Metabolic genes show excellent prognostic ability for clear cell renal cell carcinoma

Chengjian Ji (kkkingdream@163.com)
The First Affiliated Hospital of Nanjing Medical University https://orcid.org/0000-0001-9110-104X

Yichun Wang
The First Affiliated Hospital of Nanjing Medical University

Liangyu Yao
The First Affiliated Hospital of Nanjing Medical University

Jiaochen Luan
The First Affiliated Hospital of Nanjing Medical University

Rong Cong
The First Affiliated Hospital of Nanjing Medical University

Yamin Wang (wangyamin231@163.com)
The First Affiliated Hospital of Nanjing Medical University

Yuan Qin (qinyuan0304@163.com)
The Second Affiliated Hospital of Nanjing University of Chinese Medicine

Primary research

Keywords: MGs; ccRCC; prognosis; metabolic genes

Posted Date: May 6th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-25880/v1

Abstract

Background

Renal cell carcinoma (RCC) is one of the major malignant tumors of the urinary system, with a high mortality rate and a poor prognosis. Clear cell renal cell carcinoma (ccRCC) is the most common subtype of RCC. Although the diagnosis and treatment methods have been significantly improved, the incidence and mortality of ccRCC are high and still increasing. The occurrence and development of ccRCC are closely related to the changes of classic metabolic pathways. This article aims to explore the relationship between metabolic genes and the prognosis of patients with ccRCC.

Patients and methods:

Gene expression profiles of 63 normal kidney tissues and 446 ccRCC tissues from TCGA database and gene expression profiles of 39 ccRCC tissues from GEO database were used to obtain differentially expressed genes (DEGs) in ccRCC. Through the the KEGG gene sets of GSEA database, we obtained metabolic genes (MGs). Univariate Cox regression analysis was used to identify prognostic MGs. Lasso regression analysis was used to eliminate false positives because of over-fitting. Multivariate Cox regression analysis was used to established a prognostic model. Gene expression data and related survival data of 101 ccRCC patients from ArrayExpress database were used for external validation. Survival analysis, ROC curve analysis, independent prognostic analysis and clinical correlation analysis were performed to evaluate this model.

Results

We found that there were 479 abnormally expressed MGs in ccRCC tissues. Through univariate Cox regression analysis, Lasso regression analysis and multivariate Cox regression analysis, we identified 4 prognostic MGs (P4HA3, ETNK2, PAFAH2 and ALAD) and established a prognostic model (*riskScore*). Whether in the training cohort, the testing cohort or the entire cohort, this model could accurately stratify patients with different survival outcomes. The prognostic value of *riskScore* and 4 MGs was also confirmed in the ArrayExpress database. Results of GSEA analysis show that DEGs in patients with better prognosis were enriched in metabolic pathways. Then, a new Nomogram with higher prognostic value was constructed to better predict the 1-year OS, 3-year OS and 5-year OS of ccRCC patients. In addition, we successfully established a ceRNA network to further explain the differences in the expression of these MGs between high-risk patients and low-risk patients

Conclusion

We have successfully established a risk model (*riskScore*) based on 4 MGs, which could accurately predict the prognosis of patients with ccRCC. Our research may shed new light on ccRCC patients' prognosis and treatment management.

Background

In the United States, the death rate of cancer ranks second among all diseases[1]. Renal cell carcinoma (RCC) is a common tumor in the urinary system, ranking among the top ten cancer diagnoses in the world. Clear cell renal cell carcinoma (ccRCC) is the most common histological subtype in kidney cancer, accounting for about 80%. Compared with other subtypes of renal cell carcinoma (including papillary renal cell carcinoma, chromogenic renal cell carcinoma and collecting duct carcinoma), ccRCC showed worse survival outcomes[2]. Due to the lack of obvious clinical symptoms, many patients are already in advanced stages at the time of diagnosis. Therefore, the prognosis of patients becomes more complicated[3].

Glycolysis, fatty acid metabolism, tyrosine metabolism, nucleotide anabolism and other metabolic pathways are necessary to maintain normal cell homeostasis[4, 5]. Many scientists believed that abnormal metabolic processes could be a key factor in the development of cancer[6–8]. In breast cancer and colorectal cancer, studies have reported that abnormal metabolic pathways are associated with poor prognosis[9, 10]. Marin-Rubio et al believed that since abnormally expressed metabolites could be detected in various stages of tumor tissues, metabolic genes (MGs) could be used as prognostic markers for cancer[11]. The prognostic value of metabolic genes has been verified in a variety of cancers including prostate cancer, lung adenocarcinoma, liver cancer, head and neck squamous cell carcinoma, rectal cancer and so on[12–17].

Clear cell renal cell carcinoma has been demonstrated that it is usually accompanied by reprogramming of glucose metabolism, reprogramming of fatty acid metabolism and reprogramming of the tricarboxylic acid cycle. The metabolism of tryptophan, arginine and glutamine is also reprogrammed in many ccRCCs[18]. Therefore, these changes in metabolic pathways provide new possibilities for ccRCC treatment strategies and biomarkers. However, few studies have reported the prognostic role of metabolic genes in ccRCC.

In our study, we used gene expression data and related clinical data of ccRCC patients to explore the relationship between metabolic genes and ccRCC prognosis. Subsequently, we used the metabolic genes to establish a reliable prognostic model of ccRCC, which was verified in The Cancer Genome Atlas (TCGA) database, in GEO database and in ArrayExpress database respectively.

Materials And Methods

Acquisition and preparation of data

Transcriptome profiling data and related clinical information of ccRCC were downloaded from The Cancer Genome Atlas (TCGA) Data Portal (https://tcga-data.nci.nih.gov/tcga/; accessed January 2020), GEO database (https://www.ncbi.nlm.nih.gov./geo/; GSE29609) and ArrayExpress database (https://www.ebi.ac.uk/arrayexpress/; E-MTAB-1980). From the KEGG gene sets of GSEA database (https://www.gsea-msigdb.org/gsea/downloads.jsp; accessed January 2020), we obtained 944 genes related to metabolism. For the obtained transcriptome data, we used R software for data processing. The "sva" package was used to correct the transcriptome data from different databases. The "Limma" package was used for further difference analysis.

Identification of MGs with prognostic value

Firstly, we excluded patients with a follow-up time of less than 90 days. Then, the univariate Cox regression analysis was used to identify genes with prognostic value from differentially expressed MGs. We used the "survival" package in R software for univariate Cox regression analysis.

Establishment of an independent prognostic index (PI, riskScore**) based on MGs in training cohort**

We divided patients in TCGA cohort with complete records of clinical information such as age, gender, pathology stage, histological grade and TMN stage into training cohort and testing cohort. From the perspective of the distribution of patients in various clinical features, the patient composition in the training cohort and the patient composition in the testing cohort were similar.

In order to further identify the key prognostic MGs, the lasso regression and the multivariate Cox regression analysis were performed. According to the weight of each gene in multivariate Cox regression analysis, we obtained the correlation coefficient in the model formula for predicting the prognosis of patients. Combined with the expression of various prognosis-related genes, we established an independent prognostic model. The PI was calculated using the following formula: $\beta_1 \times gene_1$ expression $+ \beta_2 \times gene_2$ expression $+ \cdots\cdots + \beta_n \times gene_n$ expression, where β corresponded to the correlation coefficient.

Validation of key genes at the protein level

We used The Human Protein Atlas (HPA) database (https://www.proteinatlas.org/) to verify the protein level of key MGs in the established prognosis index.

Evaluation of the prognostic index in the training cohort

According to our prognostic model, each patient would get a risk score. We set the median risk score as the cutoff value for dividing ccRCC patients into a high-risk group and a low-risk group. Kaplan–Meier (K-M) method was utilized to plot the survival curves. The log-rank test was performed to assess differences in the survival rates between the high-risk group and the low-risk group. The time-dependent receiver operating characteristic curves (ROC) were created by the "survivalROC" package. The area under the curve (AUC) values were calculated to evaluate the specificity and sensitivity of the model. The risk score distribution of patients, Survival status scatter plots for patients in the prognostic model and the heatmap of prognosis-related MGs were also displayed.

Verification of the prognostic ability of this model in the testing cohort and the entire cohort

First of all, to verify the performance of the prognostic model in the testing cohort, the survival curve was draw with the "survival" package and "survminer" package. A time-dependent ROC curve was also made. Next, the risk score distribution of patients in the prognostic model, Survival status scatter plots for patients in the prognostic model and the heatmap of prognosis-related MGs were also displayed. We then repeated the verification work done on the testing cohort throughout the entire cohort (including TCGA cohort and GEO cohort).

Then, the correlation between the prognostic model (risk genes and *riskScore*) and each clinical feature was analyzed to illustrate the reliability of the model we built.

To further evaluate whether our model could be used as an independent prognostic factor, we included age, gender, pathology stage, histological grade, T, M, N and *riskScore* as independent variables. And then we did univariate Cox regression analysis and multivariate Cox regression analysis on the changes of overall survival time and overall survival outcome.

Verification of the prognostic ability of this model in ArrayExpress cohort

According to our prognostic model, each patient in ArrayExpress cohort will get a risk score. Then, we use the cutoff value from training cohort to divide ccRCC patients into a high-risk group and a low-risk group. Kaplan–Meier (K-M) method was utilized to plot the survival curves. And the log-rank test was performed to assess differences in the survival rates between high-risk group and low-risk group. The time-dependent receiver operating characteristic curves (ROC) were created by the “survivalROC” package. The area under the curve (AUC) values were calculated to evaluate the specificity and sensitivity of the model. Next, K-M survival curve analysis was used to assess the prognostic value of each risk gene.

GSEA enrichment analysis

Gene-set enrichment analysis was used to explore the mechanisms that lead to different outcomes between patients in the high-risk group and patients in the low-risk group.

Nomogram development and validation for prognostic risk prediction

By means of “rms” package of R software, A prognostic nomogram was also performed to visualize the relationship between individual predictors and overall survival rates in patients with ccRCC based on the Cox proportional hazard regression model.

Establishment of a ceRNA network

In order to further explore the possible mechanism for the difference in the expression levels of key risk genes between the two groups, we downloaded microRNA expression data and lncRNA expression data

of ccRCC patients from the TCGA database.

According to the prediction model we established, we divided TCGA-ccRCC patients into high-risk and low-risk groups. Then we used R software to standardize the data and performed a difference analysis with the "Limma" package.

For lncRNAs and microRNAs differentially expressed between high-risk group and low-risk group, we used miRcode (http://www.mircode.org/mircode/) to predict the target microRNAs of differential lncRNAs. Then, Targetscan (http: //www.targe tscan.org/), miRTarBase (http: // mirta rbase.mbc.nctu.edu.tw/) and miRDB (http://www.mirdb.org/) were used to find the target mRNAs of miRNAs. Finally, we screened lncRNA-microRNA pairs that regulate the expression of risk genes.

The lncRNA-miRNA interactions and miRNA-mRNA interactions were integrated together to form the molecular mechanism axis of lncRNA-miRNA-mRNA by using Cytoscape v3.7.1.

Statistical analysis

Statistical analyses of all data utilized in this article were completed by R software (version 3.5.1, https://www.r-project.org/). When the difference met a joint satisfaction of FDR < 0.05 and |log2fold changes (FC)| >1, it was regarded to be statistically significant. Student's t test was used for continuous variables, while categorical variables were compared with the chi-square ($\chi 2$) test. The Wilcoxon rank-sum test was utilized to compare ranked data with two categories. The Kruskal-Wallis test was utilized for comparisons among three or more groups. The univariate Cox regression analysis and multivariate Cox regression analysis were used to evaluate the relationship between MGs expression and survival data to establish a prognostic model. "rms" package of R software was used to create the nomogram. The receiver operating characteristic curves were created by the "survivalROC" package of R software and AUC values were also calculated by this package. If the AUC > 0.60, we would consider this model to have a certain predictive ability. If the AUC > 0.75, we would consider this prediction model to have excellent predictive value. All statistical tests were two-sided and $P < 0.05$ was considered to be statistically significant.

Results

Differentially expressed prognostic MGs

Through the online TCGA database, we obtained the mRNA expression of 446 ccRCC tissues and 63 normal kidney tissues. We combined the RNA sequencing results of 39 other ccRCC patients in the GSE29609 dataset with TCGA cohort.

After correcting and standardizing the data, we obtained the differentially expressed gene profile between the ccRCC group and the normal group through the "Limma" package. Then, we extracted the information on the differential expression of 944 metabolic genes. The filtration conditions ($|\log_2$fold-change$|>1$, FDR < 0.05) were set. Next, 479 differentially expressed MGs were identified, including 196 up-regulated genes

and 283 down-regulated genes (Fig. 1). Based on the obtained differentially expressed MGs, we carried out the univariate Cox regression analysis to identify MGs related to the ccRCC patient's overall survival (OS). Finally, a total of 187 genes were considered to be prognostic MGs.

Construction a prognostic model index based on MGs

Due to the lack of clinical data such as pathology stage and histological grade in the GEO database, we use the patients in the TCGA database to divide them into training cohort and testing cohort. Finally, the patients in the TCGA database and the patients in the GEO database were merged together as the entire cohort.

After excluding patients with a follow-up survival time of less than 90 days (n = 15) and excluding patients with incomplete clinical information (n = 11), we divided the remaining TCGA-ccRCC patients into two groups with similar composition ratios (296 patients for training cohort, 124 patients for testing cohort). Table 1 shows the clinical characteristics of patients included in the study.

Immediately afterwards, the lasso regression and the multivariate Cox regression analysis were performed in the training cohort to target key risk genes (Fig. 2). Finally, we identified 4 optimal risk genes. Among these four risk genes, P4HA3 was considered as predictors of poor prognosis. The higher the expression of P4HA3, the worse the prognosis of patients. Three other genes, ETNK2, PAFAH2 and ALAD, were protective factors. According to the results of multivariate Cox regression analysis, we obtained the risk coefficient of each differentially expressed MGs and then constructed a prognostic model index to predict the prognosis of patients with ccRCC. The 4 prognostic MGs related PI formula was as follows: (P4HA3 expression) * (0.07090771) + (ETNK2 expression) * (-0.0497429) + (PAFAH2 expression) * (-0.1753559) + (ALAD expression) * (-0.0880467).

Evaluation of the prognostic model index based on MGs in training cohort and testing cohort

After obtaining the prognostic model index based on MGs for predicting the prognosis of ccRCC patients, we took a series of measures to evaluate the model.

Firstly, we searched The Human Protein Atlas (HPA) database (https://www.proteinatlas.org/) for risk genes and found that at the protein level, the expression of these risk genes was consistent with the results of the mRNA differential analysis we obtained (Fig. 3). This indicated that the model we built is to some extent credible.

Then, we created Kaplan-Meier curves based on the log-rank test to visualize the prognostic value of our established prognostic model in training cohort and in testing cohort (Fig. 4A, Fig. 5A). From the survival curves, we could see that whether in the training cohort or in the testing cohort, patients in the high-risk group have worse prognosis than those in the low-risk group (HR = 1.1, 95%CI = 1.1–1.2, $p < 0.001$; HR = 1.1, 95%CI = 1.1–1.2, $p < 0.001$). Figure 4B and Fig. 5B respectively show the time-dependent ROC curves of *riskScore* in predicting the prognosis of training cohort patients and the time-dependent ROC curves of

riskScore in predicting the prognosis of testing cohort patients. In the training cohort, the AUC of the prognostic model at 1 year, 3 years and 5 years were 0.709, 0.719 and 0.708 respectively. In the testing cohort, the AUC of the prognostic model at 1 year, 3 years and 5 years were 0.781, 0.769 and 0.703 respectively. Figure 4C and Fig. 5C show the result of risk classification of patients in training cohort and in testing cohort according to *riskScore* respectively. From Fig. 4D and Fig. 5D we found that as the risk score increases, the number of dead patients increases. The expression patterns of the risk genes in the high-risk group and the low-risk group were shown in the form of a heat map (Fig. 4E, Fig. 5E), from which we found that whether in the training cohort or in the testing cohort, P4HA3 was up-regulated in the high-risk group, down-regulated in the low-risk group. The patterns of ETNK2, PAFAH2 and ALAD were the opposite.

Evaluation of the prognostic model index based on MGs in entire cohort

FIGURE 6 shows the preliminary validation results of the performance of the prognostic model in all patients. A Kaplan-Meier curve based on the log-rank test and the ROC curve of multiple prognostic indicators were created to visualize the prognostic value of our established prognostic model in all patients (Fig. 6A-D). It was worth mentioning that the predictive ability of the prognostic model we established was better than the predictive ability of the pathology stage. Figure 6E shows the result of risk classification of patients according to *riskScore*. From Fig. 6F we could found that as the risk score increases, the number of deaths increases. The expression patterns of the risk genes in the high-risk group and the low-risk group were shown in the form of a heat map (Fig. 6G), from which we found that P4HA3 was up-regulated in the high-risk group, down-regulated in the low-risk group. The patterns of ETNK2, PAFAH2 and ALAD were the opposite.

Evaluation of the prognostic model index based on MGs in ArrayExpress cohort

In order to verify whether our model was reliable, we evaluated the prognostic value of risk score in the external cohort from ArrayExpress database (E-MTAB-1980). The external cohort contained 101 patients with ccRCC. Similarly, we calculated the risk score of each patient based on *riskScore*. Then we divided the patients into a high-risk group and a low-risk group according to the cutoff value we obtained in the training cohort. A Kaplan-Meier curve based on the log-rank test and the time-depedent ROC curve were created to visualize the prognostic value of our established prognostic model in external cohort (Fig. 7A **and** Fig. 7B). The areas under the ROC (AUC) values of PI were 0.763 for 1-Year-OS, 0.808 for 3-Year-OS and 0.752 for 5-Year-OS. In addition, we further investigated whether each risk gene is related to the prognosis of ccRCC. Figure 7C-F show that risk genes are significantly related to prognosis. Among them, the higher the expression of P4HA3, the worse the prognosis of patients. ETNK2, PAFAH2 and ALAD all show the role of protective prognostic factors, which is consistent with the conclusions we have obtained before.

Clinical correlation analysis

Based on the information of all ccRCC patients from TCGA database we obtained, the correlation analysis between risk factors in the prognostic model (*riskScore* and each component gene) and clinical characteristics such as age, gender, pathology stage, histological grade, TMN was performed. Figure 8 showed the results. We found that there was a significant correlation between the *riskScore* and gender, pathology stage, histological grade, T, M (Fig. 8A).

Independent prognostic factor evaluation and GSEA enrichment analysis

To further evaluate whether our model could be used as an independent prognostic factor, we included some key clinical characteristics containing age, gender, pathology stage, histological grade, TMN and r*iskScore* as independent variables. By means of univariate and multivariate Cox regression analysis, our established PI (*riskScore*) remained significant (both $P < 0.001$, Table 2). At the same time, the results of multivariate Cox regression analysis also showed that histological grade and M could be used as independent prognostic indicators (both $P < 0.05$). Looking only at the results of multivariate Cox regression analysis, we could find that *riskScore*, histological grade and M were related to prognosis ($p < 0.001$, $p = 0.041$, $p = 0.014$ respectively).

In addition, in order to further explore the possible mechanisms that caused different outcomes in the high-risk group and the low-risk group, we performed Gene-set enrichment analysis (GSEA) on the gene expression profiles of the two groups of patients. Figure 9A plots enriched pathways in the high-risk group, while Fig. 9B plots enriched pathways in the low-risk group. The results of GSEA suggested that most of the differentially expressed genes in the low-risk group were genes related to metabolic pathways.

Nomogram development and validation

Finally, to better predict the 1-year OS, 3-year OS and 5-year OS of ccRCC patients, we constructed a new Nomogram based on the results of the multivariate Cox regression analysis of independent prognostic factors (Fig. 10A). Figure 10B-D show the Calibration curves of the nomogram for the probability of OS at 1, 3 and 5 year. The C-index of the nomogram for OS prediction was 0.763 (95%CI = 0.701–0.825), while the C-index of *riskScore* for OS prediction was 0.722(95%CI = 0.659–0.785).

Establishment of a ceRNA network regulating risk genes

For the lncRNA expression profile and microRNA expression profile of ccRCC patients obtained from the TCGA database, we used the “Limma” package to analyze the differentially expressed lncRNAs and microRNAs between patients in the high-risk group and patients in the low-risk group. We obtained 535 differentially expressed lncRNAs (425 up-regulated and 110 downregulated) and 176 differentially expressed microRNAs (114 up-regulated and 62 down-regulated).

Then, we used miRcode online software to predict possible matching pairs between differential lncRNAs and differential microRNAs. Finally, 26 of 176 differentially expressed miRNAs and 340 of 535

differentially expressed lncRNAs were successfully matched, and there was a correlation between them. According to the 26 target miRNAs and 4 risk genes, 23 microRNA–mRNA interactions were found in the three different databases of Targetscan, miRTarBase and miRDB, 159 lncRNA-microRNA interactions were also identified. In general, a lncRNA-microRNA-mRNA ceRNA network was established to further explore the possible mechanism for the difference in the expression levels of key risk genes between the two groups (Fig. 11).

Discussion

In recent years, the incidence of asymptomatic ccRCC has been rising. Patients are prone to poor prognosis and high mortality, which brings new challenges to early clinical detection and treatment. Studies have shown that the effects of chemotherapy and radiotherapy on ccRCC are not ideal. Although surgical resection can suppress the metastasis and development of the tumor to a certain extent, the postoperative recurrence rate is persistent and it is difficult to achieve the desired therapeutic effect[19, 20]. Studies have shown that after surgical removal of cancer tissue, many ccRCC patients still have poor prognosis such as cancer metastasis and death. The 5-year survival rate is low, indicating that the treatment effect of ccRCC patients is limited[21]. Therefore, biological indicators of disease occurrence and development have become the focus of research in recent years[22–24].

Changes in metabolic pathways exist in many diseases, including tumors. Metabolic reprogramming is very important to maintain abnormal proliferation of tumor cells[25]. More and more cancer researchers are focusing their attention on the mechanism of tumor metabolic changes. A recent study pointed out that the heterogeneity of metabolic reprogramming in the tumor is clearly related to the outcome of the tumor[26]. New cancer treatments targeting metabolic pathways are being studied in many tumors. A global regulator of glucose metabolism (MUC1), which was found to be a target for pancreatic cancer treatment recently. After MUC1 was knocked out in pancreatic cancer cells, Fu et al found that the metabolic activity of the cancer cells was significantly reduced, the cancer cells were better able to receive chemotherapy[27]. Similarly, in soft tissue sarcoma, targeting glutamine metabolism can well inhibit tumor growth[28]. Drugs that inhibit glucose metabolism and drugs that inhibit lipid metabolism can be used as a new treatment for lung cancer[29]. Recently, some researchers have suggested that ccRCC is a metabolic disease[30]. They believe that some metabolic pathways in ccRCC can be used as potential biomarkers for therapeutic targets, diagnosis and prognosis. Lucarelli et al pointed out that the changes of metabolic activities in cancer are related to the changes of gene expression. They believed that the changes of metabolic genes are closely related to the occurrence and development of tumors. They further determined that NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 4-like 2 (NDUFA4L2) plays an important role in the occurrence and development of ccRCC[31]. However, few studies have studied the relationship between metabolic genes and the prognosis of ccRCC patients.

In our study, we analyzed and screened out new metabolic genes with prognostic value from the TCGA database and the GEO database. We further verified the feasibility of the method through the data downloaded from the ArrayExpress database. Finally, we have successfully established a risk model

(*riskScore*) based on 4 MGs, which could accurately predict the prognosis of patients with ccRCC. The scientific model construction methods, comprehensive evaluation methods of prognosis ability and verification of multiple databases make the prognostic model we have established with high credibility. According to the prognostic model we established, patients with different survival outcomes were accurately classified into high-risk patients and low-risk patients. Which also indicates that metabolic genes have a good prognostic value in ccRCC. In addition, we have further conducted GSEA analysis to explore the possible mechanisms leading to different survival outcomes. The results of GSEA analysis show that many metabolic pathways changed in low-risk patients, which further suggests that metabolism plays a very important role in ccRCC.

P4HA3, ETNK2, PAFAH2 and ALAD are the four risk metabolic genes we identified. Among these four risk genes, P4HA3 was considered as predictors of poor prognosis. The higher the expression of P4HA3, the worse the prognosis of patients. Three other genes, ETNK2, PAFAH2 and ALAD, were protective factors. P4HA3, encoding a component of prolyl 4-hydroxylase, can promote the proliferation and invasion of melanoma cells[32]. In addition, P4HA3 has been shown to be closely related to the poor survival outcomes of breast cancer and gastric cancer[33, 34]. ETNK2 (Ethanolamine Kinase 2) is a protein Coding gene. Among its related pathways are phosphatidylethanolamine biosynthesis II and Glycerophospholipid biosynthesis. The decreased expression of ETNK2 is related to the progression of prostate cancer[35]. PAFAH2 is closely related to ether lipid metabolism. Kono et al revealed that PAFAH2 is involved in the metabolism of esterified 8-isoprostaglandin F (2alpha) and protects tissue from oxidative stress-induced injury[36]. However, no one has pointed out its relationship with the survival of cancer patients. Our analysis of patients in the TCGA database and ArrayExpress database showed that PAFAH2 could well distinguish the survival outcome of ccRCC patients and could be a protective prognostic factor. ALAD Catalyzes an early step in the biosynthesis of tetrapyrroles. ALAD has been found to be related to the favorable survival outcome of breast cancer patients. Overexpression of ALAD can inhibit the proliferation and invasion of breast cancer cells[37]. In summary, the four risk genes we identified show a relatively consistent prognostic value in the TCGA database and ArrayExpress database, and have been verified to a certain extent in the IHC data of the HPA database and in other tumors. All these suggest that the prognostic model we established has a high degree of credibility. In addition, we successfully established a lncRNA-microRNA-mRNA ceRNA network to further explain the differences in the expression of these genes between high-risk patients and low-risk patients.

Due to the lack of obvious clinical manifestations of early ccRCC, many ccRCC patients are already in the advanced stage at the time of diagnosis. At the same time, ccRCC has a worse prognosis than other types of renal cell carcinoma. Our results suggest that changes in the expression of some metabolic genes are closely related to the prognosis of patients with ccRCC. Considering that changes in metabolic pathways and metabolic genes play an important role in the occurrence, development and prognosis of ccRCC, we wondered whether some metabolic biomarkers can be used to improve the early diagnosis rate of ccRCC, which requires further research to explore. In addition, our research also sheds new light on the treatment of ccRCC.

However, our study still has some limitations: The results of our study were only validated in the TCGA database, GEO database and ArrayExpress database. Whether it had real clinical application value, the model requires more data support from clinical patients. In addition, the mechanism by which metabolic genes affect the prognosis of patients with ccRCC needs to be further explored through in vivo and in vitro experiments.

Conclusions

All in all, we have successfully established a risk model (*riskScore*) based on 4 MGs, which could accurately predict the prognosis of patients with ccRCC. A nomogram combining clinical variables and *riskScore* was also drawn to improve the accuracy of the prediction. Our research may shed new light on ccRCC patients' prognosis and treatment management. However, further experiments are also required to validate our findings.

Abbreviations

RCC: Renal cell carcinoma; ccRCC: Clear cell renal cell carcinoma; TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus; HPA: The Human Protein Atlas; DEGs: Differentially expressed genes; MGs: Metabolic genes; PI: Prognostic index; KEGG: Kyoto Encyclopedia of Genes and Genomes; ROC: Receiver operating characteristic; AUC: Area under the curve; OS: Overall survival; K-M: Kaplan–Meier; FC: Fold change; LASSO: Least absolute shrinkage and selection operator; IHC: Immunohistochemistry, HR: Hazard ratio; CI: Confidence interval.

Declarations

Acknowledgements

We would like to thank the researchers and study participants for their contributions.

Author's Contribution

Y.Q,YM.W: Protocol/project development;

LY.Y: Data collection or management;

CJ.J, YC.W: Data analysis;

CJ.J, YC.W: Manuscript writing/editing.

Funding

None declared.

Availability of data and material

All the data used to support the findings of this study are included within the article. Please contact author for data requests.

Competing interests

None declared.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. Cancer J Clin. 2019;69(1):7–34.
2. Hakimi AA, Voss MH, Kuo F, Sanchez A, Liu M, Nixon BG, Vuong L, Ostrovnaya I, Chen YB, Reuter V, et al. Transcriptomic Profiling of the Tumor Microenvironment Reveals Distinct Subgroups of Clear Cell Renal Cell Cancer: Data from a Randomized Phase III Trial. Cancer discovery. 2019;9(4):510–25.
3. Shingarev R, Jaimes EA. Renal cell carcinoma: new insights and challenges for a clinician scientist. American journal of physiology Renal physiology. 2017;313(2):F145-f154.
4. Kreuzaler P, Panina Y, Segal J, Yuneva M. Adapt and conquer: Metabolic flexibility in cancer growth, invasion and evasion. Molecular metabolism. 2020;33:83–101.
5. Woolbright BL, Rajendran G, Harris RA, Taylor JA. 3rd: **Metabolic Flexibility in Cancer: Targeting the Pyruvate Dehydrogenase Kinase:Pyruvate Dehydrogenase Axis**. Mol Cancer Ther. 2019;18(10):1673–81.
6. El Hassouni B, Granchi C, Valles-Marti A, Supadmanaba IGP, Bononi G, Tuccinardi T, Funel N, Jimenez CR, Peters GJ, Giovannetti E, et al. The dichotomous role of the glycolytic metabolism pathway in cancer metastasis: Interplay with the complex tumor microenvironment and novel therapeutic strategies. Sem Cancer Biol. 2020;60:238–48.
7. da Costa IA, Hennenlotter J, Stuhler V, Kuhs U, Scharpf M, Todenhofer T, Stenzl A, Bedke J: **Transketolase like 1** (**TKTL1**) **expression alterations in prostate cancer tumorigenesis**. *Urologic oncology* 2018, **36**(10):472.e421-472.e427.
8. Meng Y, Xu X, Luan H, Li L, Dai W, Li Z, Bian J. The progress and development of GLUT1 inhibitors targeting cancer energy metabolism. Future medicinal chemistry. 2019;11(17):2333–52.
9. Qiu C, Zhang Y, Chen L. Impaired Metabolic Pathways Related to Colorectal Cancer Progression and Therapeutic Implications. Iranian journal of public health. 2020;49(1):56–67.
10. Trilla-Fuertes L, Gamez-Pozo A, Lopez-Camacho E, Prado-Vazquez G, Zapater-Moros A, Lopez-Vacas R, Arevalillo JM, Diaz-Almiron M, Navarro H, Main P, et al. Computational models applied to

metabolomics data hints at the relevance of glutamine metabolism in breast cancer. BMC Cancer. 2020;20(1):307.

11. Marin-Rubio JL, Vela-Martin L, Fernandez-Piqueras J, Villa-Morales M. **FADD in Cancer: Mechanisms of Altered Expression and Function, and Clinical Implications**. *Cancers* 2019, 11(10).
12. Roberto D, Selvarajah S, Park PC, Berman D, Venkateswaran V. Functional validation of metabolic genes that distinguish Gleason 3 from Gleason 4 prostate cancer foci. Prostate. 2019;79(15):1777–88.
13. Zhang S, Lu Y, Liu Z, Li X, Wang Z, Cai Z: **Identification Six Metabolic Genes as Potential Biomarkers for Lung Adenocarcinoma**. *Journal of computational biology: a journal of computational molecular cell biology* 2020.
14. Liu GM, Xie WX, Zhang CY, Xu JW. Identification of a four-gene metabolic signature predicting overall survival for hepatocellular carcinoma. Journal of cellular physiology. 2020;235(2):1624–36.
15. Ye F, Jia D, Lu M, Levine H, Deem MW. Modularity of the metabolic gene network as a prognostic biomarker for hepatocellular carcinoma. Oncotarget. 2018;9(19):15015–26.
16. Xing L, Guo M, Zhang X, Zhang X, Liu F. A transcriptional metabolic gene-set based prognostic signature is associated with clinical and mutational features in head and neck squamous cell carcinoma. J Cancer Res Clin Oncol. 2020;146(3):621–30.
17. Zhang ZY, Yao QZ, Liu HY, Guo QN, Qiu PJ, Chen JP, Lin JQ. **Metabolic reprogramming-associated genes predict overall survival for rectal cancer**. *Journal of cellular and molecular medicine* 2020.
18. Wettersten HI, Aboud OA, Lara PN Jr, Weiss RH. Metabolic reprogramming in clear cell renal cell carcinoma. Nature reviews Nephrology. 2017;13(7):410–9.
19. Keegan KA, Schupp CW, Chamie K, Hellenthal NJ, Evans CP, Koppie TM. Histopathology of surgically treated renal cell carcinoma: survival differences by subtype and stage. The Journal of urology. 2012;188(2):391–7.
20. Lee BH, Feifer A, Feuerstein MA, Benfante NE, Kou L, Yu C, Kattan MW, Russo P. Validation of a Postoperative Nomogram Predicting Recurrence in Patients with Conventional Clear Cell Renal Cell Carcinoma. European urology focus. 2018;4(1):100–5.
21. Gershman B, Thompson RH, Boorjian SA, Lohse CM, Costello BA, Cheville JC, Leibovich BC. Radical Versus Partial Nephrectomy for cT1 Renal Cell Carcinoma. European urology. 2018;74(6):825–32.
22. Wan B, Liu B, Huang Y, Yu G, Lv C. Prognostic value of immune-related genes in clear cell renal cell carcinoma. Aging. 2019;11(23):11474–89.
23. Zhou J, Wang J, Hong B, Ma K, Xie H, Li L, Zhang K, Zhou B, Cai L, Gong K. Gene signatures and prognostic values of m6A regulators in clear cell renal cell carcinoma - a retrospective study using TCGA database. Aging. 2019;11(6):1633–47.
24. Wan B, Liu B, Yu G, Huang Y, Lv C. Differentially expressed autophagy-related genes are potential prognostic and diagnostic biomarkers in clear-cell renal cell carcinoma. Aging. 2019;11(20):9025–42.

25. Hoxhaj G, Manning BD. The PI3K-AKT network at the interface of oncogenic signalling and cancer metabolism. Nature reviews Cancer. 2020;20(2):74–88.
26. Faubert B, Solmonson A, DeBerardinis RJ. **Metabolic reprogramming and cancer progression**. *Science (New York, NY)* 2020, 368(6487).
27. Fu X, Tang N, Xie WQ, Mao L, Qiu YD. MUC1 promotes glycolysis through inhibiting BRCA1 expression in pancreatic cancer. Chinese journal of natural medicines. 2020;18(3):178–85.
28. Lee P, Malik D, Perkons N, Huangyang P, Khare S, Rhoades S, Gong YY, Burrows M, Finan JM, Nissim I, et al. Targeting glutamine metabolism slows soft tissue sarcoma growth. Nature communications. 2020;11(1):498.
29. Wu L, Wang W, Dai M, Li H, Chen C, Wang D. PPARalpha ligand, AVE8134, and cyclooxygenase inhibitor therapy synergistically suppress lung cancer growth and metastasis. BMC Cancer. 2019;19(1):1166.
30. Lucarelli G, Loizzo D, Franzin R, Battaglia S, Ferro M, Cantiello F, Castellano G, Bettocchi C, Ditonno P, Battaglia M. Metabolomic insights into pathophysiological mechanisms and biomarker discovery in clear cell renal cell carcinoma. Expert review of molecular diagnostics. 2019;19(5):397–407.
31. Lucarelli G, Rutigliano M, Sallustio F, Ribatti D, Giglio A, Lepore Signorile M, Grossi V, Sanese P, Napoli A, Maiorano E, et al. Integrated multi-omics characterization reveals a distinctive metabolic signature and the role of NDUFA4L2 in promoting angiogenesis, chemoresistance, and mitochondrial dysfunction in clear cell renal cell carcinoma. Aging. 2018;10(12):3957–85.
32. Atkinson A, Renziehausen A, Wang H, Lo Nigro C, Lattanzio L, Merlano M, Rao B, Weir L, Evans A, Matin R, et al. Collagen Prolyl Hydroxylases Are Bifunctional Growth Regulators in Melanoma. J Invest Dermatol. 2019;139(5):1118–26.
33. Winslow S, Lindquist KE, Edsjo A, Larsson C. The expression pattern of matrix-producing tumor stroma is of prognostic importance in breast cancer. BMC Cancer. 2016;16(1):841.
34. Song H, Liu L, Song Z, Ren Y, Li C, Huo J. P4HA3 is Epigenetically Activated by Slug in Gastric Cancer and its Deregulation is Associated With Enhanced Metastasis and Poor Survival. Technology in cancer research treatment. 2018;17:1533033818796485.
35. Kamdar S, Isserlin R, Van der Kwast T, Zlotta AR, Bader GD, Fleshner NE, Bapat B. Exploring targets of TET2-mediated methylation reprogramming as potential discriminators of prostate cancer progression. Clinical epigenetics. 2019;11(1):54.
36. Kono N, Inoue T, Yoshida Y, Sato H, Matsusue T, Itabe H, Niki E, Aoki J, Arai H. Protection against oxidative stress-induced hepatic injury by intracellular type II platelet-activating factor acetylhydrolase by metabolism of oxidized phospholipids in vivo. J Biol Chem. 2008;283(3):1628–36.
37. Ge J, Yu Y, Xin F, Yang ZJ, Zhao HM, Wang X, Tong ZS, Cao XC. Downregulation of delta-aminolevulinate dehydratase is associated with poor prognosis in patients with breast cancer. Cancer Sci. 2017;108(4):604–11.

Tables

Table 1: Clinical information of included patients.

Clinical parameters	Variables	TCGA cohort	GEO cohort	ArrayExpress cohort
Survival status	Dead	145	16	23
	Alive	275	23	78
Age(years)	<65	262	20	52
	>65 or =65	158	19	49
Gender	Male	277	NR	77
	Female	143	NR	24
	NR	0	39	0
Pathology stage	I	215	NR	NR
	II	44	NR	NR
	III	95	NR	NR
	IV	66	NR	NR
	NR	0	39	101
Histological grade	G1	7	NR	13
	G2	185	NR	59
	G3	164	NR	22
	G4	57	NR	5
	GX	5	NR	0
	NR	2	39	2

Abbreviations: NR, not recorded.

Table 2. Univariate and multivariate cox regression analyses of OS in the TCGA cohort.

Variables	Univariate Cox		Multivariate Cox	
	Hazard ratio (95% CI)	pvalue	Hazard ratio (95% CI)	pvalue
Age	1.014 (0.996−1.032)	0.132	1.027(1.005−1.049)	0.015
Gender	1.168(0.728−1.876)	0.519	1.667(0.978−2.842)	0.060
Histological grade	2.336(1.692−3.226)	<0.001	1.497(1.017−2.204)	0.041
Pathological stage	1.831(1.488−2.254)	<0.001	1.029(0.520−2.039)	0.934
T	1.952(1.508−2.527)	<0.001	1.164(0.595−2.278)	0.658
M	4.382(2.670−7.192)	<0.001	3.811(1.314−11.054)	0.014
N	2.701(1.235−5.908)	0.013	1.503(0.373−2.972)	0.922
Riskscore	1.134(1.085−1.184)	<0.001	1.114(1.053−1.179)	<0.001

Figures

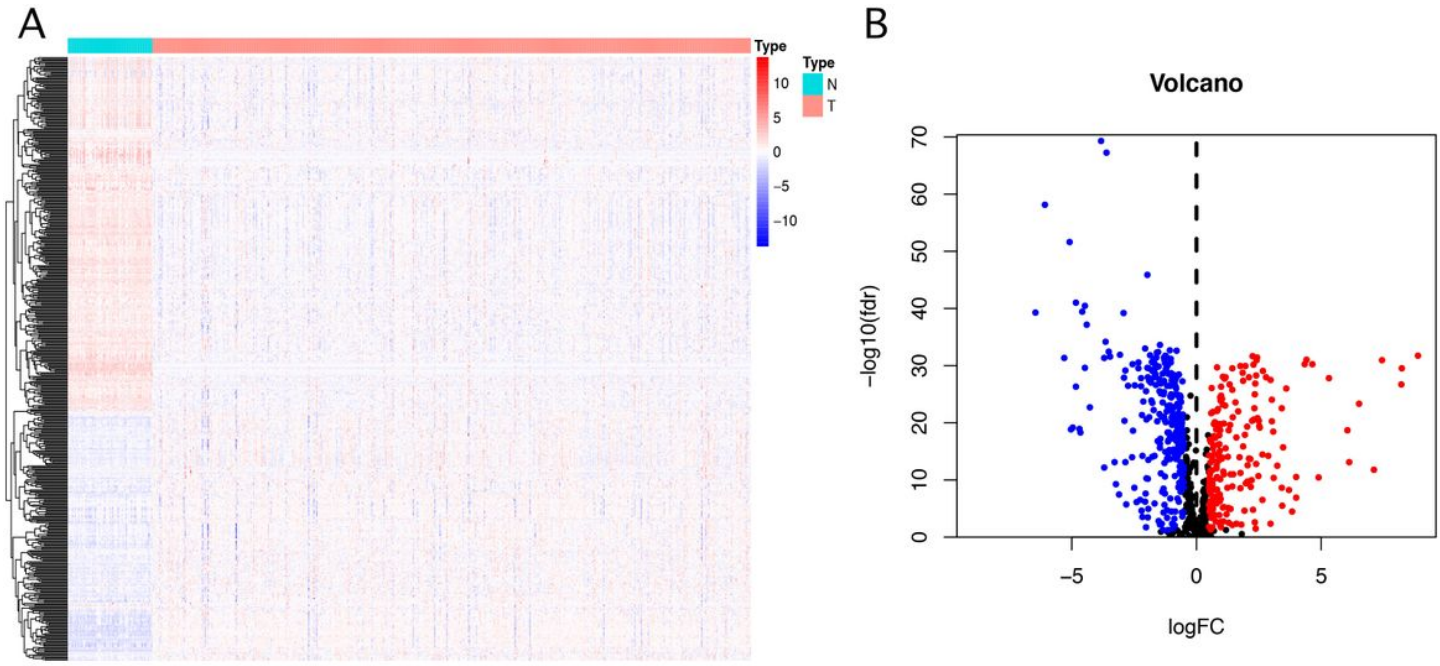

Figure 1

Identification of differentially expressed metabolic genes. (A) Heat map of MGs; the blue to red spectrum indicates low to high gene expression. (B) Volcano plot of MGs; the blue dots represent downregulated MGs, the red dots represent upregulated MGs and the black dots represent MGs that were not significantly differentially expressed.

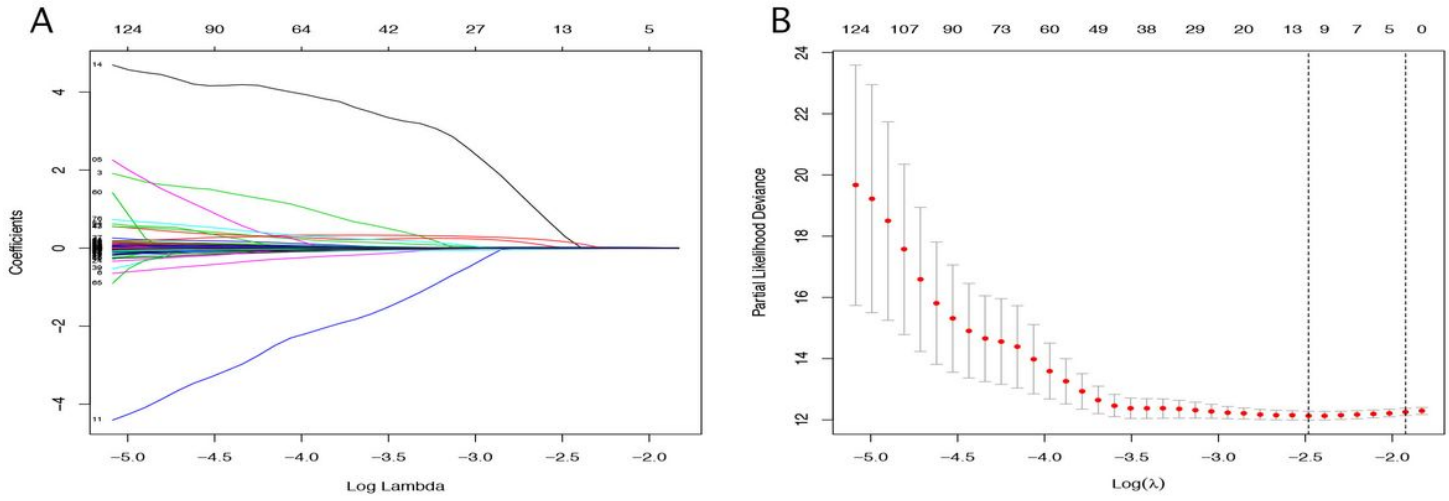

Figure 2

Further analysis of the prognostic differentially expressed MGs in the training cohort. (A and B) prognostic differentially expressed MGs selected through Lasso regression.

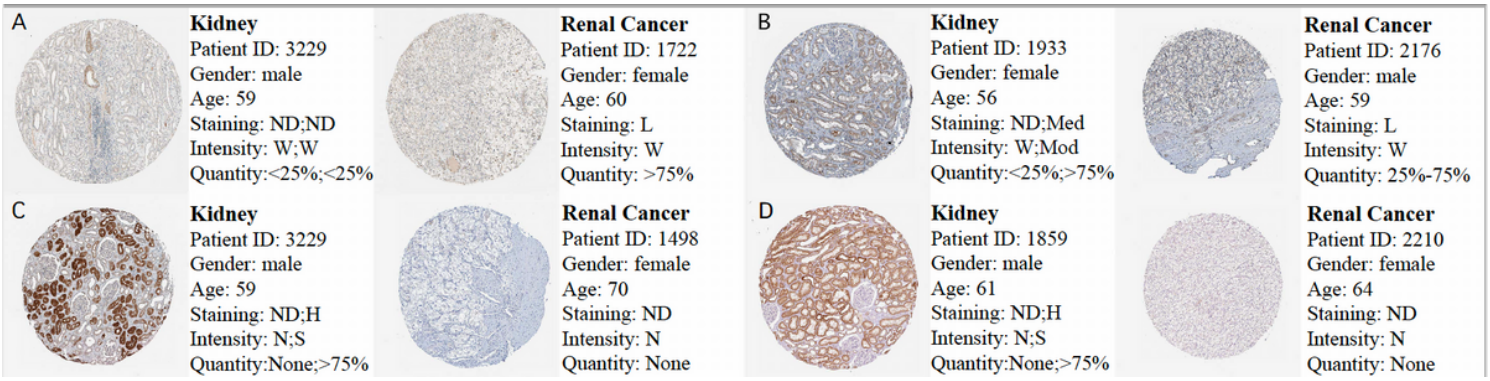

Figure 3

Validation of risk genes at the protein level by The Human Protein Atlas database (IHC). (A) Expression of P4HA3 in normal kidney tissue and renal cancer tissue (40x). (B) Expression of PAFAH2 in normal kidney tissue and renal cancer tissue (40x). (C) Expression of ALAD in normal kidney tissue and renal cancer tissue (40x). (D) Expression of ETNK2 in normal kidney tissue and renal cancer tissue (40x). In normal tissues, there are two types of IHC. The former refers to the staining state of cells in glomeruli, the latter refers to the staining state of cells in tubules. Abbreviations：L, Low; Med, Medium; H, High; ND, not detected; W, Weak; Mod, Moderate; S, Strong; N, None.

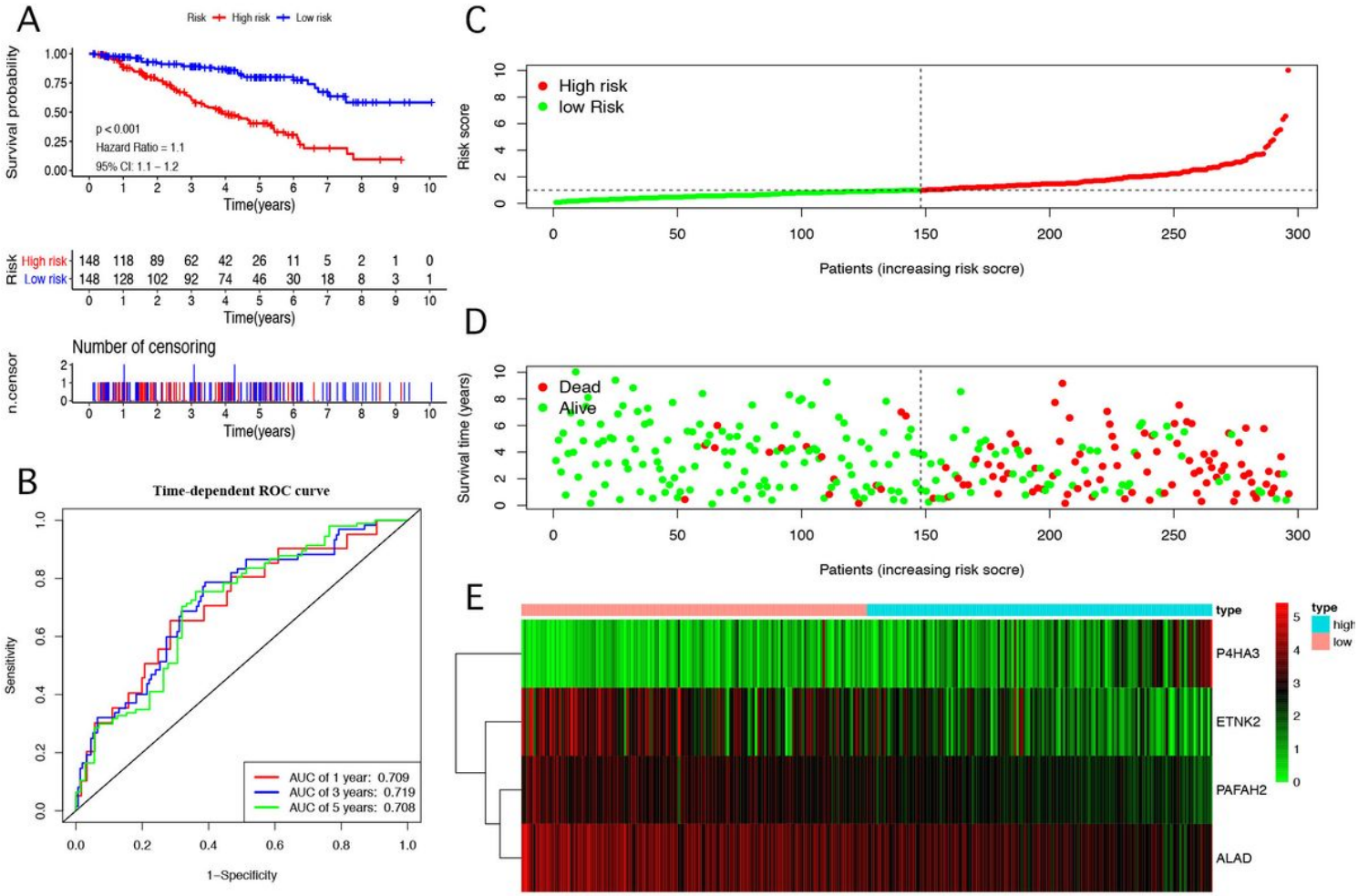

Figure 4

Prognostic analysis of the training cohort. (A) Kaplan-Meier curve analysis of the high-risk and low-risk groups. (B) Time-dependent ROC curve analysis of the prognostic model. (C) Risk score distribution of patients in the prognostic model. (D) Survival status scatter plots for patients in the prognostic model. (E) Expression patterns of risk genes in the prognostic model.

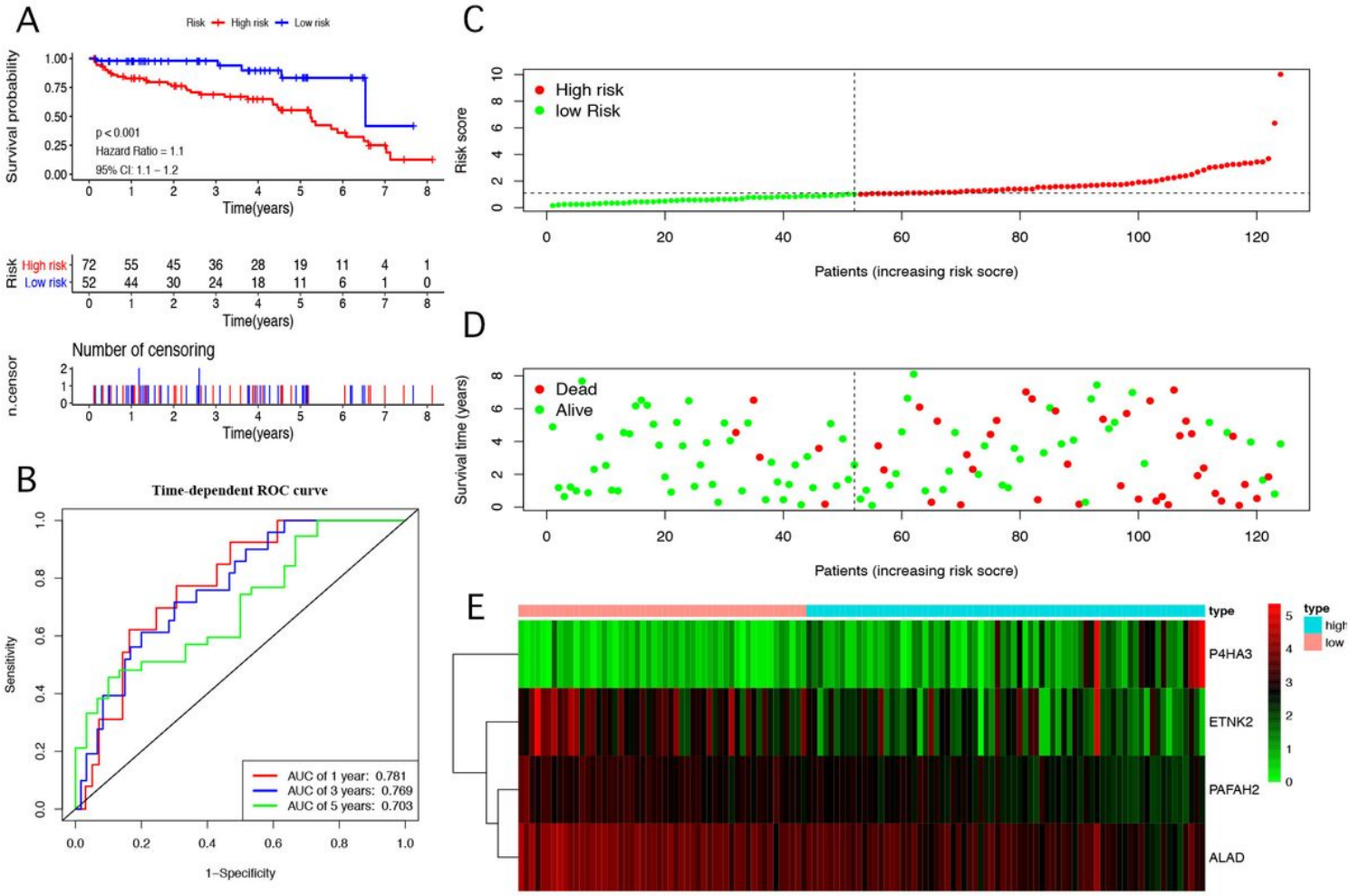

Figure 5

Prognostic analysis of the testing cohort. (A) Kaplan-Meier curve analysis of the high-risk and low-risk groups. (B) Time-dependent ROC curve analysis of the prognostic model. (C) Risk score distribution of patients in the prognostic model. (D) Survival status scatter plots for patients in the prognostic model. (E) Expression patterns of risk genes in the prognostic model.

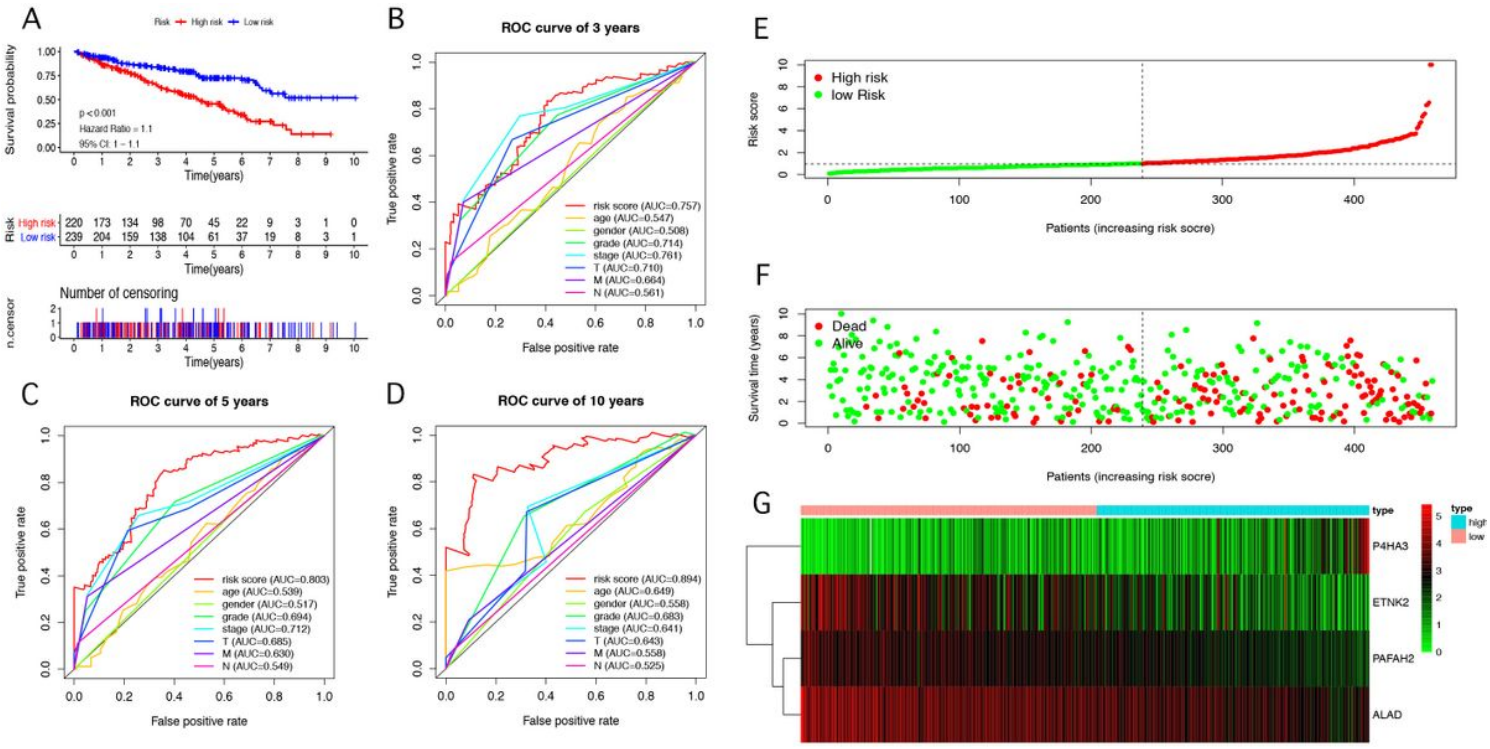

Figure 6

. Prognostic analysis of the entire cohort (TCGA database and GEO database). (A) Kaplan-Meier curve analysis of the high-risk and low-risk groups. (B) ROC curve analysis of different variables in the TCGA cohort at three years. (C) ROC curve analysis of different variables in the TCGA cohort at five years. (D) ROC curve analysis of different variables in the TCGA cohort at ten years. (E) Risk score distribution of patients in the prognostic model. (F) Survival status scatter plots for patients in the prognostic model. (G) Expression patterns of risk genes in the prognostic model.

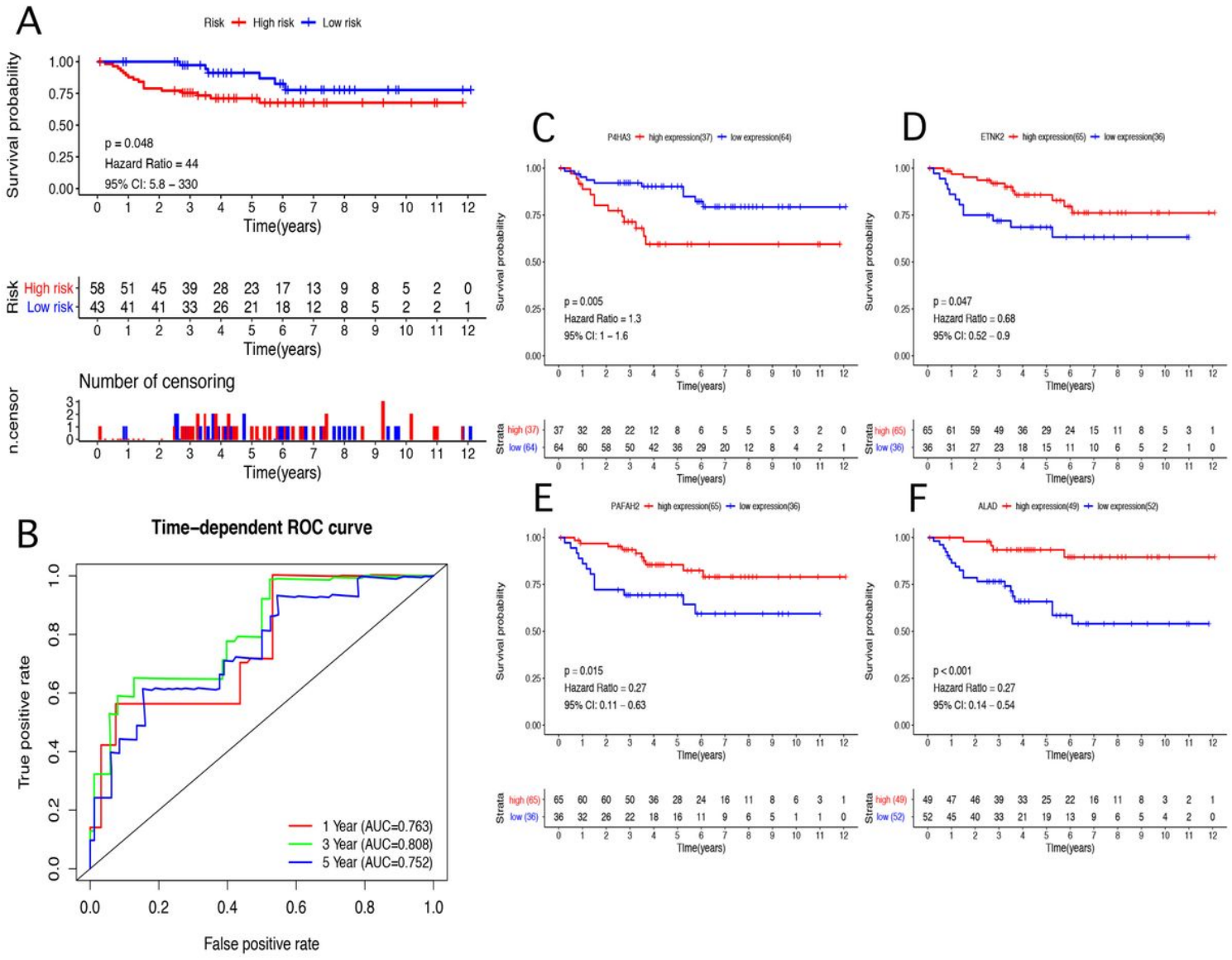

Figure 7

Prognostic analysis of the ArrayExpress cohort. (A) Kaplan-Meier curve analysis of the high-risk and low-risk groups. (B) Time-dependent ROC curve analysis of the prognostic model. (C) Kaplan-Meier survival curve analysis in the high P4HA3 expression group and low P4HA3 expression group. (D) Kaplan-Meier survival curve analysis in the high ETNK2 expression group and low ETNK2 expression group. (E) Kaplan-Meier survival curve analysis in the high PAFAH2 expression group and low PAFAH2 expression group. (F) Kaplan-Meier survival curve analysis in the high ALAD expression group and low ALAD expression group.

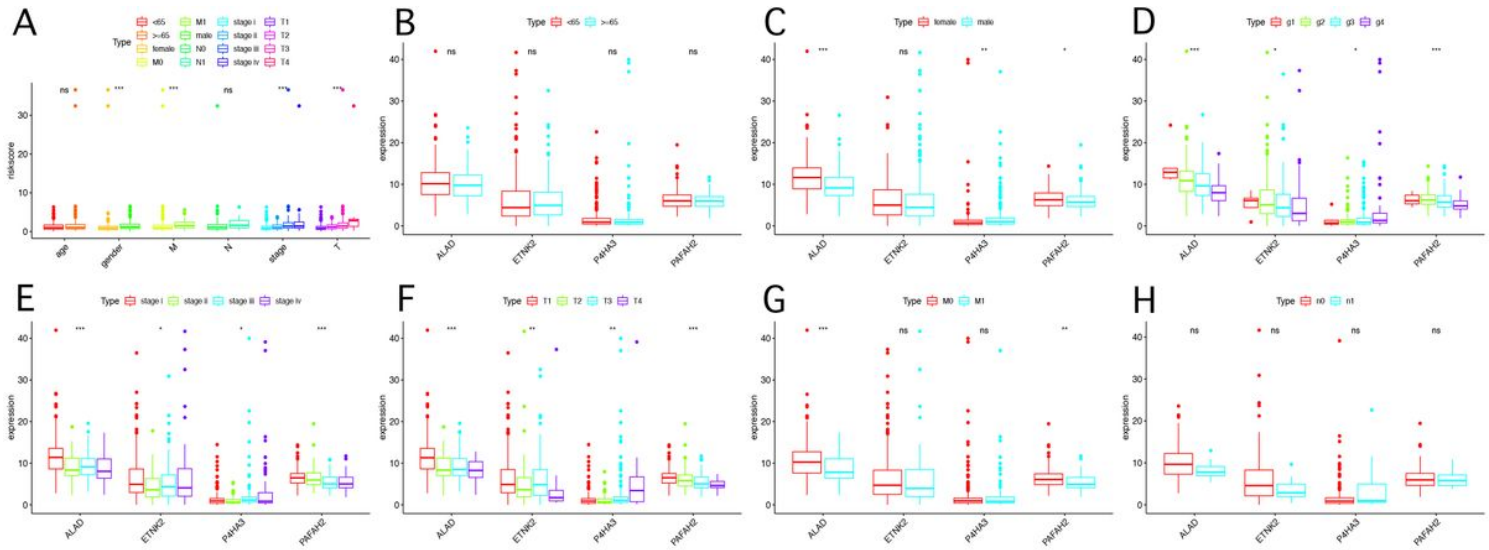

Figure 8

Relationships of the variables in the model with the clinical characteristics of patients in the TCGA cohort. (A) riskScore and clinical variables. (B) expression of risk genes and age. (C) expression of risk genes and gender. (D) expression of risk genes and grade. (E) expression of risk genes and stage. (F) expression of risk genes and T. (G) expression of risk genes and M. (H) expression of risk genes and N. Abbreviations： *：P < 0.05; **, P < 0.01; ***, P < 0.001.

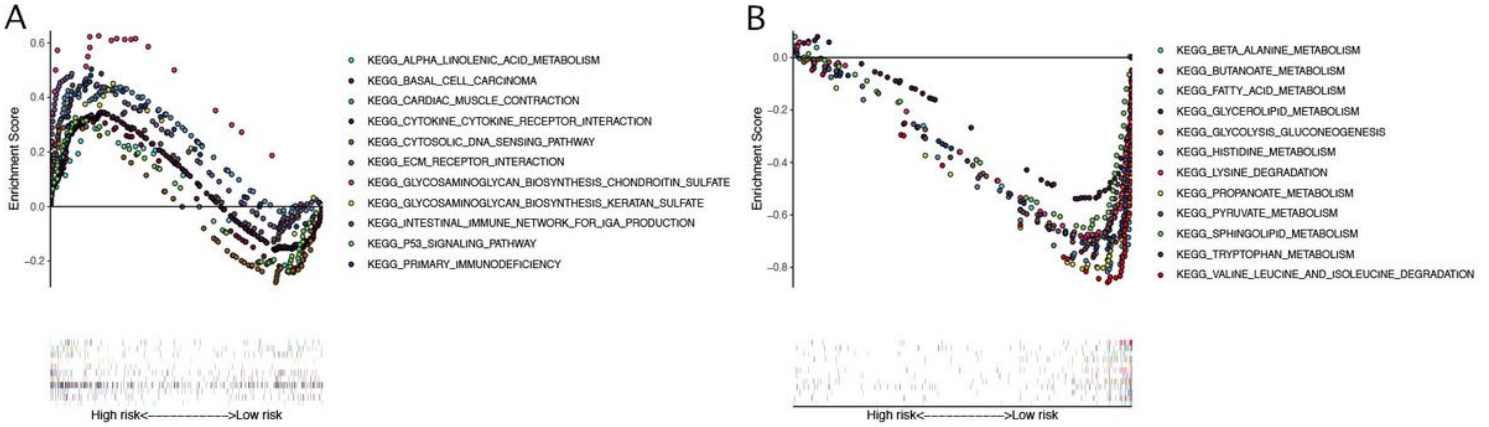

Figure 9

GSEA enrichment analysis. (A) Gene-set enrichment analysis of genes that are differentially expressed in high-risk group. (B) Gene-set enrichment analysis of genes that are differentially expressed in low-risk group.

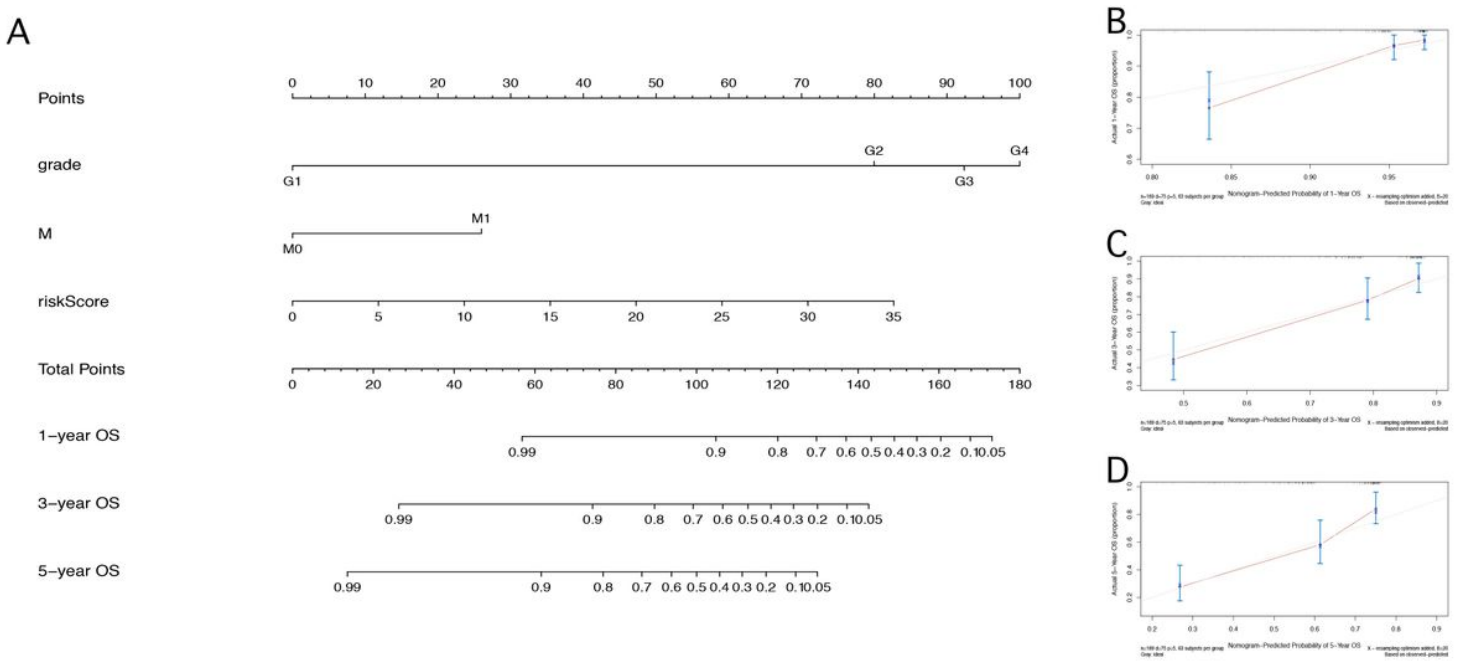

Figure 10

Nomogram with Calibration curves for the prediction of prognosis at one, three and five years in the TCGA cohort. (A) Nomogram for OS. (B) Calibration curves at 1 year. (C) Calibration curves at 3 years. (B) Calibration curves at 5 years.

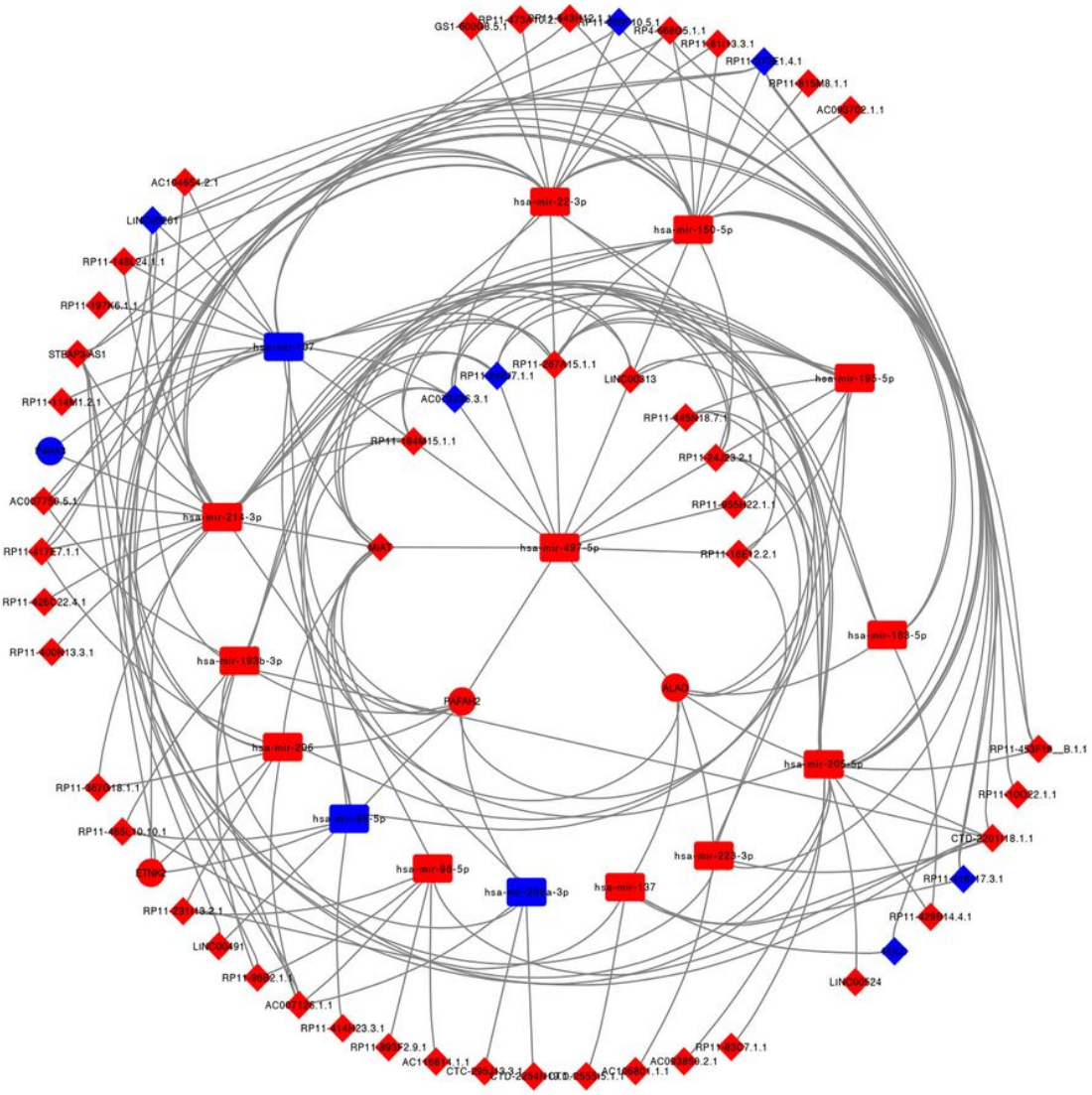

Figure 11

A lncRNA-microRNA-mRNA ceRNA network. Rectangles represent miRNA, circles represent mRNA, diamonds represent lncRNA. Red represents high expression and blue low expression.