High expression of the LAMA3/AC245041.2 gene pair associated with KRAS mutation and poor survival in pancreatic adenocarcinoma: A comprehensive TCGA analysis

Chengming Tian (384914620@qq.com)
The First Hospital of China Medical University: The First Affiliated Hospital of China Medical University
https://orcid.org/0000-0002-1220-9093

XIYAO LI
The First Hospital of China Medical University: The First Affiliated Hospital of China Medical University

CHUNLIN GE
The First Hospital of China Medical University: The First Affiliated Hospital of China Medical University
https://orcid.org/0000-0001-7413-435X

Research Article

Keywords: Pancreatic adenocarcinoma, KRAS, WGCNA, mRNA, lncRNA, prognosis model

Posted Date: April 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-361646/v1

Abstract

Background

Pancreatic adenocarcinoma (PAAD) is one of the worst cancers with high morbidity and mortality. Given that KRAS mutations may be an early event in PAAD. The present study aimed to identify differentially expression lncRNAs (DE-lncRNAs) and differentially expression mRNAs (DE-mRNAs) in KRAS-mutant PAAD. To explore the pathogenesis and molecular mechanism of PAAD development.

Methods

Clinical data of TCGA-PAAD patients were downloaded from the TCGA database and survival analysis was carried out in combination with KRAS mutation information. Through weighted gene correlation network analysis (WGCNA) and univariate cox regression analysis, we constructed the prognostic risk models to identify the hub DEmRNAs and DElncRNAs associated with PAAD prognosis. GO and KEGG enrichment analysis of DE-mRNA was performed. Multivariate cox regression was used to analyze the overall prognosis of age, gender, pathologic_T and KRAS mutations, and then the differences in clinical characteristics of risk score1 and risk score2 were analyzed. Finally, the mRNAs-lncRNA-TFs regulatory network was constructed.

Results

The functional enrichment analysis was performed after screening 1671 DE-mRNAs and 324 DE-lncRNA. It was found that the related pathways were mainly focused on the modulation of chemical synaptic transmission, synaptic membrane, ion-gated channel activity, ligand – receptor interactions that stimulate neural tissue, and so on. Univariate Cox regression analysis was used to screen 117 mRNAs and 36 lncRNAs, and the risk ratio models of mRNAs and lncRNAs was constructed. The LAMA3 (mRNA) and the AC245041.2 (lncRNA) have strong expression correlation in two risk models. The genes in the samples with high expression of the two genes were enriched into many transcription factors (TFs) related pathways, among which transcription factors related pathways including ATF5, CSHL1, NR1I2, SIPA1, HOXC13, HSF2, and HOXA10 were shared by the two groups, the core enrichment genes in the common TFs pathway were collated, and the regulatory network of mRNAs-lncRNAs-TFs was drawn.

Conclusion

In our study, new prognostic mRNAs and lncRNAs were identified, prognostic models were constructed respectively, and nomograms were constructed to guide clinical practice. We had drawn the regulatory network of mRNAs-lncRNAs-TFs, which can provide clues for further research in the future.

Introduction

Pancreatic adenocarcinoma (PAAD) is a malignant tumor occurring in the exocrine glands of the pancreas. Pancreatic malignancies may originate from pancreatic exocrine, endocrine or non-epithelial tissues, 95% of which are pancreatic adenocarcinoma. This disease a very poor prognosis, with high morbidity and mortality. Although the incidence and mortality of other common cancers have been decreasing, the mortality and number of deaths caused by pancreatic tumours have been increasing (1,2). Early surgery is the main treatment for PAAD, but pancreatic adenocarcinoma is difficult to diagnose at an early stage and most of it has metastasized by the time of initial diagnosis (3). Only 9.7% of them were localized at the time of diagnosis (4). Most patients die from liver, lung, and/or peritoneal metastasis, which is the most common site of spread (5). In addition, PAAD does not respond well to most chemotherapy drugs (6). Thus, exploring the molecular mechanisms underlying the pathogenesis and development of PAAD is urgently needed.

Molecular biology studies have shown that proto-oncogene activation and tumor suppressor gene inactivation and abnormal DNA repair genes are closely related to the occurrence of PAAD (7). Many important genes were mutated, among which the mutation rate of P16 in PAAD patients was 95%, KRAS 90%, P53 75% and DPC4 55% (8). KRAS is one of the most common mutant oncogene in human cancers. Experiments in cell culture and animal models confirmed that the development of many cancers depends on the sustained expression and signal transduction of KRAS (9,10). McCormick F believe that it was effective that attack KRAS cancers, by direct attack on the protein, or by indirect approaches such as siRNA or harnessing the immune system (11). In recent years, gene profiling and next-generation sequencing technologies have become indispensable tools for cancer research, as they allow detection of cancer-related genetic and epigenetic changes, such as mutations, copy number variations, and DNA methylation changes across more extensive genomic regions (12,13). The bioinformatics analysis of these data can provide valuable information for PAAD research. For example, Cheng Y synthesized several sets of public data and preliminarily clarified pathways and functions of pancreatic adenocarcinoma. Candidate molecular markers were identified for the diagnosis and prediction of prognosis of pancreatic adenocarcinoma, and candidate proteins were suggested to be attributable to the clonal and invasive nature of pancreatic cancer cells (14). Considering that KRAS mutation may be an early event of pancreatic adenocarcinoma, this study mainly started from the KRAS mutation grouping and constructed prognostic models based on WGCNA to predict the prognosis of the disease. Moreover, a line map was constructed to guide clinical practice, and the hub lncRNAs and mRNAs in the model were further analyzed, so as to construct a mRNA-lncRNA-TFs regulatory network, providing clues for subsequent analysis.

Materials And Methods

1. Data download:

A workflow chart of this study is shown in Figure 1.

PAAD expression data, clinical data and phenotypic data were downloaded from the TCGA database

Data of expression: https://gdc.xenahubs.net/download/TCGA-PAAD.htseq_counts.tsv.gz

Data of clinical: https://gdc.xenahubs.net/download/TCGA-PAAD.survival.tsv.gz

Phenotypic matrix:

https://gdc.xenahubs.net/download/TCGA-PAAD.GDC_phenotype.tsv.gz

Data of sample mutation :

https://portal.gdc.cancer.gov/files/fea333b5-78e0-43c8-bf76-4c78dd3fac92

Human gtf file from the Ensembl database (Homo_sapiens.GRCh38.99.gtf.gz), lncRNA and symbol information:

http://www.ensembl.org/info/data/ftp/index.html

Table 1. Phenotype statistics of TCGA-PAAD

	Tumor
KRAS	
Yes	128
No	49
Gender	
Male	97
Female	80
Age (Mean/Median = 65years)	
> 65	84
<= 65	93
Tumor_stage	
Stage I	24
Stage II	34
Stage III	35
Stage IV	17
Pathologic_M	
M0	79
M1	5
MX	93
pathologic_N	
N0	50
N1	122
NX	4
pathologic_T	
T1	7
T2	24
T3	141
T4	3
TX	1

2. Survival analysis

The KRAS mutation information was extracted from the mutation data of PAAD samples, phenotypic data was integrated, and the group was grouped according to whether KRAS mutated or not. Then KM curve was plotted, and the P value of the curve was less than 0.05, indicating significant survival difference among the groups.

3. Screening for differentially expressed lncRNAs and mRNAs

R package “edgeR” was used to obtain differentially expressed genes (DEGs). The expression matrix in the database was in the form of log2(count 1), so round(2^a-1) was used to get the counts of the sample. Then, the low-expression genes were filtered according to the CPM (count-per million) greater than 1 in at least 10 samples, and the differentially expressed genes were extracted with the threshold of |logFC| > 1 and FDR < 0.05. After that, the standardized expression matrix was further extracted, and the expression spectrum used in the subsequent analysis was all the standardized expression matrix. Then, the genetic information was obtained from the human gtf files, and DE-mRNAs and DE-lncRNAs were extracted for subsequent analysis.

4. Functional enrichment analysis:

First, R package “clusterProfiler” was used to conduct functional enrichment analysis of differential DE-mRNAs (15,16), and P value < 0.05 and Q value < 0.2 were used as thresholds for screening. After the enrichment pathway was obtained, R package “GOplot” was used for result visualization.

5. Weighted gene co-expression network analysis (WGCNA):

The weighted gene co-expression network analysis aims to find the gene modules of co-expression, explore the association between gene networks and phenotypes of concern, and mine the hub genes in the network. The main principle is to take the correlation coefficient of the expression quantity between genes to the power of n, and the direct result is to amplify the difference of the correlation between genes. Take a specific value β for the power of the correlation coefficient between each pair of genes (i□j), so as to calculate the correlation between all genes, that is, the adjacency matrix: $a_{i,j} = |cor(i, j)|^{\beta}$. In order to better determine whether two genes have similar expression profiles, WGCNA adopts a method based on soft threshold. Since the result of adjacency ($a_{i□j}$) directly depends on the value of β, which directly affects the construction of the module and the division of adjacent genes, WGCNA calculates the value β according to the adjacent lowest value of the scale free network. A scale-free network is characterized by a small number of nodes with degrees significantly higher than the general points, which are called hub. A few hubs are associated with other nodes and ultimately constitute the entire network. Selected β value of network construction, and then carried out network construction and module identification from four steps: The similarity between each gene was calculated by topological overlap; Gene cluster tree was obtained; Genes with the same expression were divided into the same module by cutting the tree; Merge similar modules, after the module classification was obtained, the correlation between different modules

and phenotypes was calculated, and the more relevant modules were used for subsequent analysis. The R package “WGCNA” was used to analyze the weighted co-expression network of all differentially expressed genes (17), and the module with the strongest correlation with prognostic traits was obtained for subsequent analysis.

6. Univariate cox regression analysis:

In order to deeply explore genes related to prognosis in differentially expressed genes and combine survival data, R-package “survival” and “survminer” were used for batch univariate Cox regression analysis. After regression analysis, the significantly correlated genes were screened with $P < 0.05$ as the threshold value for subsequent model construction, and the top 6 genes were selected for Kaplan-Meier analysis.

7. Construction of prognostic risk model:

LASSO regression dimensionality reduction was further carried out for mRNAs and lncRNAs in Cox regression results, and a risk scoring model was constructed, which mainly depended on R package “glmnet”. In order to build a more accurate regression model, lambda screening was first carried out by cross-validation, then the corresponding model of lamdba.min was selected to further extract the expression matrix of relevant genes in the model, and the risk score of each sample was calculated based on the following formula:

$$\mathrm{RScore_i} = \sum_{j=1}^{n} \mathrm{exp_{ji}} \times \beta_j$$

Where exp represents the expression level of the corresponding gene, represents the regression coefficient (coef) of the corresponding gene in the multivariate regression results, RScore represents the expression level of the significantly related gene in each sample multiplied by the coef of the corresponding gene and summed over, i represents the sample, and j represents the gene. Based on the risk scores of the samples, the high and low risk groups were divided by the median of the nodes, so as to carry out the later model performance detection.

On the basis of the above analysis, the high and low risk group was obtained, Kaplan-meier analysis was carried out in combination with the survival data. Then, the ROC curve was drawn with the sample risk score as the model prediction result. The curve AUC value was greater than 0.6, indicating that the model was of good performance.

8. Multivariate cox regression analysis and nomogram construction:

To verify the mRNA and lncRNA prognostic model as an independent prognostic factor of the disease, cox multivariate regression analysis was used to analyze the overall prognosis of age, gender,

pathologic_T, and KRAS mutations.

A nomogram can visualize the results of Cox regression. It sets the scoring standard according to the regression coefficient of all independent variables, gives a score for each value level of each independent variable, calculates an total score for each patient, and calculates the probability of the outcome time of each patient through the conversion function between the score and the probability of the outcome.

The nomogram was drawn mainly using the R package “rms” and “survival”. First, the scale risk regression model was constructed by cph(), then the survival probability was calculated by survivalcph() function, and finally the nomogram object was constructed by nomogramcph() function and displayed by plotcph().

9. The difference analysis of clinical characteristics of risk score1 and risk score2

Clinical indicators age, Gander, pathologic_M, pathologic_N, pathologic_T and Tumor_stage were selected to further detect the differences between risk score1 and risk score2 in these indicators, and the “ggpubr” package was used to draw boxplot to represent the results. After that, the difference in distribution within the group was further detected with t test, so as to verify whether the risk score was consistent with the clinical indicators.

10. Analysis of important regulatory relationships in risk prognostic models:

The mRNA and lncRNA with the strongest correlation with prognostic traits were obtained from the two models respectively, KRAS mutation samples were extracted, and mutation samples were grouped based on the expression of hub mRNA and lncRNA respectively. Then the KM curve was drawn to explore the relationship between the expression of hub mRNA and lncRNA and prognostic traits. After that, the expression levels of hub mRNA and lncRNA were further used to predict the prognosis of the samples, and the ROC curve was drawn. Finally, multivariate COX regression analysis was carried out in combination with the clinical phenotypes to verify the independent prognostic efficacy of hub mRNA and lncRNA.

GSEA analysis of hub mRNA and lncRNA was further performed (18,19). The nodes were divided into high and low expression groups based on the median expression of hub mRNA and lncRNA, and then GSEA analysis was carried out. The results were filtered with P value < 0.05 and FDR< 0.25.

11. Construction of TF regulation network related to hub mRNA-lncRNA regulatory axis:

According to the expression levels of hub mRNA and lncRNA, differentially expressed genes were extracted with edgeR, and the core genes enriched into the TFs pathway by the key single gene GSEA were integrated to draw the regulatory network of mRNAS-lncRNA-TFs.

Results

1. Data download:

PAAD related expression data and clinical data were downloaded from UCSC TGCA database to remove the missing clinical information samples and integrate the KRAS mutation phenotype of the samples, and finally the data information of 177 cancer samples were obtained, including 128 KRAS mutation samples and 49 non-mutation samples. (Table 1)

2. Survival analysis:

According to KRAS mutated or not, the samples were divided into two groups, and then survival analysis was carried out. KM curve showed that the survival difference between the two groups was significant (P = 0.013), and the survival curve of KRAS mutation samples decreased faster. (Figure 2) Here we also confirmed that the survival rate of patients in the KRAS-mutant group was significantly lower than that in the KRAS-wildtype group.

3. DEmRNAs and DElncRNAs between KRAS-mutant and KRAS-wildtype PAAD:

To further explore KRAS mutation-related mRNAs and lncRNAs in PAAD, we analyzed the differential expression of lncRNA and mRNA in patients with KRAS-mutant group and KRAS-wildtype group. The DE-mRNAs and DE-lncRNAs were obtained from the expression data of TCGA-PAAD cancer samples, using |logFC| > 1 and FDR < 0.05 as threshold screening ,1671 DE-mRNAs (up-regulated 368, down-regulated 1302) and 324 DE-lncRNAs (up-regulated 56, down-regulated 171) was obtained. (Figure 3)

4. DE-mRNA GO□KEGG enrichment analysis:

The functional enrichment analysis of DE-mRNAs was further carried out, and the GO enrichment analysis was divided into three parts: Biological Process (BP), Cell Components (CC), Molecular Function (MF), in which BP enrichment pathway is mainly the modulation of chemical synaptic transmission, regulation of trans-synapic signaling and the signal release, CC enrichment pathway is mainly synaptic membrane, neuronal Cell body, etc., MF enrichment pathway is mainly channel activity and passive transmembrane transporter activity and ions gated channel activity, etc. The pathway of KEGG enrichment are mainly neuroactive ligand-receptor interaction, insulin secretion and so on. (Figure 4)

5. WGCNA and identification of prognosis-associated module

The WGCNA was further constructed based on DEGs and identified prognosis-related modules. Firstly, the soft threshold was calculated, and R^2 > 0.85 was taken as the filtering threshold to get power=6. The network was further constructed by one-step method with Power =6, and similar modules were combined with height < 0.25 as the threshold. Finally, 4 modules were obtained, among which the genes in blue module (338) had the highest correlation with OS_status and OS_time, and also had the highest correlation with KRAS mutation. Therefore, the genes in this module were selected for subsequent analysis. (Figure 5)

6. Univariate cox regression analysis:

The differential expression matrix of 177 cancer samples was extracted for Cox regression analysis. After screening by P value, 153 genes significantly related to PAAD were obtained, and the KM curve was drawn for the top 6. The survival curve of MYEOV, WNT7A and FAM83A-AS1 high-expression samples decreased faster, and the hazard ratio was greater then 1, 95% Cl interval lower than 1, indicating that the high expression of these three genes might threaten survival. The survival curves of KATNAL2、GLTPD2 and KCNJ2-AS1 samples with high expression decreased more slowly, and the hazard ratio was greater then 1, 95%Cl interval was also less than 1, indicating that the low expression of these three genes might threaten survival. (Figure 6)

7. Construction of mRNA and lncRNA prognostic risk model:

Seven mRNAs significantly related to the prognosis of PAAD were screened out from 117 mRNAs using LASSO regression, and then built a risk ratio model based on the expression of 7 markers and regression coefficients, that is, risk score1 = GLTPD2 * (-0.113) + RP1 * 0.008 + MUC21 * 0.016 + FAM83A * 0.029 + MYEOV * 0.035 + ZNF488 * 0.083 + LAMA3 * 0.153. Eight lncRNAs significantly related to the prognosis of PAAD were screened out from 36 lncRNAs using LASSO regression, and then built a risk ratio model based on the expression of 8 markers and regression coefficients, that is, risk score2 = AC068580.2 * 0.003 + LINC01910 * 0.035 + AC245041.2 * 0.044 + AC107959.3 * 0.058 + CASC8 * 0.100 + UCA1 * 0.102 + LINC00520 * 0.126 + AL033384.1 * 0.170. After that, the risk scores of each sample were calculated, and the median was further divided into high and low risk groups, and the KM curve of high and low risk groups was drawn. The results showed that the difference of high and low risk groups was significant, ($p < 0.0001$) and the AUC values of 1 year ,3 years and 5 years in ROC curve are all greater than 0.75, which indicates that mRNAs and lncRNA model has good prediction efficiency. (Figure 7,8)

8. Multivariate cox regression analysis and nomogram construction:

Further combined with age, Gender, pathologic_T and KRAS mutations, cox multivariate regression was used to verify the mRNAs and lncRNAs prognostic model. Multivariate cox regression results showed that the risk score1 (2.597) and risk score2 (3.698) had the highest HR, respectively, and pathologic_T also had better predictive efficacy. However, the predictive efficacy of age, gender and KRAS mutation was poor. HR < 1 and P value was not significant in multivariate regression, indicating that KRAS mutated or not was not appropriate to be used as a single factor for prognostic analysis. Therefore, the factor was removed in subsequent construction of nomogram. The high risk, high age, and high pathologic_T grade samples scored higher and had a higher survival risk in the nomogram. (Figure 9,10)

9. Difference analysis of clinical characteristics of risk score1 and risk score2:

It could be seen from Figure 11 and 12 that the risk scores of the two models are significantly different in different KRAS mutation states, and the risk scores of patients with KRAS mutation were generally higher, indicating that our models were closely related to KRAS mutation. Further observation of the difference of

risk score in the Pathologic_T grouping showed that the statistical results were meaningless due to the small number of T1 and T4 patients, while the risk score of T2 patients was significantly lower than that of T3 patients, which indicated that the prediction results of the two models were in good agreement with the diagnosis of Pathologic_T. Finally, we observed the difference of risk score in tumor stage grouping. Also, because the sample size of Stage III and Stage IV was too small, the statistical results were meaningless, so we focused on Stage I and Stage II. We could see that the risk score of Stage II patients was significantly higher than Stage I, which indicated that risk score and tumor stage had a good consistency. Therefore, we can conclude that the risk score of the two models are consistent with the clinical diagnosis results as well as the mutation results of KRAS, which further indicates that our two models have good accuracy.

10. Important regulatory relationships in risk - prognostic models:

In the two risk models, we used Pearson's correlation to analyze the correlation of each node in the two groups and it was found that LAMA3 (mRNA) and AC245041.2 (lncRNA) were located in the two risk models respectively, and the expression correlation of these two genes was the highest. ROC curve was drawn to further verify whether the expression of these two genes had independent prognostic effect. It was found that the AUC values of both genes were greater than 0.65, indicating that the expression values of both genes could be used as independent prognostic factors. Then GSEA analysis was performed on the two genes, and the enrichment results were filtered with P value < 0.05 and FDR < 0.25 as the threshold. The results showed that the genes in the high expression samples of the two genes were enriched to many TFs related pathways, among which the ATF5、CSHL1、NR1I2、SIPA1、HOXC13、HSF2、HOXA10 were common in the two groups. (Figure 13、14)

11. Construction of TF regulatory network related to hub mRNA-lncRNA regulatory axis:

Considering the high expression samples of the two genes were enriched many TFs related pathways, we constructed the TF regulatory network related to the hub mRNNA-lncRNA regulatory axis based on LAMA3 and AC245041.2. According to the enrichment results of genes in the LAMA3 high expression samples, the core enrichment genes in the common TFs pathway were sorted out, and the differently expressed genes based on the high and low expression groups of LAMA3 and AC245041.2 were obtained and integrated. Then the mRNA-lncRNA-TFs regulatory network was plotted. (Figure 15)

Discussion

As one of the most malignant tumors in the digestive system, PAAD has an alarming mortality rate in both eastern and western societies (20,21). Current clinical data show that surgery remains the only chance of cure, yet only 20% of patients will be alive at 5 years after pancreatic resection, and the benefits of chemotherapy are also limited (22,23). Therefore, it is important to identify new molecular biomarkers and to understand the underlying mechanisms of the occurrence and progression of PAAD. The importance of KRAS activation in PAAD was demonstrated in earlier sequencing in PAAD (24). Among them, about 90% of the pancreatic cancer genomes sequenced by targeted sequencing, whole exome or

whole genome sequencing showed carcinogenic KRAS. Activation of oncogenic KRAS in PAAD is associated with the occurrence and progression of tumors in many aspects, including deregulation of key signal transduction pathways, metabolic changes, metastasis, and drug resistance (25).

However, the activation mechanism of mutant KRAS in PAAD has not been clearly confirmed (26). It is of great significance to further elucidate the potential genes related to KRAS mutation and PAAD prognosis. Mining data from the TCGA-PAAD dataset can help identify prognostic factors that may be involved in cancer occurrence and progression. In this study, using the TCGA-PAAD dataset, we identified DElcnRNAs and DEmRNAs between KRAS-mutant and KRAS-wildtype PAAD. The results of enrichment analysis showed that the enrichment mainly included the pathways for the modulation of chemical synaptic transmission, regulation of trans-synapic signaling, the signal release, synaptic membrane, neuronal Cell body, channel activity and passive transmembrane transporter activity and ions gated channel activity etc. Weighted co-expression network analysis of all differentially expressed genes was performed to obtain the module with the strongest correlation with prognostic traits. Using multivariate Cox regression analysis, we constructed prognostic risk models for lcnRNA and mRNA to identify hub differentially expressed genes associated with PAAD prognosis. Moreover, we found that the distribution of these differentially expressed genes was related to the development of PAAD in the analysis of the differences in clinical characteristics of risk score1 and risk score2.

We screened out mRNAs: GLTPD2、RP1、MUC21、FAM83A 、 MYEOV、 ZNF488、LAMA3 and LncRNA、 AC068580.2、LINC01910、AC245041.2、AC107959.3、CASC8、UCA1、LINC00520 、AL033384.1 with significant prognostic correlation. Among them, FAM83A gene was amplified in many human cancers, while silencing FAM83A in related cancer cell lines inhibited the activation of WNT /β-catenin and TGF-β signaling pathways and reduced tumorigenicity (27,28). Chen S showed that overexpression of FAM83A significantly promoted the cancer stem cell-like characteristics and chemotherapyresistance of tumor cells in vitro and in vivo in mouse models of pancreatic Cancer, while inhibition of FAM83A reduced the drug resistance of tumor cells (29). Kim J used gene analysis of differential expression between normal pancreas and PAAD tissues to calculate the prognostic gene expression model through LASSO regression analysis, suggesting that LAMA3, E2F7, IFI44, SLC12A2 and LRIG1 genes might be potential drug targets for PAAD (30). In addition, Yang C used an online public database to evaluate mRNA expression and prognostic value of the laminin subunits in pancreatic ductal adenocarcinoma tissues, found that LAMA3 and LAMC2 were positively correlated with the amount of pancreatic ductal adenocarcinoma blood and were considered as potential therapeutic targets and prognostic markers for pancreatic ductal adenocarcinoma (31). However, none of the above studies has explained the specific mechanism of LAMA3 action. In our study, LAMA3 (mRNA) and AC245041.2 (lncRNA), the genes with the highest correlation, were found in the two risk models, and both genes were highly expressed in KRAS mutant PAAD as independent prognostic factors. The genes in the sample with high expression of LAMA3 have been enriched to many transcription factors related pathways. We collated the core enrichment genes in the common TFs pathway and mapped the regulatory network of mRNA-lncRNA-TFs, which may be closely related to the prognosis of PAAD.

Conclusion

Taken together, we identified hub lncRNAs and mRNAs associated with KRAS mutation and PAAD prognosis through a comprehensive bioinformatics analysis. In addition, we constructed a mRNA-lncRNA-TFs network. Our findings may deepen the understanding of the pathogenesis of KRAS mutant PAAD and provide clues and new ideas for further research.

Declarations

Acknowledgements

The results of our study are based on data from TCGA (https://portal.gdc.cancer.gov/)

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Author contributions

The authors contributed to this study and manuscript in the following manner:

data collection, CT and XL; statistical analysis, CT and XL; writing and editing, CT; supervision, CG; funding acquisition, CG. All authors read and approved the fifinal manuscript.

Funding

The study was supported by the Educational Commission of Liaoning Province, People's Republic of China (Grant Number L2014294).

Availability of data and materials

The data used to support the fndings of this study are included in the article.

Competing interests

The authors declare that they have no competing interests.

Author details

Department of Hepatobiliary and Pancreatic Surgery, The First Hospital of China

Medical University, Shenyang 110001, Liaoning, China.

Correspondence: 13998303666@139.com

Institutional email address: cmugs@cmu.edu.cn

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68:7–30. doi:10.3322/caac.21442.
2. Rahib L, Smith BD, Aizenberg R, et al. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res. 2014;74:2913–2921.
3. DeSantis CE, Lin CC, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2014. *CA Cancer J Clin*. 2014;64(4):252-271. doi:10.3322/caac.21235
4. National Cancer Institute. 2018 [cited March 14, 2018]. In: Pancreatic Cancer Cancer Stat Facts [Internet] Available from: https://seer.cancer.gov/
5. Yachida S, Iacobuzio-Donahue CA. The pathology and genetics of metastatic pancreatic cancer. Arch. Pathol. Lab. Med. 2009;133:413–422.
6. Vincent A, Herman J, Schulick R,et al. Pancreatic cancer. *Lancet*. 2011;378(9791):607-620.
7. Roberts NJ, Norris AL, Petersen GM, et al. Whole Genome Sequencing Defines the Genetic Heterogeneity of Familial Pancreatic Cancer. *Cancer Discov*. 2016;6(2):166-175.
8. Jones S, Zhang X, Parsons DW, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*. 2008;321(5897):1801-1806.
9. Haigis KM. KRAS Alleles: The Devil Is in the Detail. *Trends Cancer*. 2017;3(10):686-697.
10. Hayes TK, Neel NF, Hu C, Gautam P, et al. 2016. Long-term ERK inhibition in KRAS-mutant pancreatic cancer is associated with MYC degradation and senescence-like growth suppression. Cancer Cell 29: 75–89.
11. McCormick F. KRAS as a Therapeutic Target. *Clin Cancer Res*. 2015;21(8):1797-1801.
12. Huang CC, Du M, Wang L. Bioinformatics analysis for circulating cell-free DNA in cancer. Cancers (Basel). 2019;11(6):E805.
13. Stark R, Grzelak M, Hadfeld J. RNA sequencing: the teenage years. Nat Rev Genet. 2019;20:631–56.
14. Cheng Y, Wang K, Geng L, et al. Identification of candidate diagnostic and prognostic biomarkers for pancreatic carcinoma. *EBioMedicine*. 2019;40:382-393.
15. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-287. doi:10.1089/omi.2011.0118
16. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res*. 2010;38(Database issue):D355-D360. doi:10.1093/nar/gkp896

17. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. (2008) 9:559. doi: 10.1186/1471-2105-9-559
18. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550. doi:10.1073/pnas.0506580102
19. Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003;34(3):267-273. doi:10.1038/ng1180
20. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136(5):E359-E386. doi:10.1002/ijc.29210
21. Ryan DP, Hong TS, Bardeesy N. Pancreatic adenocarcinoma. *N Engl J Med*. 2014;371(11):1039-1049. doi:10.1056/NEJMra1404198
22. Dreyer SB, Chang DK, Bailey P, Biankin AV. Pancreatic Cancer Genomes: Implications for Clinical Management and Therapeutic Development. *Clin Cancer Res*. 2017;23(7):1638-1646.
23. Zhu H, Li T, Du Y, Li M. Pancreatic cancer: challenges and opportunities. *BMC Med*. 2018;16(1):214. Published 2018 Nov 22. doi:10.1186/s12916-018-1215-3
24. Bailey P, Chang DK, Nones K, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature*. 2016;531(7592):47-52. doi:10.1038/nature16965
25. Mann KM, Ying H, Juan J, Jenkins NA, Copeland NG. KRAS-related proteins in pancreatic cancer. *Pharmacol Ther*. 2016;168:29-42. doi:10.1016/j.pharmthera.2016.09.003
26. Grant TJ, Hua K, Singh A. Molecular Pathogenesis of Pancreatic Cancer. *Prog Mol Biol Transl Sci*. 2016;144:241-275.
27. Zhang J, Sun G, Mei X. Elevated FAM83A expression predicts poorer clincal outcome in lung adenocarcinoma. *Cancer Biomark*. 2019;26(3):367-373. doi:10.3233/CBM-190520
28. Liu C, Peng X, Li Y, et al. Positive feedback loop of FAM83A/PI3K/AKT/c-Jun induces migration, invasion and metastasis in hepatocellular carcinoma. *Biomed Pharmacother*. 2020;123:109780. doi:10.1016/j.biopha.2019.109780
29. Chen S, Huang J, Liu Z, Liang Q, Zhang N, Jin Y. FAM83A is amplified and promotes cancer stem cell-like traits and chemoresistance in pancreatic cancer. *Oncogenesis*. 2017;6(3):e300. Published 2017 Mar 13. doi:10.1038/oncsis.2017.3
30. Kim J, Jo YH, Jang M, et al. PAC-5 Gene Expression Signature for Predicting Prognosis of Patients with Pancreatic Adenocarcinoma. *Cancers (Basel)*. 2019;11(11):1749. Published 2019 Nov 7. doi:10.3390/cancers11111749
31. Yang C, Liu Z, Zeng X, et al. Evaluation of the diagnostic ability of laminin gene family for pancreatic ductal adenocarcinoma. *Aging (Albany NY)*. 2019;11(11):3679-3703. doi:10.18632/aging.102007

Figures

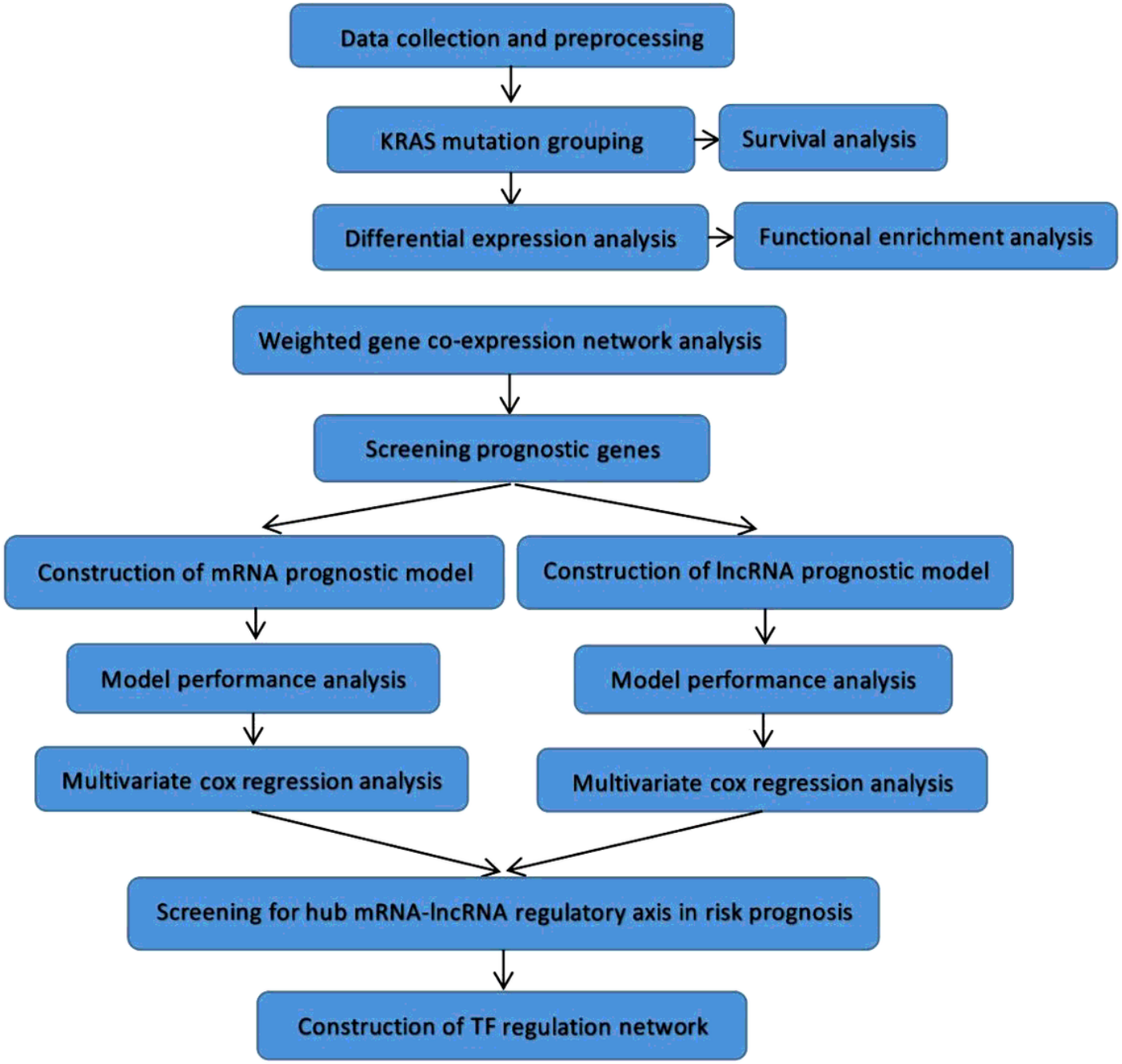

Figure 1

workflow chart.

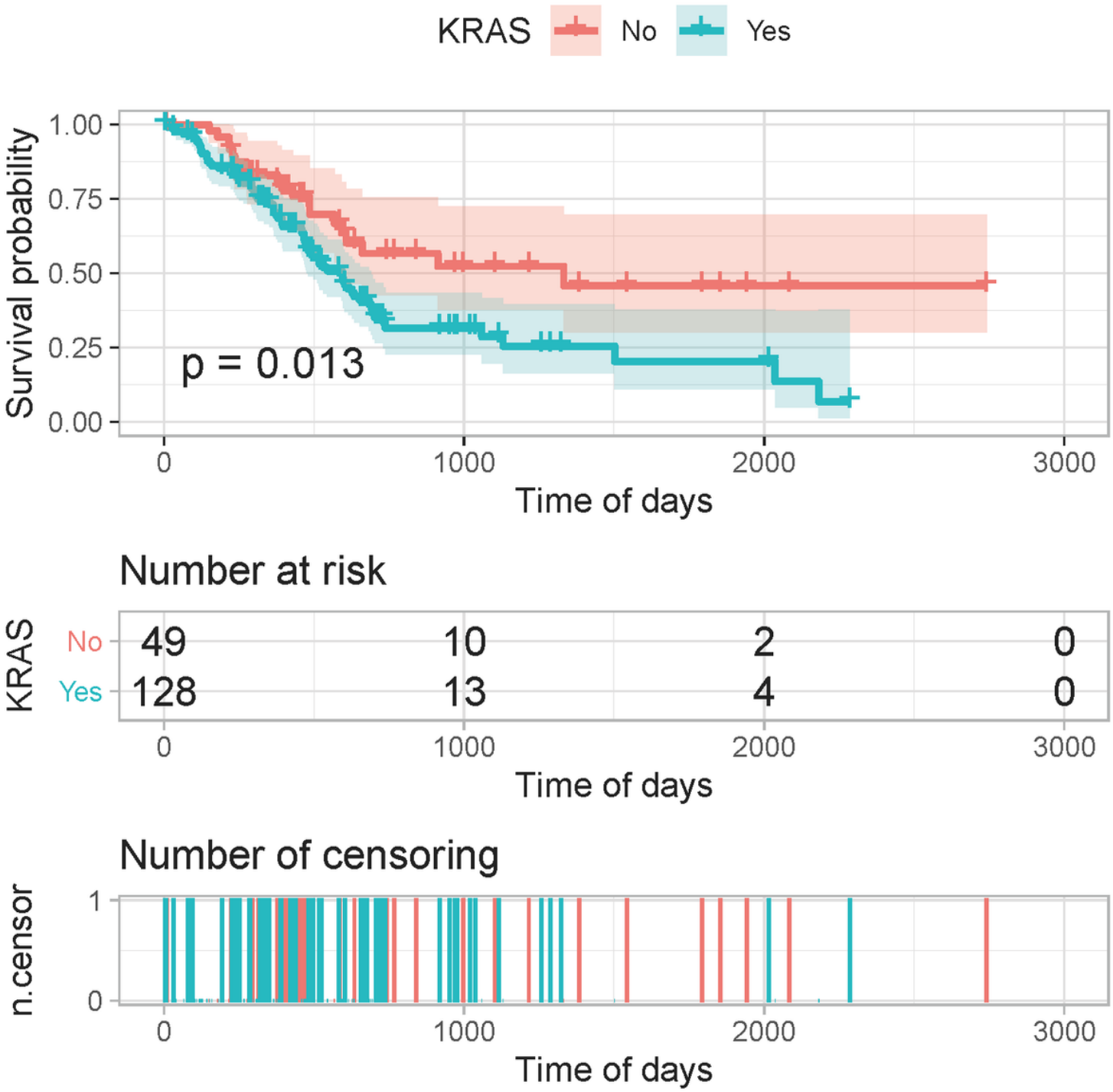

Figure 2

Survival analysis of KRAS mutation:

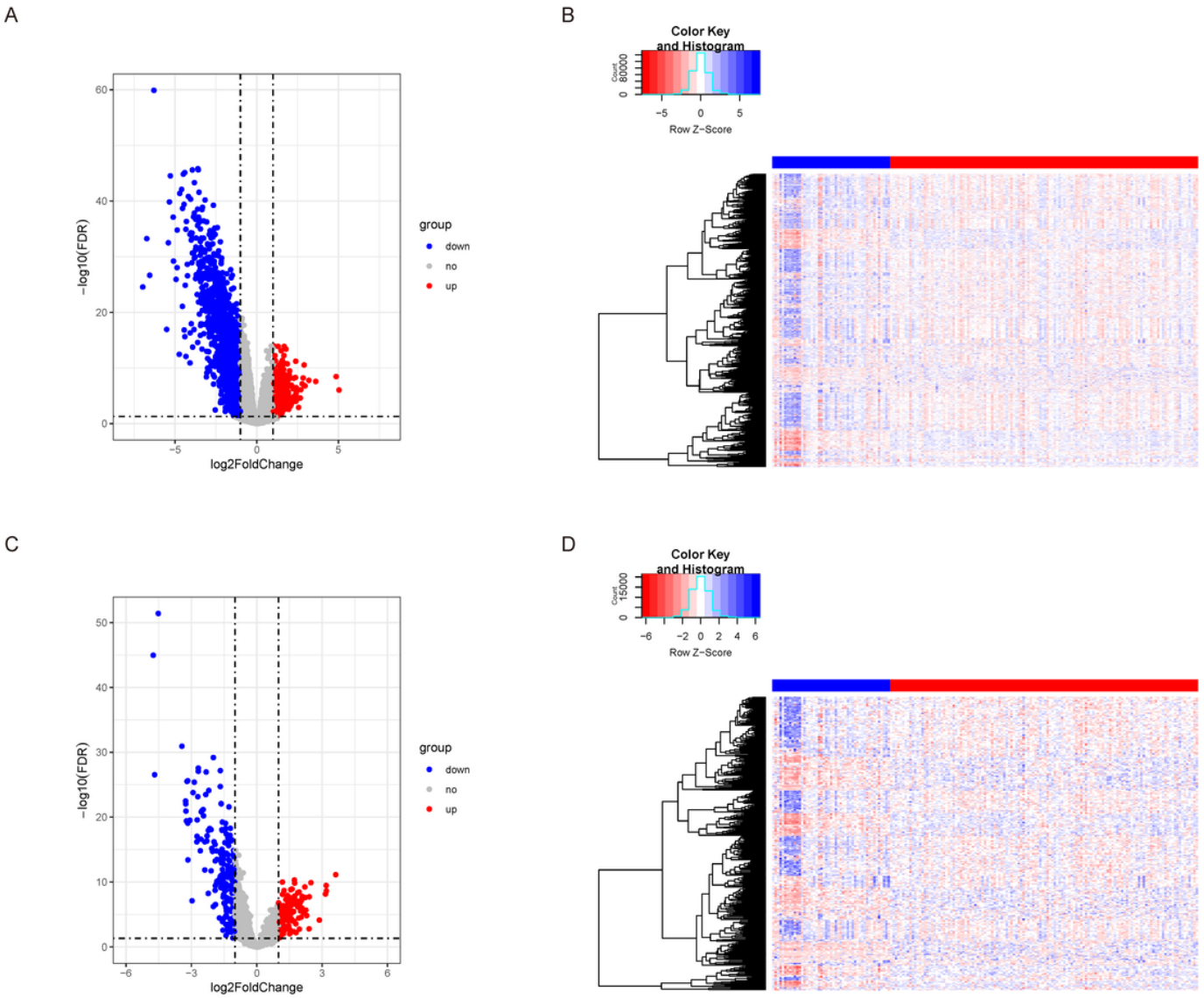

Figure 3

Statistics on differential expression. Volcano plot of differential gene expression (A：mRNAs；C：lncRNAs). Heat map of differential gene expression (B：mRNAs；D：lncRNAs).

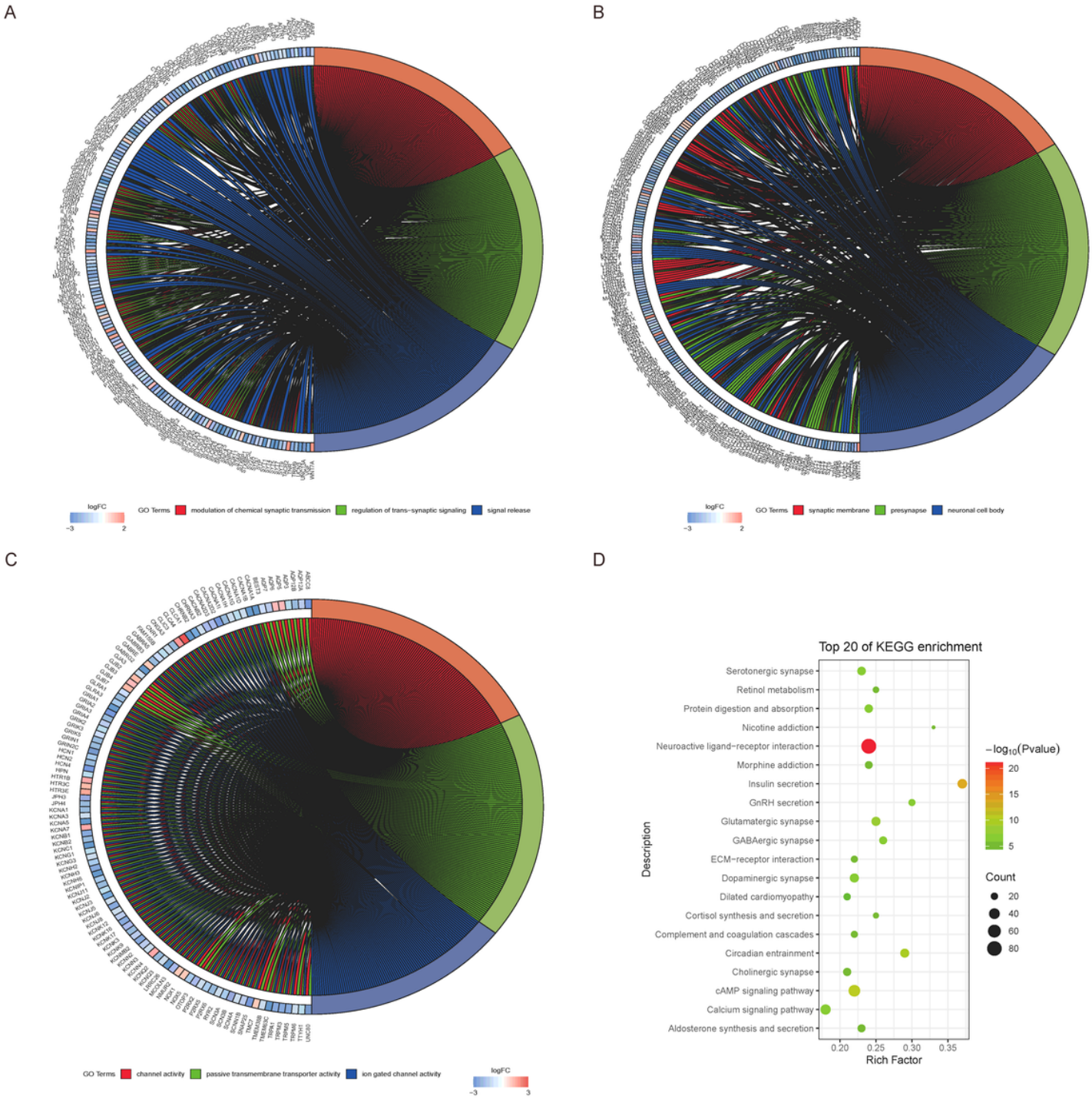

Figure 4

The results of functional enrichment analysis included three types of GO analysis (A, B, C) and KEGG analysis (D).

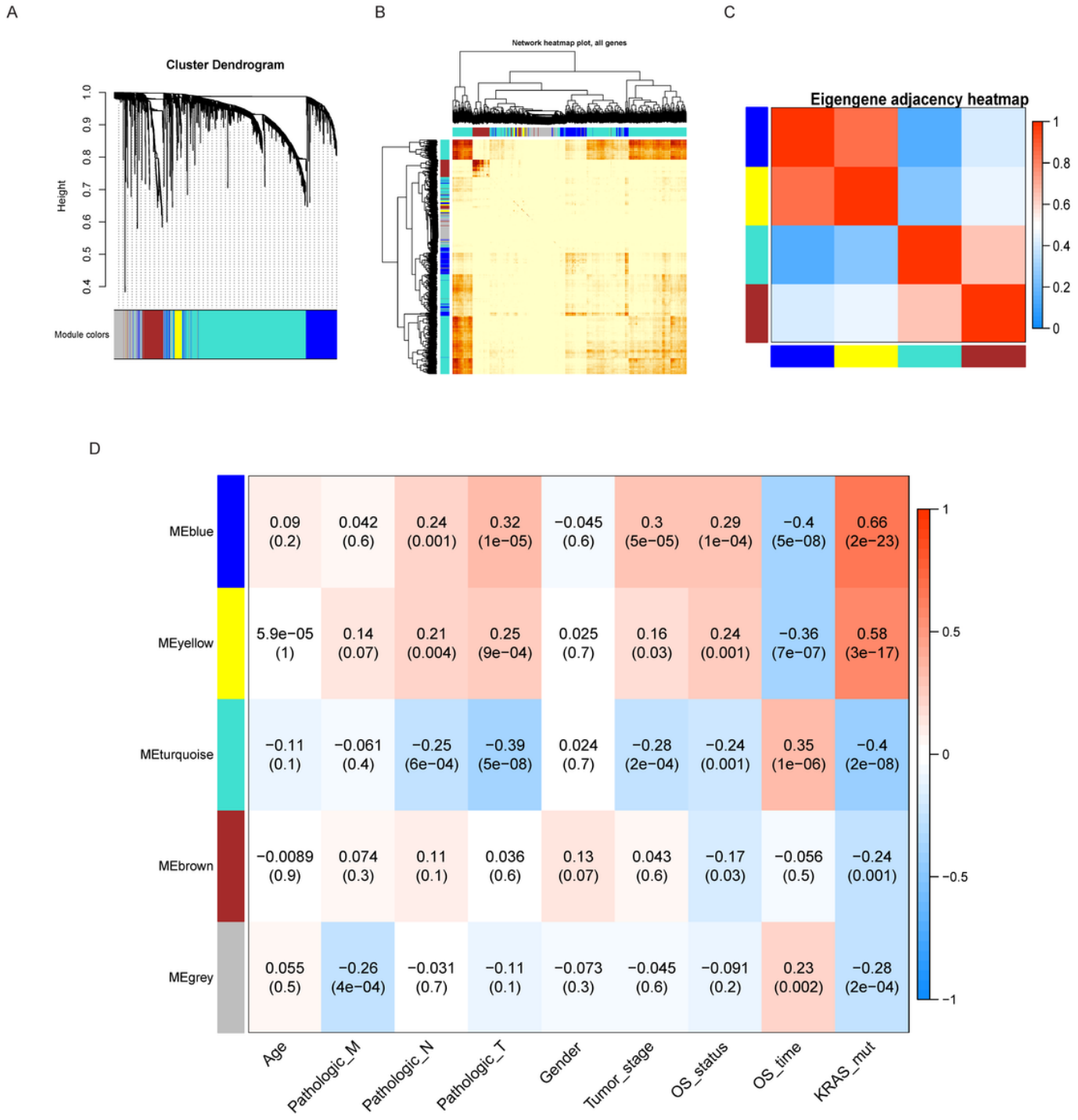

Figure 5

WGCNA analysis results. A. Clustering results; B. Weighted network heat map of all genes; C. Heat map of correlation between modules; D. Heat map of correlation between modules and phenotypes.

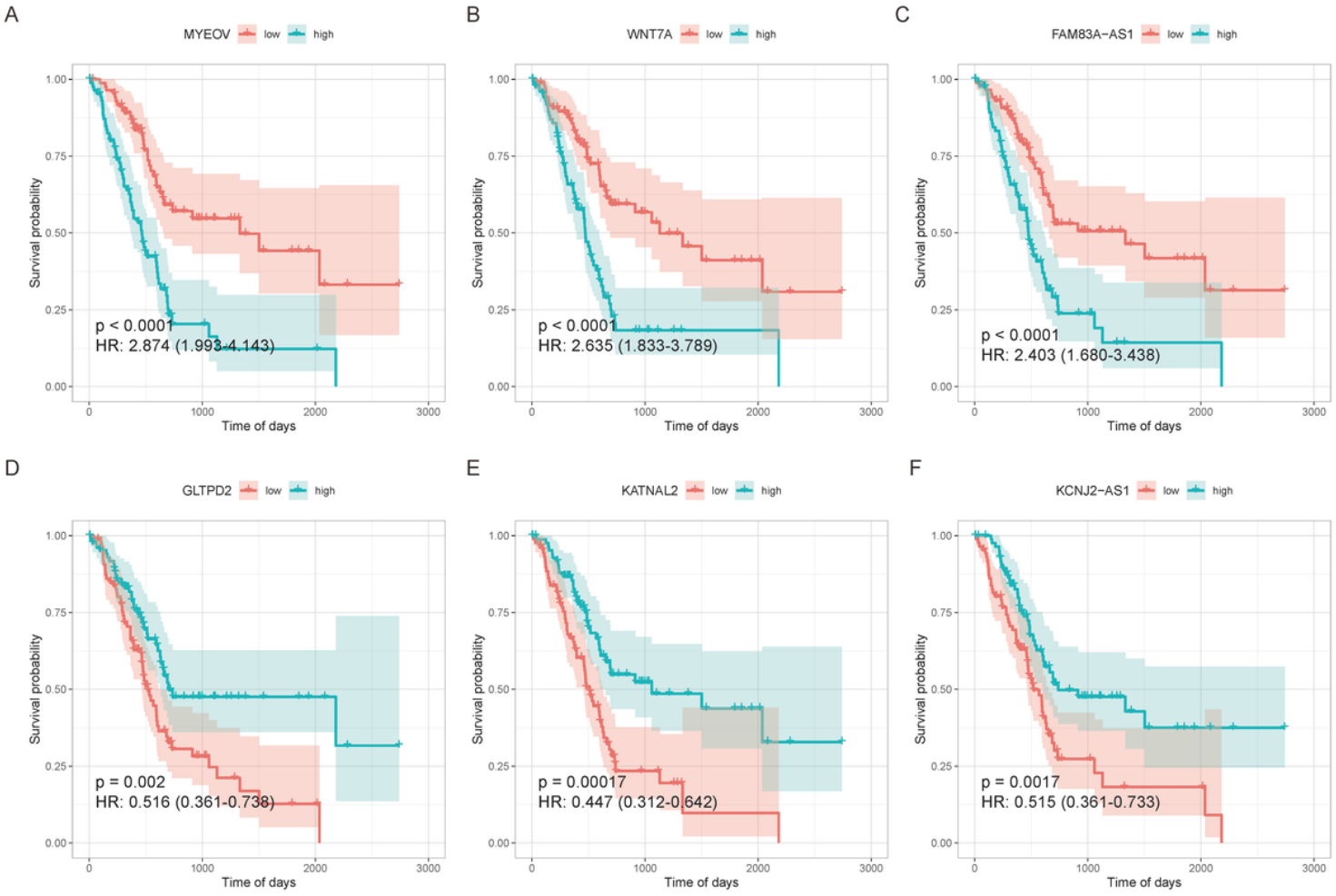

Figure 6

The KM curves of 6 genes with significant correlation to disease prognosis.

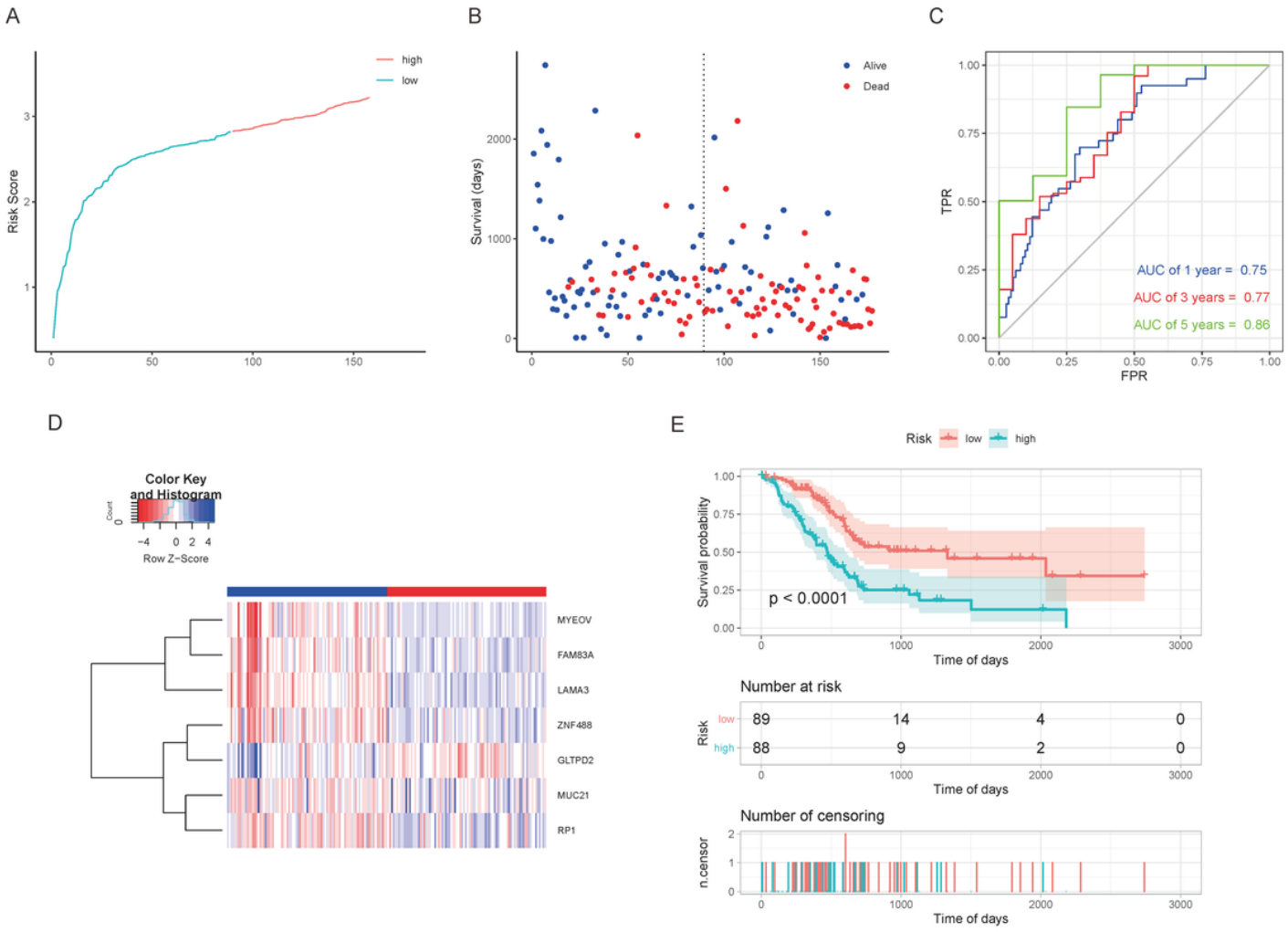

Figure 7

The mRNA prognostic risk model. A. Sample risk score curve; B. Scatter plot of sample survival time; C. Time-based ROC curve; D. Heat map of gene expression in the model; E. KM curve of high and low risk groups.

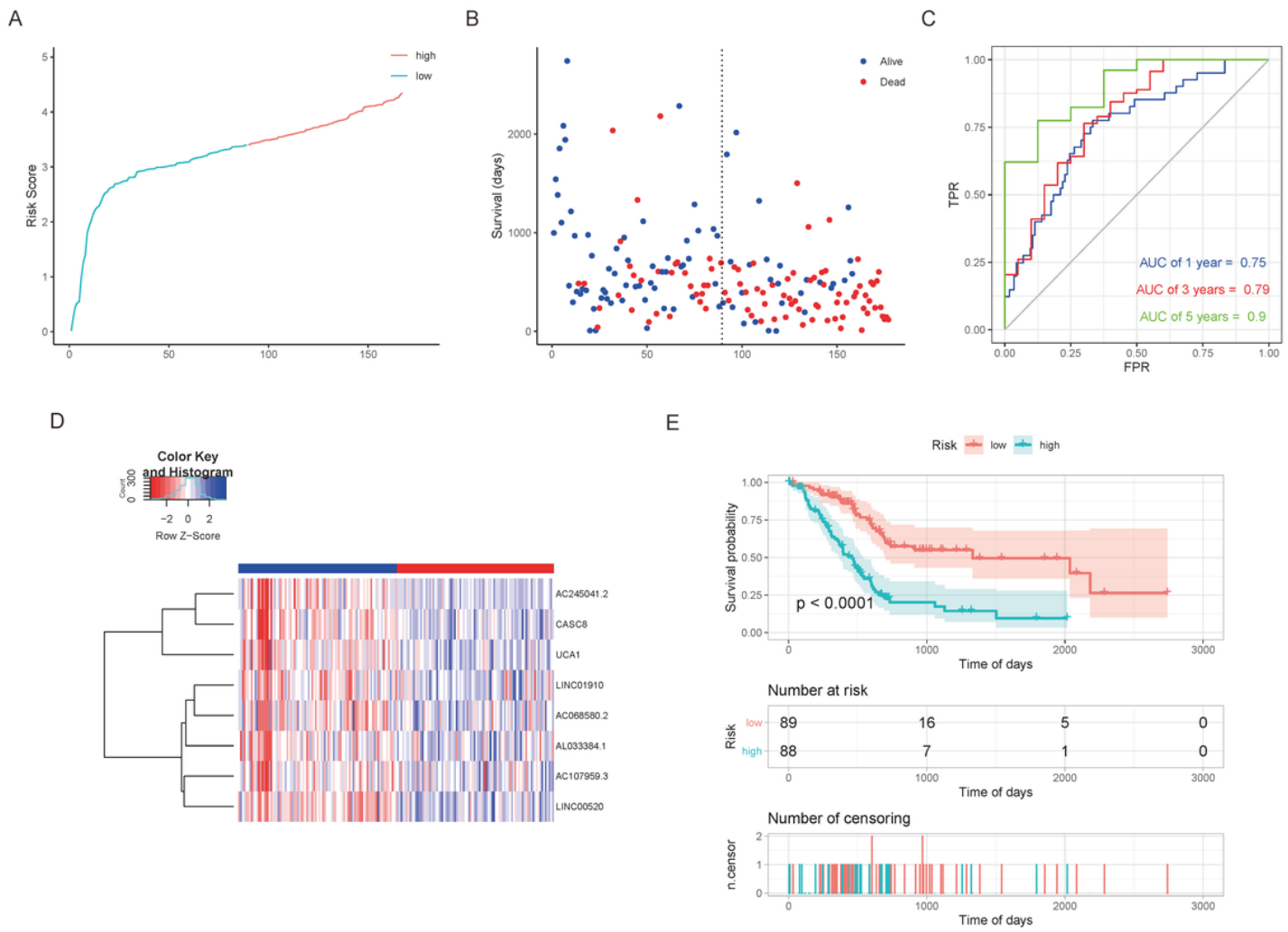

Figure 8

The lncRNA prognostic risk model: A. Sample risk score curve; B. Sample time-of-living scatter plot; C. Time-based ROC curve; D. Heat map of gene expression in the model; E. The KM curve of high and low risk group.

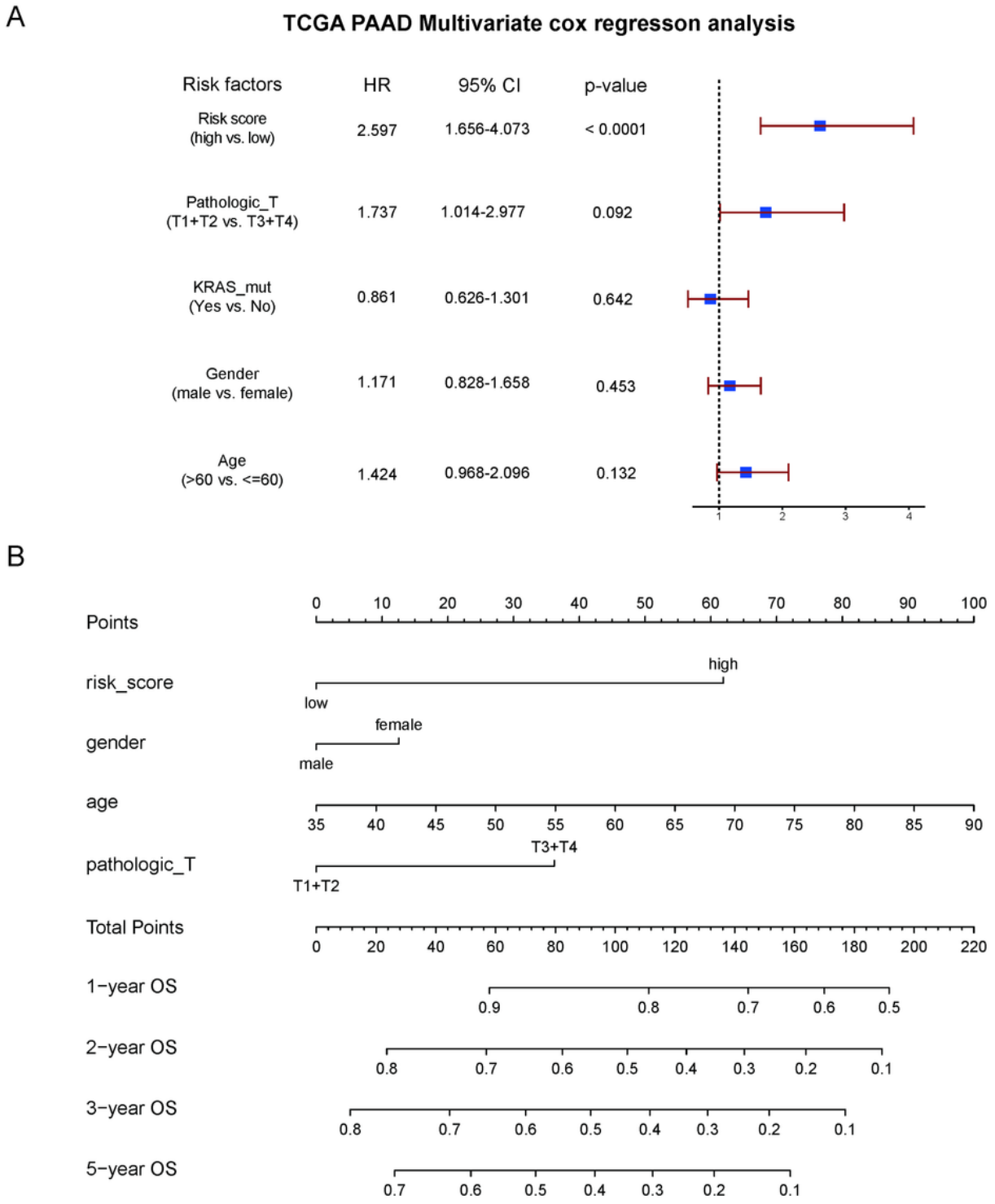

Figure 9

Results of multivariate cox regression analysis(A) and mRNA Nomogram (B).

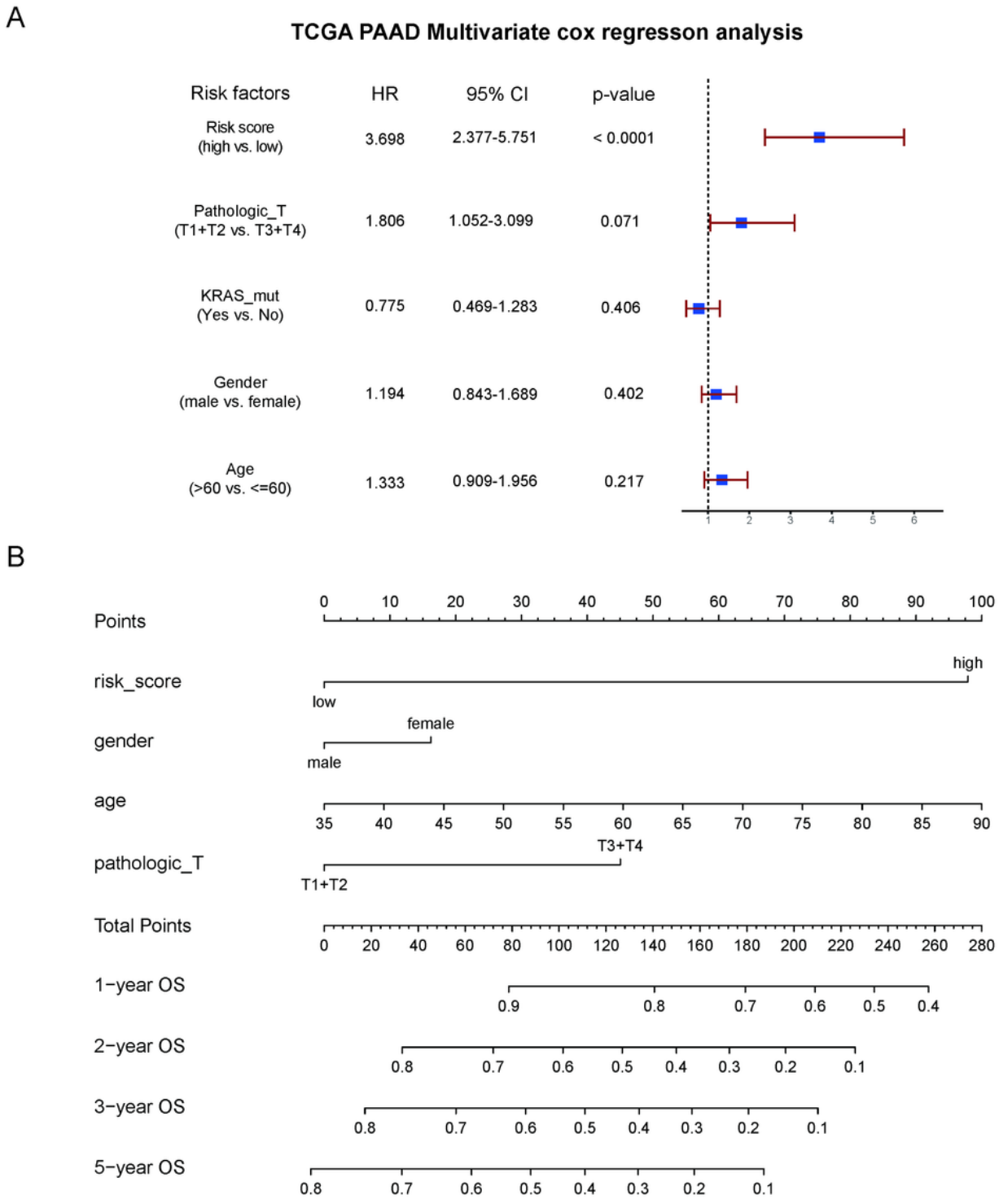

Figure 10

Results of multivariate cox regression analysis(A) and lncRNA Nomogram (B).

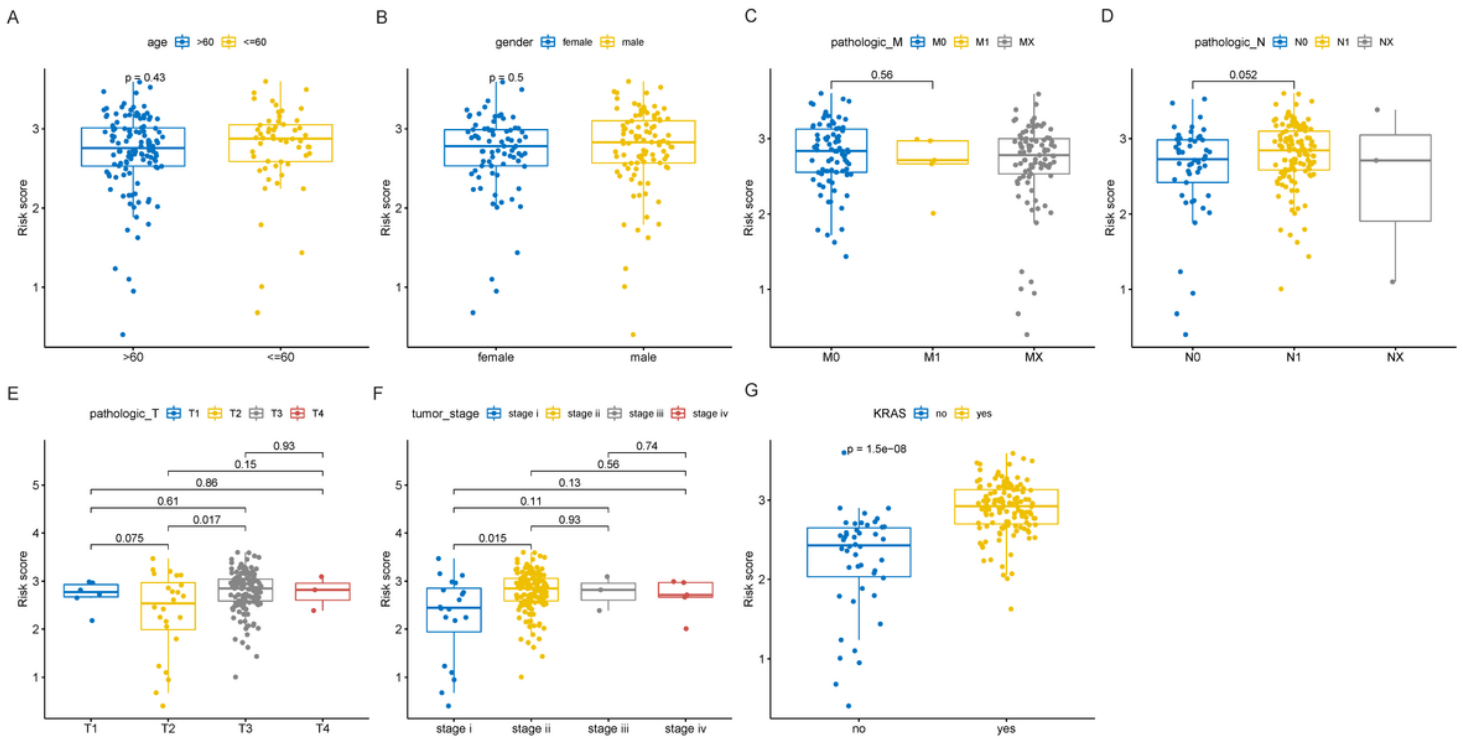

Figure 11

Distribution of risk score1 in clinical phenotypes.

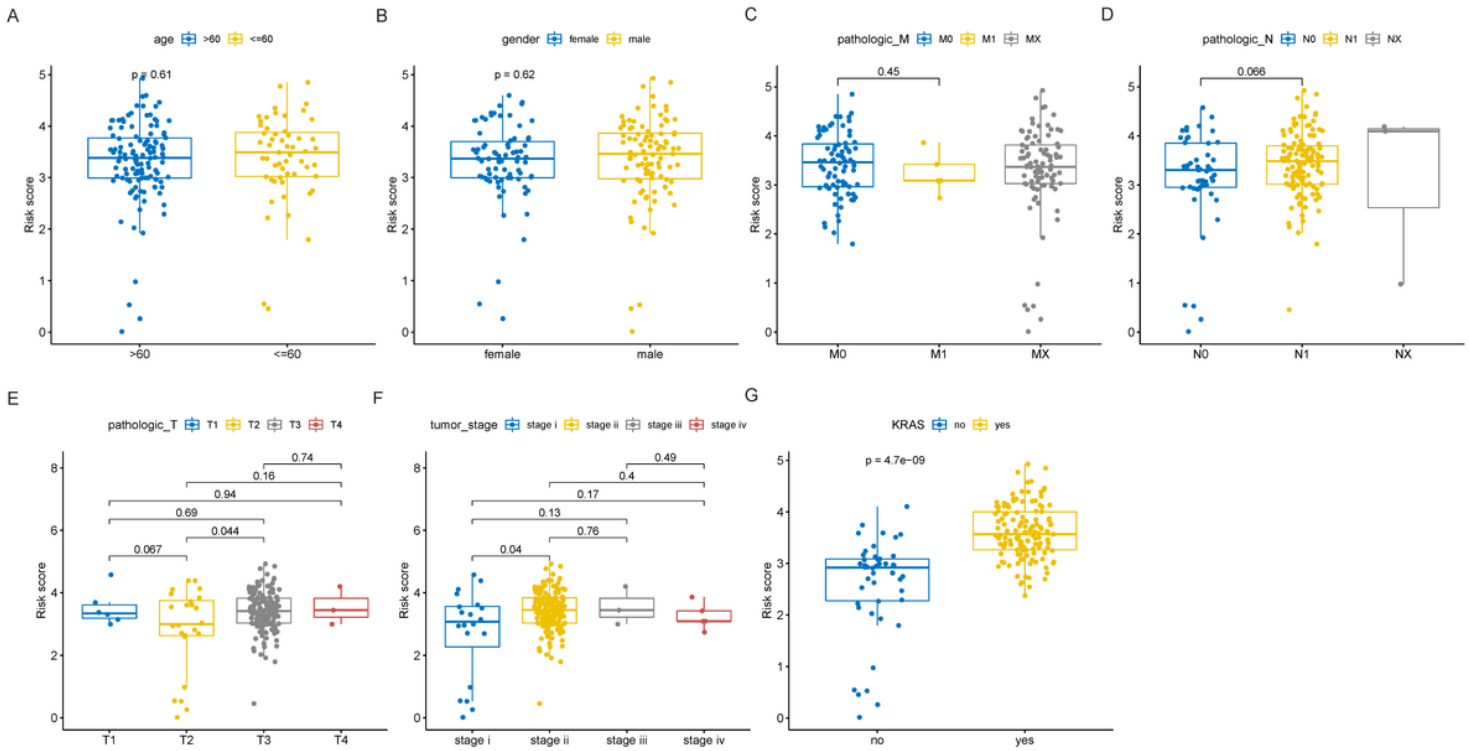

Figure 12

Distribution of risk score2 in clinical phenotypes.

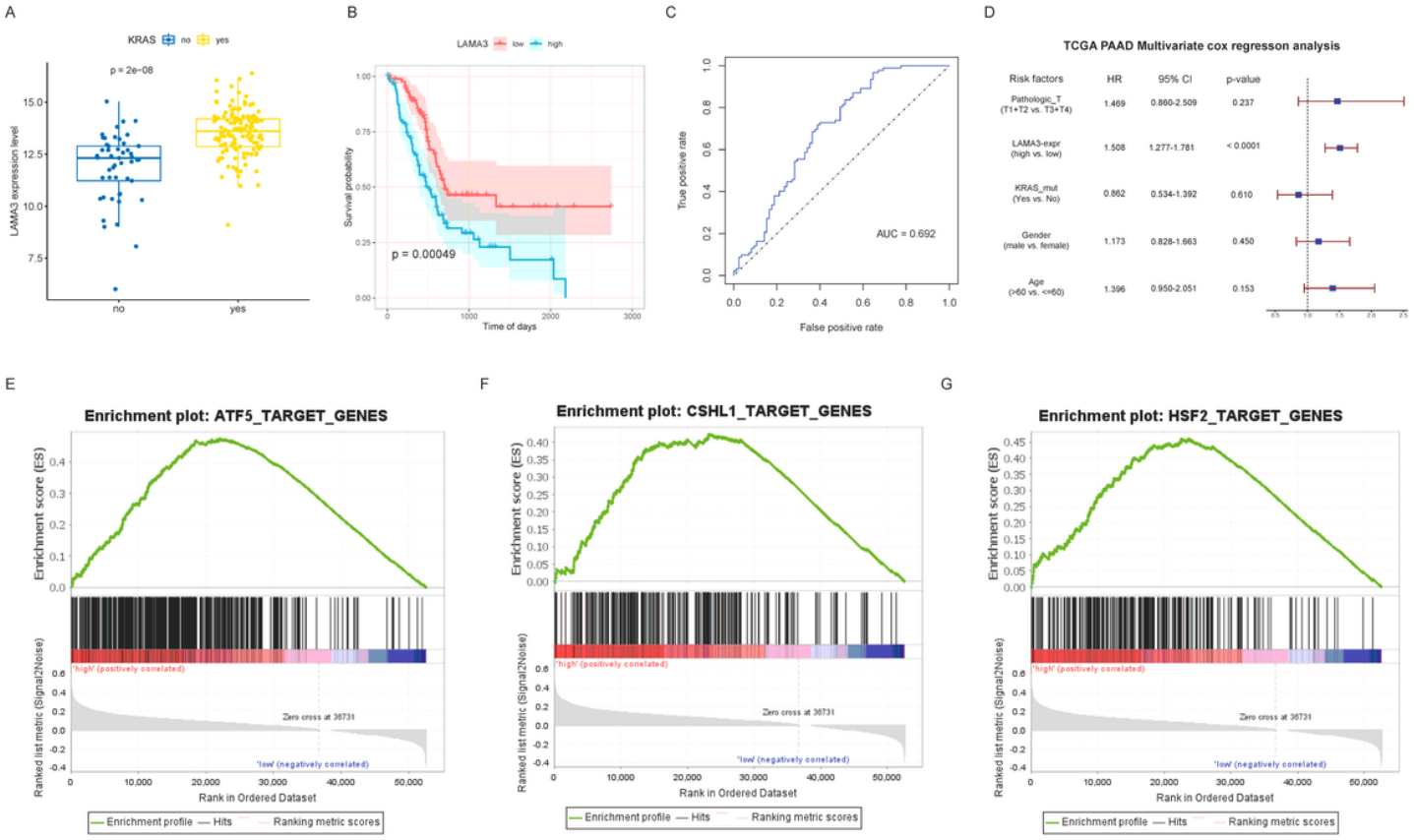

Figure 13

(A) Distribution of LAMA3 expression in KRAS mutation groups. (B) KM curves of LAMA3 expression groups. (C) ROC curve for prognosis based on LAMA3 expression. (D) Multivariate cox regression analysis based on LAMA3 expression. (E□F□G) GSEA analysis of LAMA3 expression.

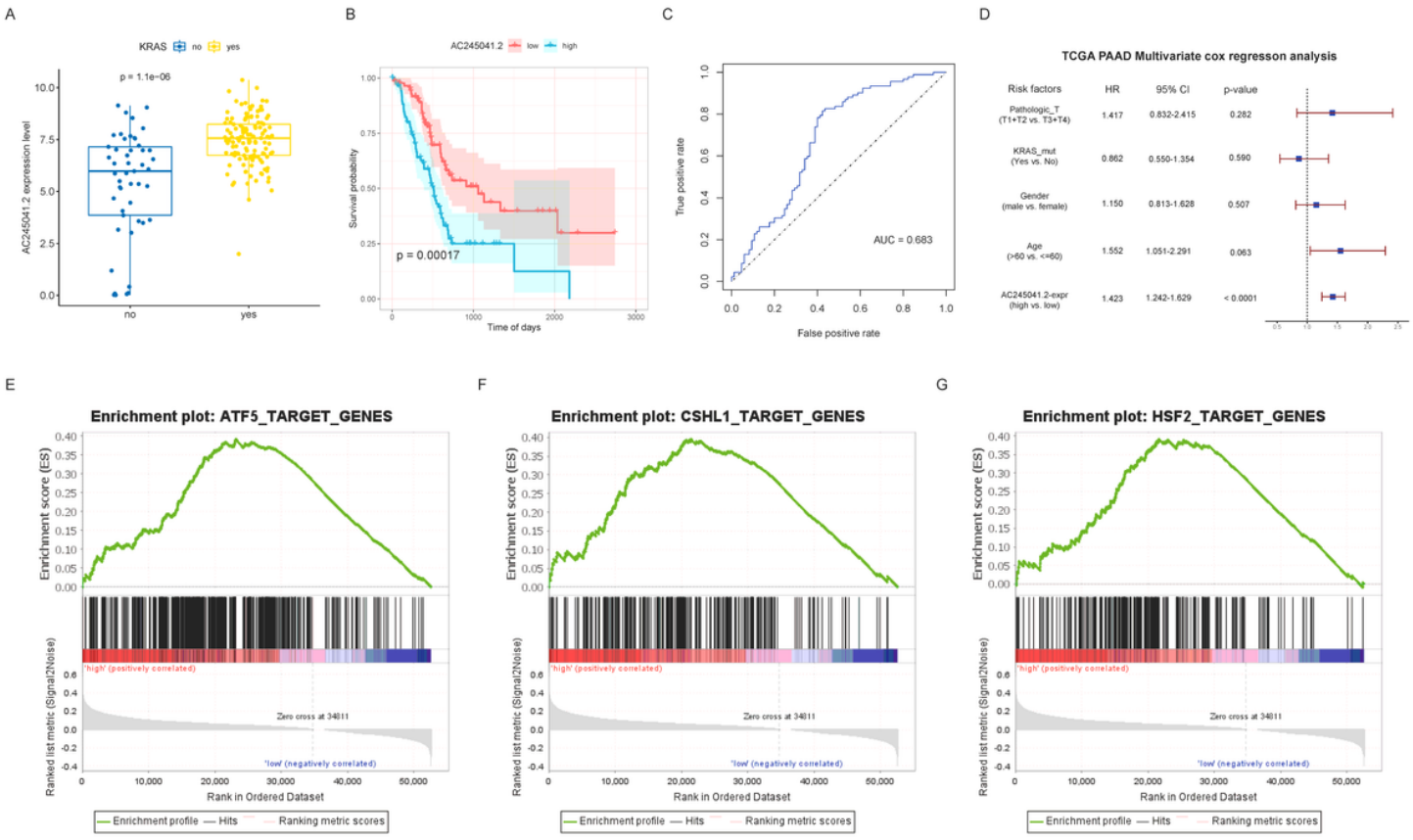

Figure 14

(A) Distribution of AC245041.2 expression in KRAS mutation groups. (B) KM curves of AC245041.2 expression groups. (C) ROC curve for prognosis based on AC245041.2 expression. (D) Multivariate cox regression analysis based on AC245041.2 expression. (E–F–G) GSEA analysis of AC245041.2 expression.

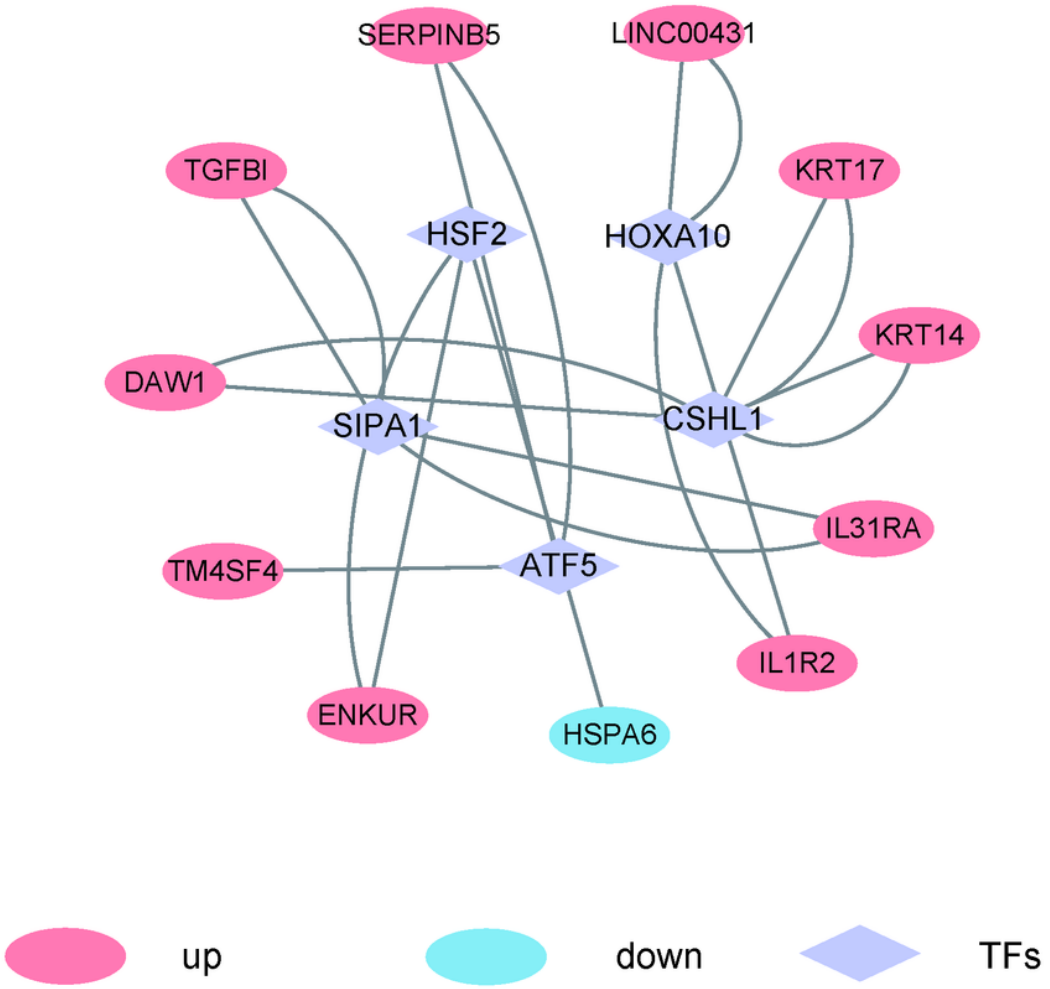

Figure 15

The mRNAs-lncRNA-TFs regulatory networks.