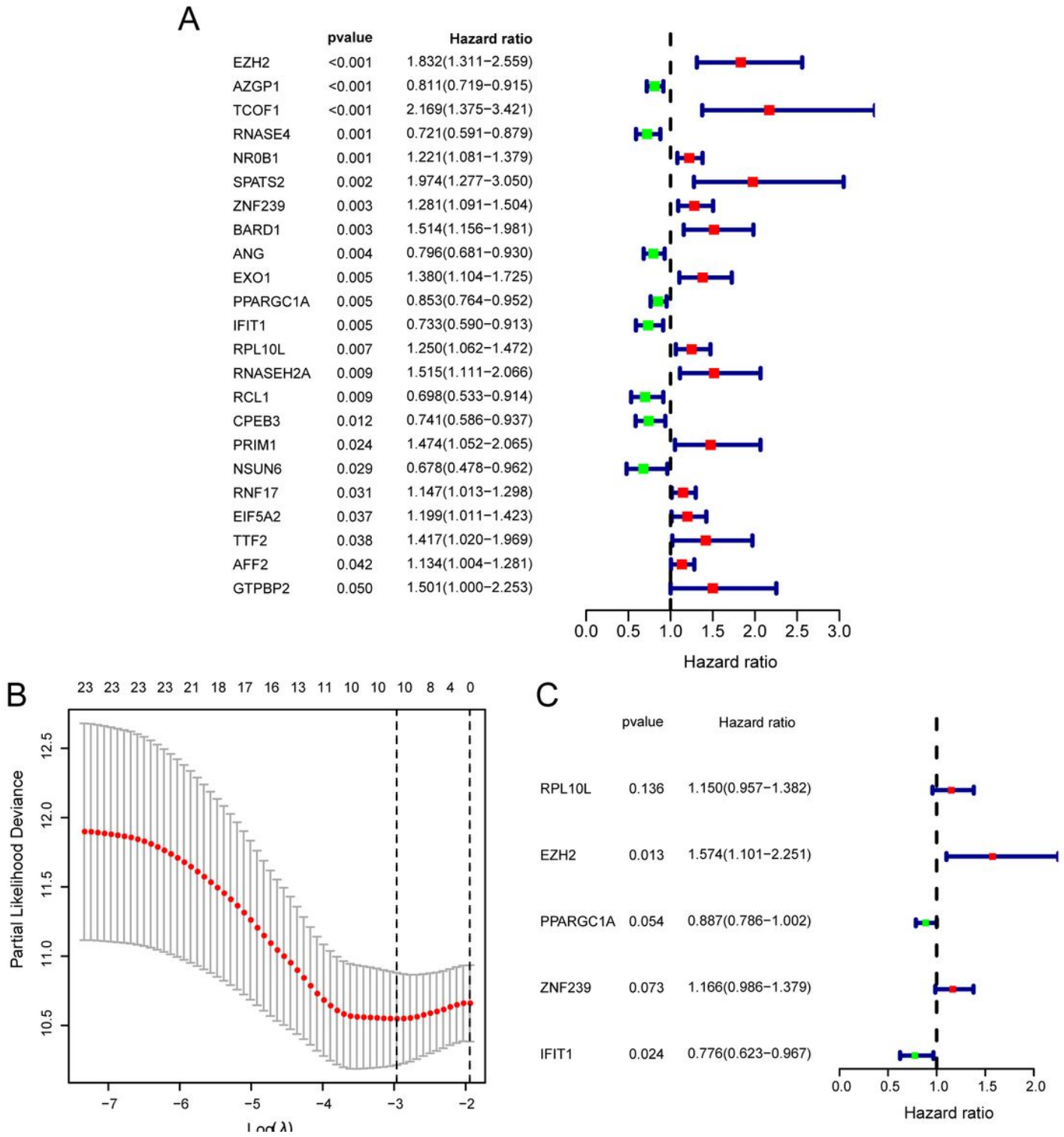

Figure 2

Identification of risk RBPs in the prognostic risk model. Identification of 23 prognostic-associated RBPs through univariate Cox regression analysis (A). Further analysis through Lasso regression analysis (B). Risk gene in the prognostic risk model (C).

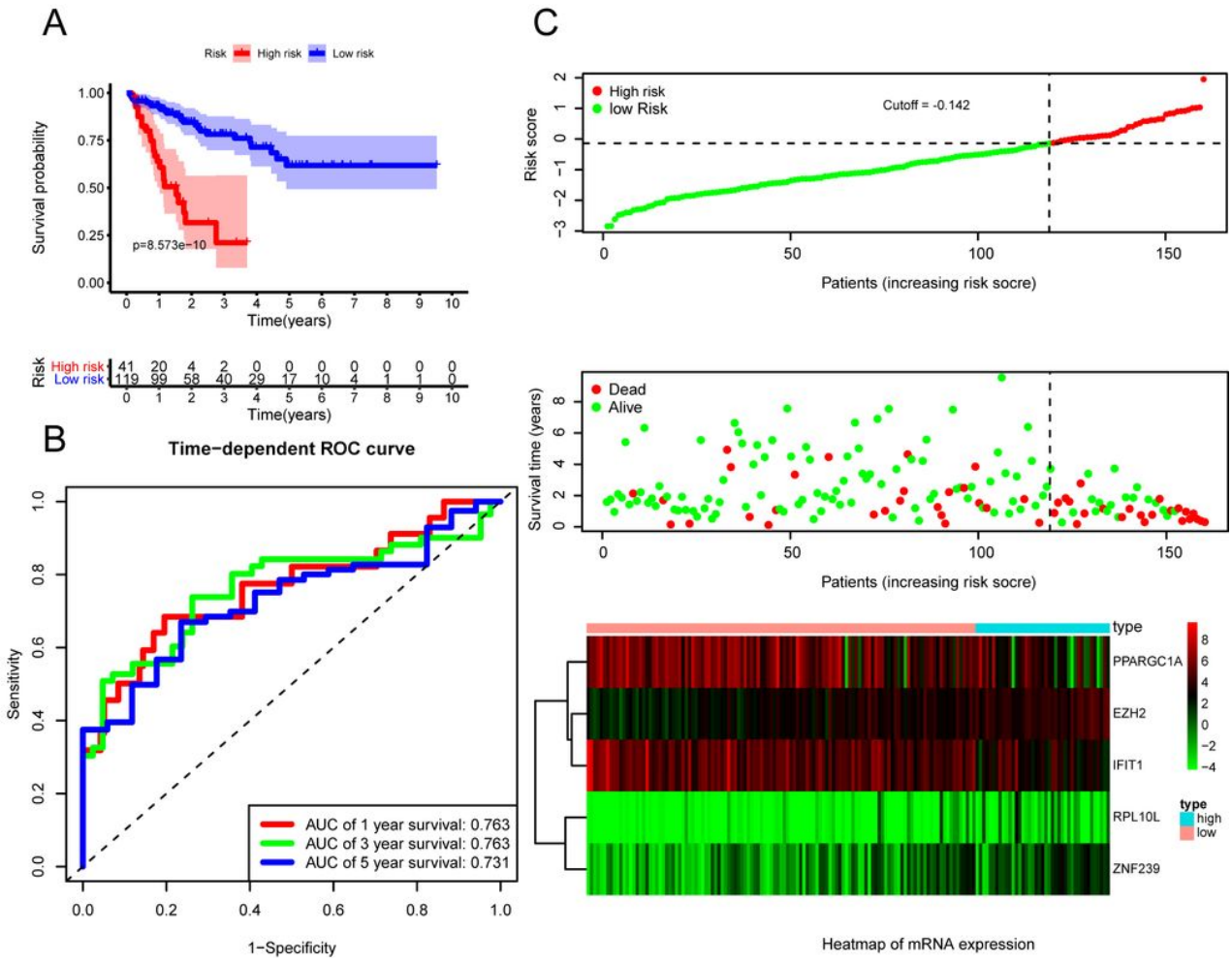

Figure 3

Validation of the risk model. Kaplan–Meier analysis (A), time-dependent ROC analysis (B) and risk score analysis (C) for the prognostic risk model in the training cohort.

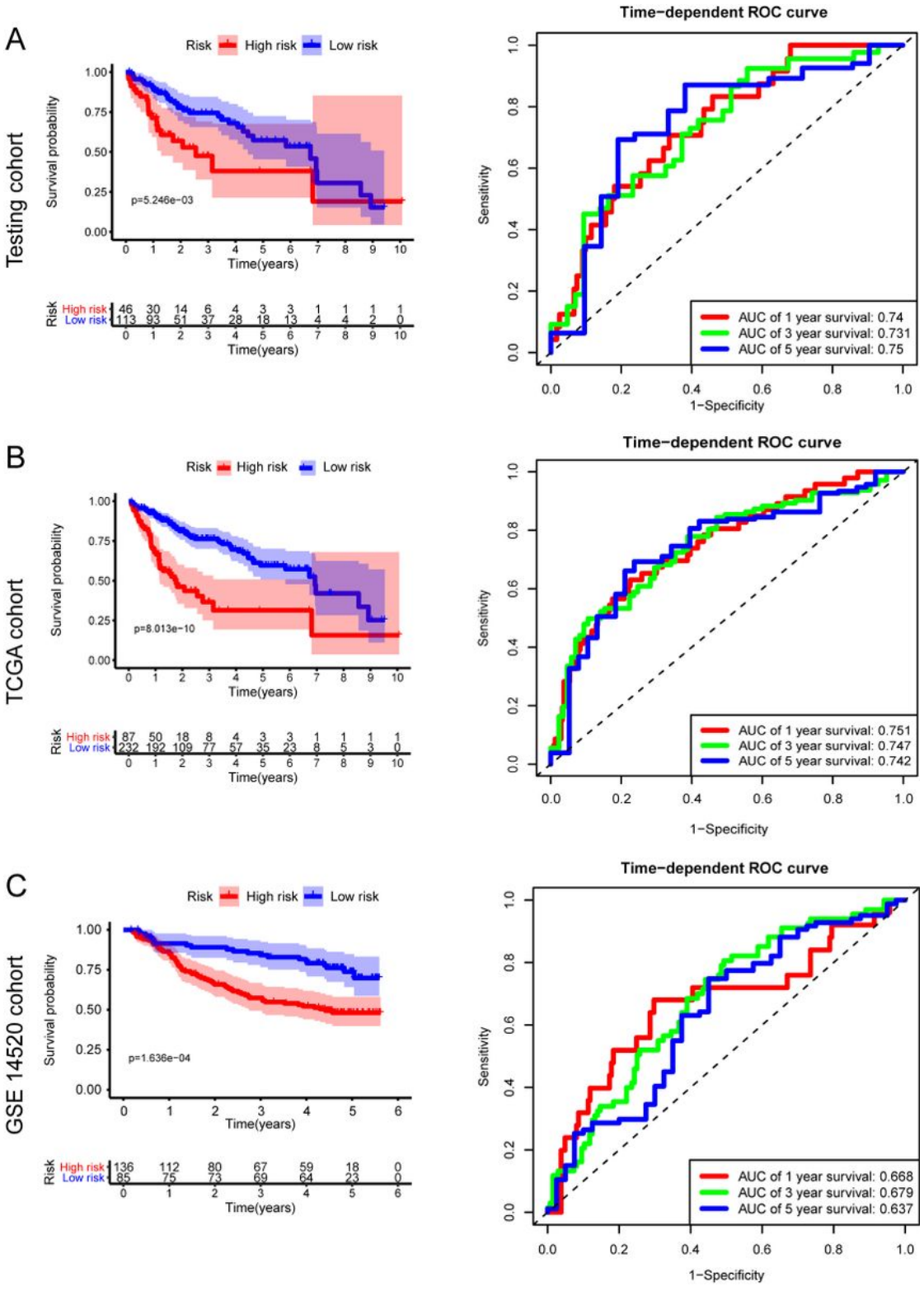

Figure 4

Prognostic analysis of the testing cohort, whole TCGA cohort and GSE14520 cohort. Kaplan-Meier analysis and time-dependent ROC analysis in the testing cohort. (A). Kaplan–Meier analysis and time-dependent ROC analysis in the whole TCGA cohort (B). Kaplan–Meier analysis and time-dependent ROC analysis in the GSE14520 cohort (C).

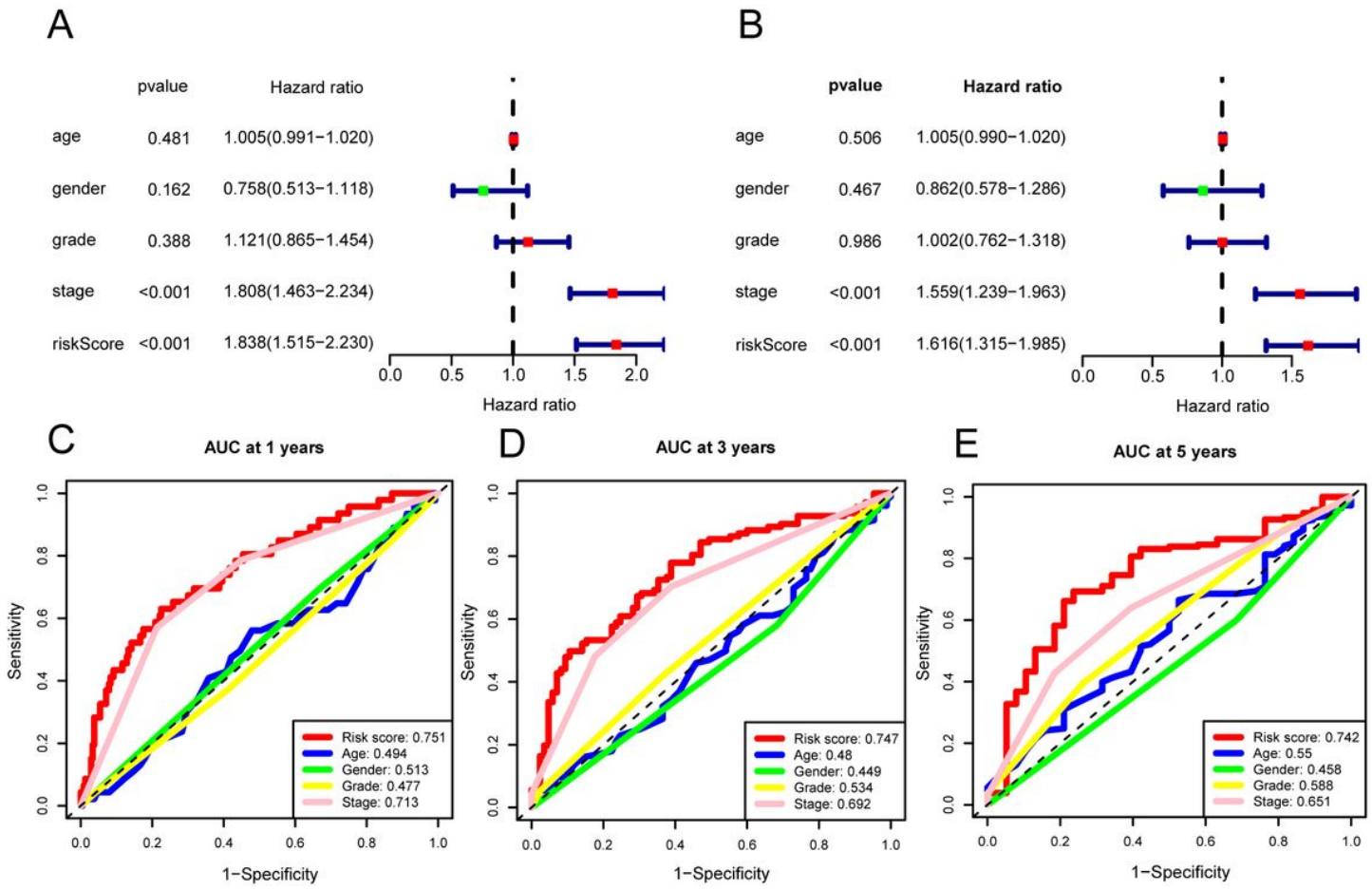

Figure 5

Independent prognostic role of the model of prognostic risk in the whole TCGA cohort. Uni- (A) and multivariate Cox regression assay (B) of the whole TCGA cohort. Time-dependent ROC analysis of different clinical parameters in the whole TCGA cohort at 1, 3 and 5 years (C, D, E).

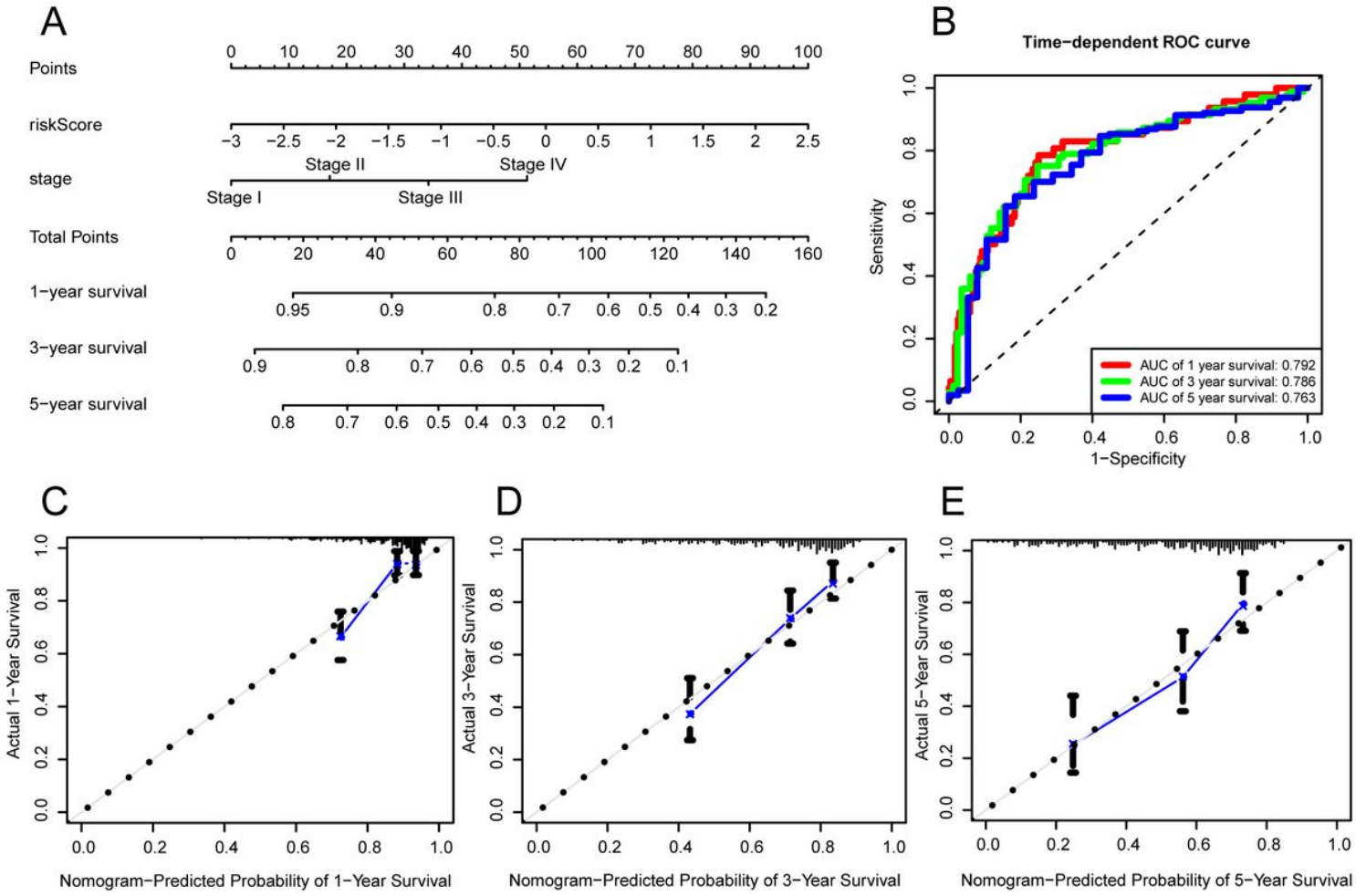

Figure 6

Establishment of a nomogram for predicting overall survival for HCC patients. Nomogram combining risk score with pathological stage (A). Time-dependent ROC analysis of the nomogram (B). The calibration plot for validation of the nomogram (C, D, E).

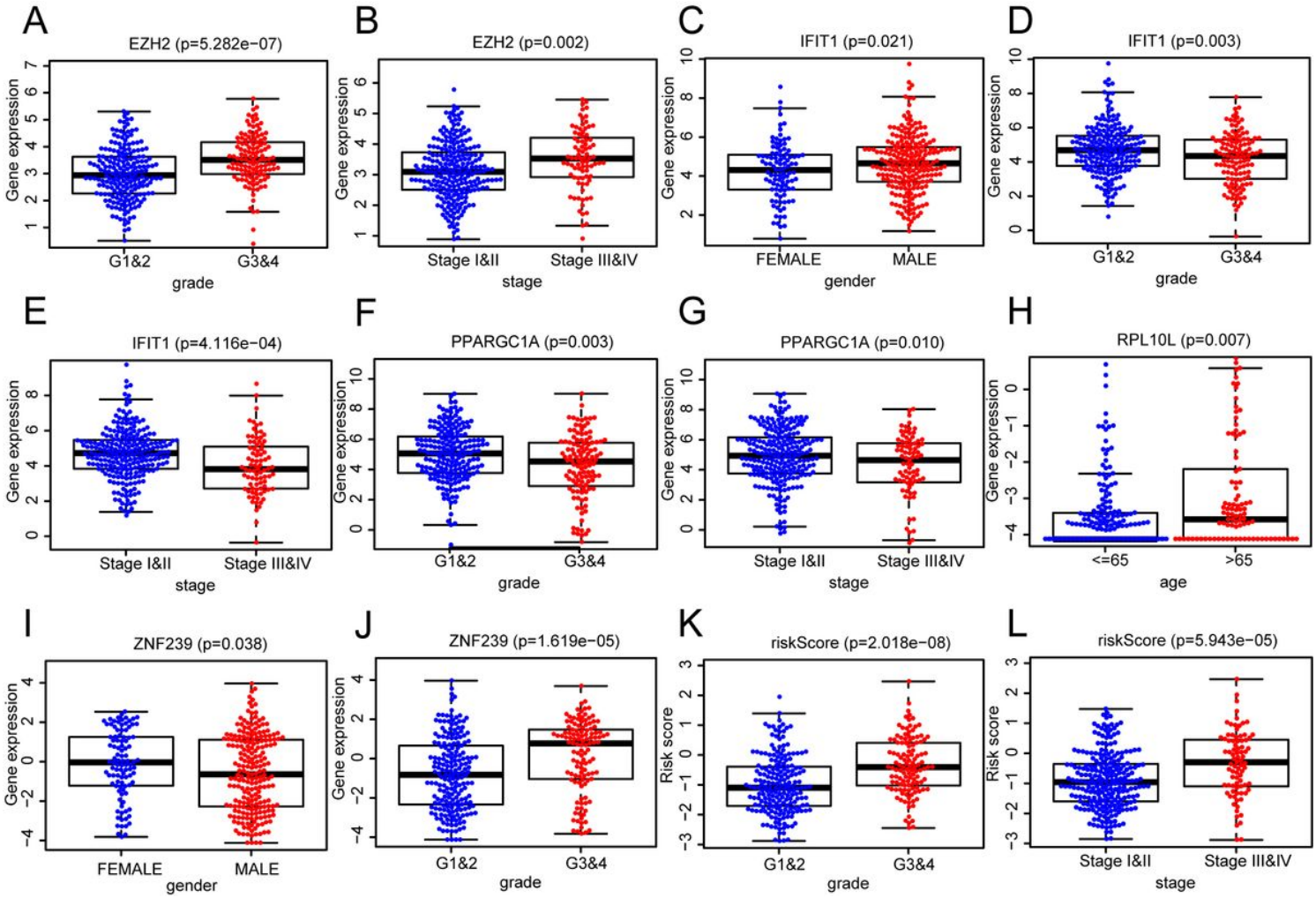

Figure 7

Association of variables in the modelling with clinical attributes of patients in entire TCGA cohort. EZH2 expression and histological grade (A). EZH2 expression and pathological stage (B). IFIT1 expression and gender (C). IFIT1 expression and histological grade (D). IFIT1 expression and pathological stage (E). PPARGC1A expression and histological grade (F). PPARGC1A expression and pathological stage (G). RPL10L expression and age (H). ZNF239 expression and gender (I). ZNF239 expression and histological grade (J). Risk score and histological grade (K). Risk score and pathological stage (L).

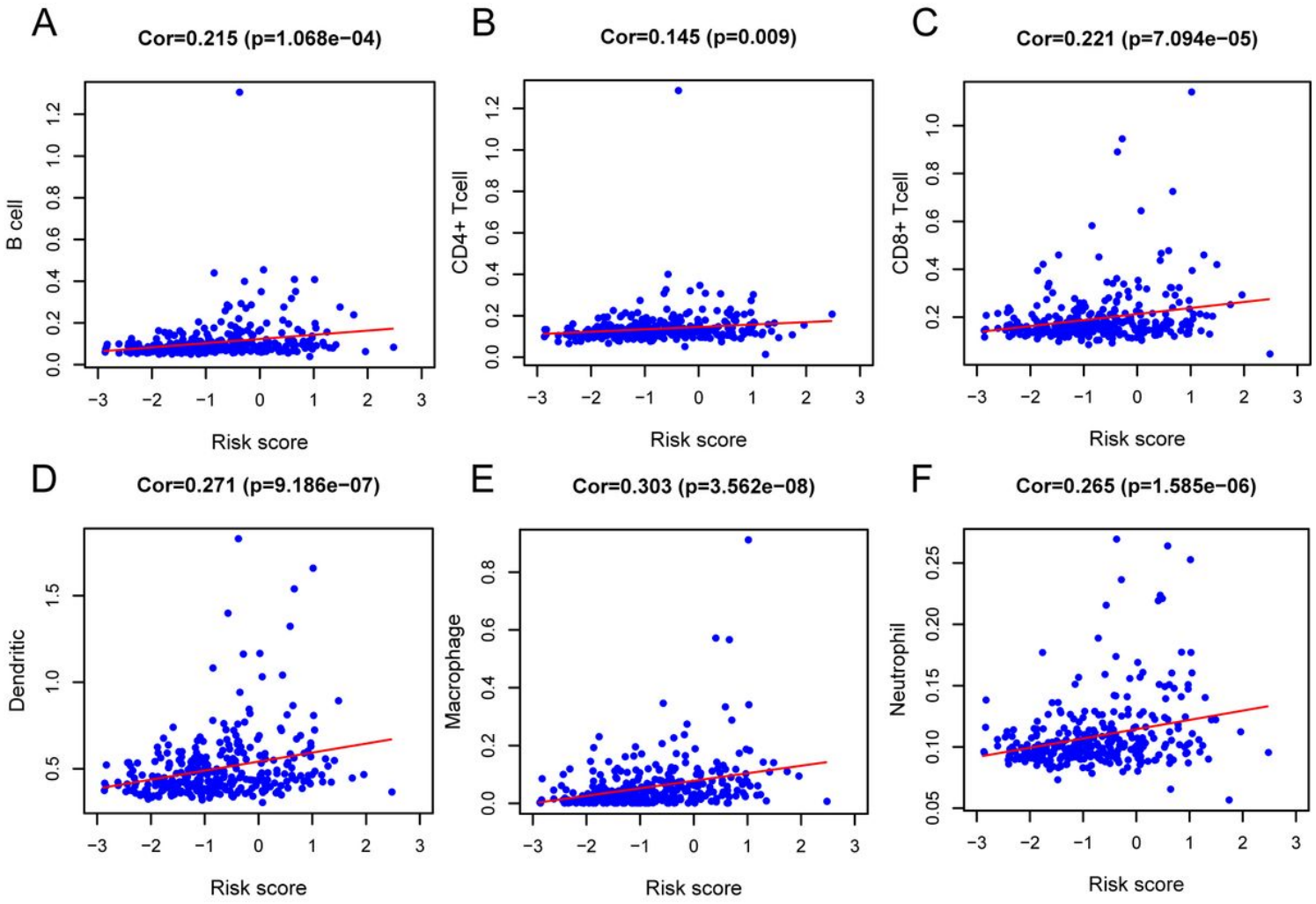

Figure 8

The association on risk score and level of immune infiltration in the entire TCGA cohort. (A). CD4+ T cell (B). CD8+ T cell (C). Dendritic (D). Macrophage (E). Neutrophil (F).

Supplementary Files

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

- Additionalfile1FigureS1.tif
- Additionalfile3TableS1.xlsx
- Additionalfile2FigureS2.tif