

Identification of a three-long noncoding RNA prognostic model involved competitive endogenous RNA in kidney renal clear cell carcinoma

Di Zhang

Beijing Chaoyang Hospital <https://orcid.org/0000-0001-7009-6296>

Xiaopeng Hu (✉ xiaopeng_hu@sina.com)

<https://orcid.org/0000-0003-4176-3605>

Song Zeng

Beijing Chaoyang Hospital

Primary research

Keywords: TCGA, KIRC, ceRNA, lncRNA prognostic model, nomogram

Posted Date: February 10th, 2020

DOI: <https://doi.org/10.21203/rs.2.23095/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published on July 17th, 2020. See the published version at <https://doi.org/10.1186/s12935-020-01423-4>.

Abstract

Background: Long noncoding RNA (lncRNA) is generally identified as competing endogenous RNA (ceRNA) that plays a vital role in the pathogenesis of kidney renal clear cell carcinoma (KIRC), the most common subtype of renal cell carcinoma with poor prognosis and unclear pathogenesis. This study established a novel ceRNA network and thus identified a three-lncRNA prognostic model in KIRC patients.

Methods: Differentially expressed genes (DEGs) were screened out from The Cancer Genome Atlas (TCGA) database. The lncATLAS was applied to determine the differentially expressed lncRNAs (DElncRNAs) located in the cytoplasm. The miRcode, miRDB, miRTarBase, and TargetScan databases were utilized to predict the interactions of DElncRNAs, DEmiRNAs, and DEMRNAs. Cytoscape was used to construct the ceRNA network. Then, a lncRNA prognostic model (LPM) was constructed based on ceRNA-related lncRNA that was significantly related to overall survival (OS), and its predictive ability was evaluated. Moreover, an LPM-based nomogram model was constructed. The significantly different expression of genes in the LPM was validated in an independent clinical cohort (N=21) by quantitative RT-PCR.

Results: A novel ceRNA regulatory network, including 73 lncRNAs, 8 miRNAs, and 21 mRNAs was constructed. Functional enrichment analysis indicated that integral components of membrane and PI3K-Akt signaling pathway represented the most significant GO terms and pathway, respectively. The LPM established based on three lncRNAs (MIAT, LINC00460, and LINC00443) of great prognostic value from the ceRNA network was proven to be independent of conventional clinical parameters to differentiate patients with low or high risk of poor survival, with the AUC of 1-, 5- and 10-year OS were 0.723, 0.714 and 0.826 respectively. Furthermore, the nomogram showed a better predictive value in KIRC patients than individual prognostic parameters. The expression of MIAT and LINC00460 was significantly upregulated in the KIRC samples, while the expression of LINC00443 was significantly downregulated compared with the adjacent normal samples in the clinical cohort, TCGA, and GTEX.

Conclusion: This LPM based on three-lncRNA could serve as an independent prognostic factor with a tremendous predictive ability for KIRC patients, and the identified novel ceRNA network may provide insight into the prognostic biomarkers and therapeutic targets of KIRC.

Background

Kidney Renal Clear Cell Carcinoma (KIRC), the most common and aggressive malignant subtype of renal cell carcinoma, has a poor prognosis and high mortality in an advanced stage due to the lack of useful biomarkers and treatments [1]. Currently, there is a multitude of established treatments for KIRC, such as surgical resection, nonspecific immune approach, targeted therapy against vascular endothelial growth factor, and novel immunotherapy agents. Despite these treatments, about 50% of KIRC patients develop metastases, and the 5-year survival rate of these patients is still lower than 10% [2]. Therefore, the

identification of useful prognostic biomarkers and therapeutic targets is urgently needed to predict more accurately and improve the outcome of KIRC patients.

Long noncoding RNA (lncRNA), broadly defined as noncoding RNA molecules longer than 200 nucleotides, has gained emerging attention in cancer biology due to their direct and indirect regulatory roles [3]. It has been reported to be aberrantly expressed in a broad spectrum of tumors, leading tumor initiation and progression [4]. Hence, they may serve as promising a new type of biomarkers for tumor diagnosis, prognosis, even in targeted gene therapy [5]. More recently, extensive research has explored the lncRNA expression profiling in the KIRC with the development of sequencing technology [6–8]. The determination of their interaction with other molecules and functional analysis is also widely investigated in recent years [9]. Notably, the subcellular localization of lncRNAs holds valuable clues to their molecular function [10]. In the cell nucleus, lncRNA could modulate nuclear functions, such as transcription, chromatin regulation, and variable splicing [11]. While in the cytoplasm, lncRNA could modulate mRNA mainly through the competitive endogenous RNA (ceRNA) regulation mechanism, according to the ceRNA hypothesis proposed by Salmena et al. in which lncRNA could compete with miRNA as a natural sponge and therefore indirectly regulate mRNA expression [12, 13]. Since proposed, numerous studies have investigated and validated that this lncRNA-miRNA-mRNA regulatory mechanism participates in tumor occurrence, progression, and metastasis and the potential prognosis of multiple cancers [9]. However, the overall regulatory functions of the lncRNA-miRNA-mRNA ceRNA network remain unclear, and the predictive accuracy of the prognostic model based on multiple lncRNAs' expression is virtually worthy of exploring in KIRC patients.

In this study, we constructed a ceRNA network in KIRC, in which the subcellular localization of lncRNA was restricted to be cytoplasm, to elucidate the potential interaction of differentially expressed lncRNAs (DElncRNAs), DEmiRNAs and DE mRNAs, and the vital role of the network in KIRC. Subsequently, we identified the potential prognostic values of lncRNAs included in the ceRNA network and confirmed a lncRNA prognostic model (LPM) signature, which was established based on three lncRNAs of great prognostic value for overall survival (OS), as an independent prognostic biomarker. Then, we demonstrated a nomogram with better prediction value in KIRC patients than individual prognostic parameters. In the end, an independent clinical cohort was used to verify the significantly different expression of genes in the LPM by RT-PCR.

Methods

Data source and processing

The level 3 gene expression files and the corresponding clinical information for 530 KIRC patients, and miRNA expression files for 516 KIRC patients were retrieved from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>) (up to October 10, 2019), sequencing by Illumina HiSeq RNA-Seq and Illumina HiSeq miRNA-Seq platforms, respectively. Among these KIRC patients, a total of 539 KIRC samples and 72 adjacent normal samples in RNA sequence data, 545 KIRC samples and 71

adjacent normal samples in miRNA sequence data were subjected to subsequent analyses. The annotation data (antisense, lincRNA, and sense_intronic/ sense overlapping, lincRNAs, 3' overlapping ncRNAs, processed transcripts, antisense, and sense intronic) of probes of the TCGA RNA sequence data was recognized as lincRNA, and annotation data (protein-coding) as mRNA by using the GENCODE database (<https://www.genencodegenes.org/>). Then, the gene symbols were annotated based on the Homo_sapiens.GRCh38.84.chr.gtf file, which was downloaded from the Ensembl database (<https://asia.ensembl.org/index.html>).

Identification of differentially expressed genes (DEGs)

We compared the KIRC samples and adjacent normal samples to identify DEGs by using DESeq2 R package (Version 1.27.19; <http://www.bioconductor.org/packages/devel/bioc/html/DESeq2.html>) with a rigorous threshold as $|\log_2\text{-fold change (FC)}| > 2.0$ and $\text{FDR} < 0.01$ [14]. Then a heat map and volcano plot were drawn by using the heatmap R package (Version 1.0.1; <https://www.rdocumentation.org/packages/pheatmap>) and ggpubr R package (Version 0.2.4; <https://www.rdocumentation.org/packages/ggpubr>) in R software (Version 3.6.0; <https://www.r-project.org/>), to visualize the hierarchical clustering analysis of the identified DEGs.

Construction of the ceRNA network

The lincAtlas database (<http://lincatlas.org.eu/>) was used to identify the DElincRNAs located in the cytoplasm [15]. Then the DEmiRNAs which potentially interacted with DElincRNAs located in the cytoplasm were predicted using the miRcode (<http://www.mircode.org/>), a comprehensive searchable map of putative microRNA target sites in the long noncoding transcriptome [16]. Subsequently, the target DErnAs of DEmiRNA were predicted using miRDB (<http://mirdb.org/>), miRTarBase (<http://mirtarbase.mbc.nctu.edu.tw/php/index.php>) and TargetScan (http://www.targetscan.org/vert_72/) databases [17–19]. After that, Cytoscape software (Version 3.7.2; <http://www.cytoscape.org/>) was utilized to visualize and construct the ceRNA network [20].

Functional enrichment analysis

The pathway and functional enrichment analysis were carried out by utilizing KO-Based Annotation System (KOBAS) (Version 3.0; <http://kobas.cbi.pku.edu.cn/>) and the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Version: 6.8; <https://david.ncifcrf.gov/>), to investigate the potential biological implications of the ceRNA network [21, 22]. Significant GO terms and pathways were visualized by the GOplot (Version 1.0.2; <https://cran.r-project.org/web/packages/GOplot/index.html>) and ggalluvial (Version 0.9.1; <https://www.rdocumentation.org/packages/ggalluvial>) R packages, respectively.

Construction and validation of a lincRNA-related prognostic model

Among 530 KIRC patients with RNA-sequencing data and clinical information, 514 KIRC patients were subjected to subsequent analyses after excluded according to the following criteria: (1) patients without

survival information including survival time and status; (2) patients without complete lncRNA expression data; (3) patients who did not meet endpoint with following time less than 30 days. The expression files of DElncRNAs involved in the ceRNA network from 514 KIRC patients were analyzed through univariate Cox regression analysis in which genes were regarded as significant at $P < 0.001$, to identify the prognostic value of the DElncRNAs for OS. Then, the least absolute shrinkage and selection operator (LASSO) model with L1-penalty, performed by using the glmnet R package (Version 2.0–16; <https://www.rdocumentation.org/packages/glmnet>), was utilized to further select crucial lncRNAs from the prognostic DElncRNAs. In this method, a sub-selection of lncRNAs involved in KIRC patient prognosis was identified by shrinkage of the regression coefficient. Eventually, quite a few indicators with a weight of nonzero remained, and most of the potential indicators were shrunk to zero. In this process, we subsampled the data set 1000 times and selected the lncRNAs that were repeated > 900 times [23]. Finally, a risk-score based LPM was established using the regression coefficients from the multivariate Cox regression analysis in which genes were regarded as significant at $P < 0.01$. The formula of the risk score was constructed as follows:

$$\text{Risk score} = \sum_{i=1}^N (\beta * x).$$

β stands for the regression coefficient of genes, X represents the expression level of genes, and N is the number of significant genes derived from the multivariate Cox regression analysis. The univariate and multivariate Cox regression analysis was performed utilizing the survival R package (Version 3.1-8; <https://www.rdocumentation.org/packages/survival>). Subsequently, the X-tile 3.6.1 software was applied to calculate the optimal cutoff to classify the KIRC patients into high- and low-risk groups [24]. Risk heatmap applying pheatmap R package (Version 1.0.12; <https://www.rdocumentation.org/packages/pheatmap>) was employed to cluster the expression files of core lncRNAs, which constructed the LPM, in the high- and low-risk groups. Then, the Kaplan–Meier (K–M) survival analysis and time-dependent receiver operating characteristic (ROC) curves were used to evaluate the ability of the LPM to predict OS and disease-specific survival (DSS) by utilizing the survminer (Version 0.4.3; <https://www.rdocumentation.org/packages/survminer>) and survivalROC (Version 1.0.3; <https://www.rdocumentation.org/packages/survivalROC>) R package, respectively.

Independence of the LPM from traditional clinical features

514 KIRC patients with complete lncRNA expression data, survival information, and complete clinical information, including age, gender, pathologic stage, and histologic grade, were subjected to subsequent univariate and multivariate Cox regression analyses, to assess the independent prognostic ability of the LPM for KIRC patients. To better visualize the prognostic value of risk score and clinical features, the forest plot was performed by using the ggplot2 R package (Version 3.2.1; <https://www.rdocumentation.org/packages/ggplot2>).

Construction and validation of the nomogram

To further determine the predictive accuracy of model efficiency for 1-, 5-, 10-year, we constructed a novel nomogram, contained significant clinical features and calibration plots, based on the results of the multivariate Cox analysis utilizing the rms R package (Version 5.1–4; <https://cran.r-project.org/web/packages/rms/index.html>). The concordance index (C-index) was applied to evaluate the discrimination of the nomogram, and it was corrected by a bootstrap method with 1000 resamples [25]. Besides, the C-index and time-dependent ROC curves were performed to compare the predictive accuracies of the nomogram and individual prognostic factors. Besides, decision curve analysis (DCA) was performed to assess the clinical utility of the nomogram by comparing the benefits of different models.

Gene set enrichment analysis (GSEA)

GSEA (Version 4.0; <http://software.broadinstitute.org/gsea/index.jsp>) was performed between high- and low-risk groups to identify the potential biological function of LPM[26]. An annotated gene set file (c5.bp.v7.0.entrez.gmt) was chosen as the reference gene set. The threshold was set at levels of $|NES| > 2$ and $P < 0.01$.

Clinical samples and Quantitative RT-PCR

In the clinical validation set, a total of 21 KIRC patients with KIRC samples and adjacent normal samples were selected according to the following criteria: (1) patients treated in the Beijing Chao-Yang Hospital; (2) patients who did not undergo treatment before surgery. Two individual experienced pathologists confirmed the final diagnosis of samples through identifying the morphology of the samples stained with H&E. This study was carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki). All patients signed the informed consent, and this study was approved by the ethics committees of Beijing Chao-Yang Hospital.

Total RNA of samples was extracted using the Trizol method, and then we synthesized cDNA via reverse transcription using the HiScript III RT SuperMix Kit (R323-01, Vazyme, China). The expression levels of lncRNAs were quantitated using the AceQ qPCR SYBR Green Master Mix (R323-01, Vazyme, China) by the ABI 7500 real-time PCR system (Applied Biosystems, Foster City, CA, USA). The primers used in this study are listed in Table S1. Target lncRNA levels were normalized against GAPDH standards and calculated using the $2^{-\Delta\Delta Ct}$ method.

Validation of lncRNAs expression

The expression levels of the three lncRNAs were verified in KIRC patients with paired KIRC samples and adjacent normal samples from the Beijing Chao-Yang cohort (N = 21) and TCGA cohort (N = 72), respectively, by using Wilcoxon signed-rank test. Additionally, the different expression analysis was further performed in KIRC samples (N = 523) from TCGA and normal samples (N = 100) from the match TCGA normal and Genotype-Tissue Expression (GTEx) data by utilizing Gene Expression Profiling Interactive Analysis (GEPIA2) (<http://gepia2.cancer-pku.cn/#index>) [27]. $P < 0.05$ was considered significant, and all statistical tests were two-sided.

Results

Identification of differentially expressed genes between KIRC samples and adjacent normal samples

A total of 539 KIRC samples and 72 adjacent normal samples were utilized to screen DElncRNAs and DEmRNAs; 545 KIRC samples and 71 adjacent normal samples were utilized to screen DEmiRNAs. The DESeq2 R package was performed to identify differentially expressed genes with a strict cutoff threshold of $|\log_2 FC| > 2$ and an adjusted $P < 0.01$. Compared with normal samples, 2015 DElncRNAs, 47 DEmiRNAs, and 2314 DEmRNAs were differentially expressed, among which 1461 lncRNAs, 19 miRNAs, and 1511 mRNAs were upregulated as well as 554 lncRNAs, 28 miRNAs and 803 mRNAs were downregulated (Table S2). The heat maps and volcano plots of DEGs between KIRC samples and adjacent normal samples were shown in Fig. 1.

Construction of the ceRNA network

Considering the nuclear-cytoplasmic localization of lncRNAs plays a vital role in its molecular function, we firstly confirmed the subcellular localization of the 2015 DElncRNAs by utilizing the lncATLAS database, and excluded 385 DElncRNAs which were located only in the nucleus because the endogenous competition role of lncRNAs is mainly exhibited in the cytoplasm. The detailed distribution information for the DElncRNAs was shown in Table S3. Then the remained DElncRNAs were put into the miRcode database to identify the potential miRNAs targeting lncRNAs. However, only 12 out of predicted miRNAs were selected after taking the intersection with 47 DEmiRNAs. We then utilized the databases of miRDB, miRTarBase, and TargetScan to identify the downstream target mRNAs with reference to the 12 DEmiRNAs. In addition, we selected potential mRNAs that only shared by all three databases to enhance the veracity of the prediction. The results showed that 21 out of 2314 DEmRNAs were identified. Finally, a total of 73 DElncRNAs, 8 DEmiRNAs, and 21 DEmRNAs were eventually incorporated into the KIRC-associated ceRNA regulatory network by applying Cytoscape software (Fig. 2A, Table S4).

To further reveal the biological functions and pathways associated with the ceRNA network, GO and KEGG enrichment analysis was performed via KOBAS and DAVID. The results of GO analysis indicated that the DEmRNAs involved in the ceRNA network were mainly enriched in the "Integral component of membrane", "Extracellular region" and "Extracellular space" (Fig. 2B). Moreover, the results from KEGG analysis showed that the DEmRNAs were particularly enriched in "ECM-receptor interaction", "Chemokine signaling pathway", "PI3K-Akt signaling pathway" and Neurotrophin signaling pathway (Fig. 2C).

Construction of an LPM and evaluation of its predictive ability

To consider whether those DElncRNAs involved in the ceRNA network were closely related to the OS of KIRC patients in the TCGA cohort, we performed the univariate Cox regression to identify the prognostic

value of the DElncRNAs for OS. The result indicated that 17 of the 73 DElncRNAs were significantly related to OS (Table S5). To further find crucial lncRNAs from the prognostic DElncRNAs, we applied LASSO estimation, and selected 8 lncRNAs which appeared > 900 times out of 1000 repetitions (Fig. 3A). Subsequently, multivariate Cox regression analysis was utilized to select lncRNAs with the best prognostic value and calculate their relative coefficients, to further establish a risk-score based LPM. Finally, we constructed the LPM to predict patient survival with the risk score of each patient calculated as follows: risk score = $(0.13383 \times \text{expression level of LINC00460}) + (-0.33667 \times \text{expression level of LINC00443}) + (0.15751 \times \text{expression level of MIAT})$ (Fig. 3B). Furthermore, we applied the X-tile software to find the optimal cutoff value of risk scores, and patients with risk scores greater than 2.08 (n = 65) were classified into the high-risk group, while those with risk scores less than or equal to 2.08 (n = 449) were allocated to the low-risk group. The risk score distribution and lncRNA expression data are shown in Fig. 3C. The K-M survival analysis indicated that the high-risk patients had a shorter OS than the low-risk patients (Fig. 3D). Additionally, the high-risk group showed a 3.768-fold higher risk [95% confidence interval (CI): 2.254–6.299, $P < 0.001$] than the low-risk group. The time-dependent ROC curve analysis validated the great prognostic value of the LPM (Fig. 3E). The area under the ROC curve (AUC) of the LPM for OS was 0.723 at 1 year, 0.714 at 5 years, and 0.826 at 10 years. Besides, the high-risk patients had a shorter DSS than the low-risk patients with a 5.561-fold higher risk (95% CI: 2.929–10.560, $P < 0.001$), and the AUC of the LPM for DSS was 0.723 at 1 year, 0.770 at 5 years and 0.793 at 10 years (Fig. 3F, G).

High risk indicated an enhanced immune response

To explore the biological pathways associated with the LPM, GSEA was conducted between the 65 high-risk and 449 low-risk KIRC patients in the TCGA KIRC cohort. The result revealed that the high-risk patients were significantly related to 223 biological processes (Table S6), in which the top 3 immune processes were HUMORAL_IMMUNE_RESPONSE_MEDIATED_BY_CIRCULATING_IMMUNOGLOBULIN (NES = 3.512, size = 136), B_CELL_MEDIATED_IMMUNITY (NES = 3.364, size = 199) and REGULATION_OF_HUMORAL_IMMUNE_RESPONSE (NES = 3.172, size = 124) ($P < 0.01$) (Fig. 4A). According to the results, the poor prognosis of high-risk group patients was strongly related to the immune response, which indicated an intense immune phenotype.

Validation of lncRNAs in the clinical cohort, TCGA, and GTEx

According to the above analysis in the TCGA KIRC cohort, the three lncRNAs (LINC00460, LINC00443, MIAT) involved in the ceRNA network could serve as potential prognostic predictors for KIRC patients. To verify the validity and reliability of the results, the expression of the three lncRNAs were analyzed between 21 diagnosed KIRC samples and 21 adjacent normal samples through RT-PCR. As shown in Fig. 4B, the results demonstrated that the expression of MIAT and LINC00460 was significantly upregulated in the KIRC samples while the expression of LINC00443 was significantly downregulated compared with the adjacent normal samples. In addition, we performed the difference analysis of three-lncRNA expression

by using the TCGA paired samples and GEPIA2, to verify the expression of MIAT, LINC00460, LINC00443 further. A consistent finding with that of the clinical KIRC cohort was elucidated (Fig. 4C, D).

Stratification analysis of OS for the LPM

Stratification analysis was performed to determine whether the prognostic value of the LPM would remain stable in different subgroups. Therefore, patients in the TCGA KIRC cohort were classified into two groups according to age, sex, tumor grade, and tumor stage, respectively. As shown in Fig. 5, patients in the high-risk group showed worse survival compared to those in the low-risk group in patients with grade low or grade median and high tumors, stage I and II or stage III and IV tumors, younger or older, and male or female patients. Besides, the LPM still remained a stable and great predictive ability for KIRC patients in distinct subgroups.

LPM is independent of traditional clinical features for KIRC Patients

To identify whether the LPM is an independent clinical prognostic factor for KIRC patients, we firstly performed the univariate Cox regression analysis and demonstrated that the LPM was significantly associated with OS [Hazard ratio (HR): 3.809, 95% CI: 2.720–5.330, $P < 0.001$; Fig. 6A]. Then clinical characteristics, including gender, age, pathologic stage, and histological grade were adjusting by multivariate Cox regression analysis, and the result indicated that the LPM remained an independent prognostic factor with an HR of 2.020 in the TCGA KIRC cohort (95% CI: 1.387–2.940, $P < 0.001$; Fig. 6B).

Construction and validation of an LPM-based nomogram model

To provide a clinically related quantitative approach to predicting the prognosis of KIRC patients, a nomogram that integrated the risk score and independent clinical prognostic factors (age, histologic grade, and tumor stage) was established (Fig. 6C). In this nomogram based on multivariate Cox analysis, a point scale was used to assign points to these variables. A straight line was drawn upward to determine the points for the variables, and the grand total of the points assigned for each variable was rescaled to a range from 0 to 100. The points of the variables were accumulated and recorded as the total points. The probability of KIRC patient survival at 1, 5, and 10 years was determined by drawing a vertical line from the total point axis straight downward to the outcome axis. For example, a KIRC patient with risk score 2 (22 points), age ≥ 65 (21 points), grade high (22 points), and stage ≥ 3 (25 points) received a total point score of 90. The probability of 1-year survival was determined by drawing a vertical line from the total point axis at a value of 90 straight downward to the outcome axis, which showed that the probability of 1-year survival was 80%. The LPM was found to contribute the most risk points (ranging from 0 to 100) compared with the other clinical information, which was consistent with our Cox multivariate regression results. Based on the total nomogram scores, we divided the patients into high-risk group and low-risk group with an optimal cutoff point 66, and the high-risk patients had a shorter OS than the low-risk

patients with a 5.574-fold higher risk (95% CI: 3.852–8.064, $P < 0.001$; Fig. 6D). We then used the time-dependent ROC curve analysis to compare the predictive accuracy between the nomogram model and individual predictors, including risk score, age, grade, and stage (Figs. 6E). The nomogram model suggested higher prognostic accuracy at 1-and 3-year OS with a larger AUC (Figure. 6F).

Moreover, we compared the C-index between the nomogram and other individual predictors. The nomogram had a higher mean C-index (0.772) than other predictors (0.551 to 0.676) (Fig. 6G), which meant that the nomogram had a better predictive ability than LPM and general clinical characteristics in prognosis prediction for KIRC patients, further validating the robust prognostic value of the nomogram. The bias-corrected line in the calibration plot was found to be close to the ideal curve (the 45-degree line), which indicated good agreement between the prediction and the observation (Fig. 6H). Ultimately, we attempted to determine the clinical benefit of the nomogram model and the corresponding scope of application via DCA. Compared with other individual predictors, the nomogram model revealed an enhanced net benefit with wider threshold probabilities and offered the best clinical utility (1-year OS: Fig. 6I; 5-year OS: Fig. 6J; 10-year OS: Fig. 6K). In sum, these findings suggest that the nomogram was a better model for predicting short-term or long-term survival in KIRC patients than individual prognostic factors.

Discussion

As the most common and lethal subtype of renal carcinoma, KIRC is driven by distinct driver gene mutations and complex molecular alterations [28]. Though many effective treatments have been developed for KIRC, the unsatisfied survival rate and intolerant of chemotherapy make it an emerging need for new therapeutic targets and prognostic biomarkers to improve the clinical outcomes of KIRC patients in the future [29]. Previous research has demonstrated that lncRNAs located in the cytoplasm could regulate mRNA transcription indispensably, primarily through ceRNA regulatory networks, making these attractive molecules targets and prognostic biomarkers [30]. However, few studies have investigated the specific functions and prognostic value of lncRNAs involved in the ceRNA regulatory network in KIRC, especially lncRNAs located in the cytoplasm.

In this study, we established a lncRNA-miRNA-mRNA ceRNA regulatory network and constructed a novel three-lncRNA-based LPM that could identify KIRC patients who had a high risk of poor prognosis. Therefore, it is feasible to divide the KIRC patients into different subgroups with a particular risk score, and such patient stratification could contribute to more individualized therapy for patients. The lncRNAs (MIAT, LINC00460, and LINC00443) that constitute the LPM could be regarded as individual targets, and they may provide better performance in combination, depending on their prognostic significance.

An increasing number of researches have found that tumor initiation and progression can be largely represented by DEGs [31, 32]. Thus, we firstly identified DEGs as candidate genes for the ceRNA network by using the DESeq2 method, which has better statistical power in sensitivity and precision than edgeR and DESeq. To the best of our knowledge, our study is the first to use the DESeq2 method in constructing

the ceRNA network in KIRC, resulting in an inconsistent finding in the DEGs screen with other studies [33, 34]. Since the lncRNA-miRNA-mRNA interaction in the ceRNA network only presented in the cytoplasm [35], we excluded the lncRNAs that only located in the nucleus to enhance the accuracy of prediction. Based on these improved methods, we constructed a novel ceRNA network.

In this study, we identified three lncRNAs of great prognostic value from the ceRNA network and further established a risk-score based LPM. Among the three lncRNAs in the LPM, LINC00460 is the most studied oncogenic lncRNA. On a large scale of cancer types, LINC00460 functions as a competing endogenous RNA, sponging multiple miRNAs, indicating that it plays a vital role in promoting tumor cell proliferation, migration, and invasion [36–38]. However, a rare study has explored the role of LINC00460 in KIRC, so this was the first study which found that high expression of LINC00460 is linked to poor prognosis in patients with KIRC. In the previous study, LINC00443 was reported as a tumor suppressor, which was associated with a better prognosis of KIRC, while MIAT was associated with worse prognosis [39–41]. In our study, expression and survival analysis of LINC00443 and MIAT revealed the high expression of MIAT and the low expression of LINC00443 in KIRC samples, which was related to poor prognosis and better prognosis, respectively. However, a further basic study needs to be undertaken to validate their molecule functions in the development of KIRC.

In addition, we demonstrated that the LPM remained an independent prognostic factor after adjusting the traditional clinical characteristics. This result suggests that LPM has the potential to improve the predictive ability of traditional features. Therefore, we propose a comprehensive assessment that combines our LPM and other clinical features (age, grade, and stage). The calibration curve showed satisfactory agreement between the observed values and the predicted values for 1-, 5-, and 10-year OS. The main advantage of this model is that it provides a complementary perspective on individual tumors and develops an individual scoring system for patients. Therefore, our nomogram could be a promising tool for clinicians in the future.

However, our study is limited because we only validated the expression of three lncRNAs in the individual clinical cohorts, and it would be better if clinical cohorts could validate the prognostic value of LPM. Besides, prospective studies are further needed to perform to confirm its predictive ability. Despite these limitations, our study provided a new insight into developing a prognostic score system for KIRC patients. This LPM can easily separate patients with poor prognosis from patients with a good prognosis by performing PCR. Furthermore, clinicians can develop more individualized treatment regimens for patients with different prognosis, and this will make treatment more targeted.

Conclusions

In conclusion, we successfully constructed a novel ceRNA regulatory network, which narrowed the scope of predicting prognostic biomarkers and therapeutic targets for KIRC. Besides, we identified and validated an LPM which is based on three lncRNAs involved in the ceRNA network, and it has independent and great prognostic value for KIRC patients.

Abbreviations

KIRC

Kidney Renal Clear Cell Carcinoma

lncRNA

Long noncoding RNA

ceRNA

competitive endogenous RNA

DElncRNAs

differentially expressed lncRNAs

LPM

lncRNA prognostic model

OS

overall survival

TCGA

The Cancer Genome Atlas

LASSO

least absolute shrinkage and selection operator

K-M

Kaplan–Meier

ROC

receiver operating characteristic

DSS

disease-specific survival

C-index

concordance index

DCA

decision curve analysis

GSEA

gene set enrichment analysis

GTEX

Genotype-Tissue Expression

GEPIA2

Gene Expression Profiling Interactive Analysis

CI

confidence interval

AUC

area under the ROC curve

HR

Hazard ratio

Declarations

Ethics approval and consent to participate

This study was carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki). All patients signed the informed consent, and this study was approved by the ethics committees of Beijing Chao-Yang Hospital.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China [grant numbers 81670679 and 81970645].

Authors' contributions

ZD initiated the study and organized; ZD, ZS and Hu XP designed and carried out bioinformatics analyses, statistical analyses, drew figures and drafted the manuscript; ZS and Hu XP participated in modifying the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

[1] Capitanio U, Montorsi F. Renal cancer. *Lancet*. 2016;387(10021):894-906. doi:10.1016/s0140-6736(15)00046-x.

[2] Barata PC, MSc, FACP. Treatment of Renal Cell Carcinoma: Current Status and Future Directions. *Ca A Cancer Journal for Clinicians*. 2017;67.

[3] Prensner JR, Chinnaiyan AM. The Emergence of lncRNAs in Cancer Biology. 2011;1(5):391.

- [4] Lin C, Yang L. Long Noncoding RNA in Cancer: Wiring Signaling Circuitry. *Trends Cell Biol.* 2018;28(4):287-301. doi:10.1016/j.tcb.2017.11.008.
- [5] Schmitt AM, Chang HY. Long Noncoding RNAs in Cancer Pathways. *Cancer cell.* 2016;29(4):452-63. doi:10.1016/j.ccell.2016.03.010.
- [6] Deng M, Blondeau JJ, Schmidt D, Perner S, Muller SC, Ellinger J. Identification of novel differentially expressed lncRNA and mRNA transcripts in clear cell renal cell carcinoma by expression profiling. *Genom Data.* 2015;5:173-5. doi:10.1016/j.gdata.2015.06.016.
- [7] Blondeau JJ, Deng M, Syring I, Schrodter S, Schmidt D, Perner S, et al. Identification of novel long non-coding RNAs in clear cell renal cell carcinoma. *Clin Epigenetics.* 2015;7:10. doi:10.1186/s13148-015-0047-7.
- [8] Fachel AA, Tahira AC, Vilella-Arias SA, Maracaja-Coutinho V, Gimba ER, Vignal GM, et al. Expression analysis and in silico characterization of intronic long noncoding RNAs in renal cell carcinoma: emerging functional associations. *Mol Cancer.* 2013;12(1):140. doi:10.1186/1476-4598-12-140.
- [9] Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet.* 2016;17(1):47-62. doi:10.1038/nrg.2015.10.
- [10] Chen LL. Linking Long Noncoding RNA Localization and Function. *Trends Biochem Sci.* 2016;41(9):761-72. doi:10.1016/j.tibs.2016.07.003.
- [11] Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. *Cell.* 2013;152(6):1298-307. doi:10.1016/j.cell.2013.02.012.
- [12] Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell.* 2011;146(3):353-8. doi:10.1016/j.cell.2011.07.014.
- [13] Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, et al. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell.* 2011;147(2):358-69. doi:10.1016/j.cell.2011.09.028.
- [14] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. doi:10.1186/s13059-014-0550-8.
- [15] Mas-Ponte D, Carlevaro-Fita J, Palumbo E, Hermoso Pulido T, Guigo R, Johnson R. LncAtlas database for subcellular localization of long noncoding RNAs. *Rna.* 2017;23(7):1080-7. doi:10.1261/rna.060814.117.
- [16] Jeggari A, Marks DS, Larsson E. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics.* 2012;28(15):2062-3. doi:10.1093/bioinformatics/bts344.

- [17] Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.* 2015;43(Database issue):D146-52. doi:10.1093/nar/gku1104.
- [18] Chou CH, Chang NW, Shrestha S, Hsu SD, Lin YL, Lee WH, et al. miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res.* 2016;44(D1):D239-47. doi:10.1093/nar/gkv1258.
- [19] Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife.* 2015;4. doi:10.7554/eLife.05005.
- [20] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-504. doi:10.1101/gr.1239303.
- [21] Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 2011;39(Web Server issue):W316-22. doi:10.1093/nar/gkr483.
- [22] Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 2003;4(5):P3.
- [23] Long J, Wang A, Bai Y, Lin J, Yang X, Wang D, et al. Development and validation of a TP53-associated immune prognostic model for hepatocellular carcinoma. *EBioMedicine.* 2019;42:363-74. doi:10.1016/j.ebiom.2019.03.022.
- [24] Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2004;10(21):7252-9. doi:10.1158/1078-0432.Ccr-04-0713.
- [25] Kiran M, Chatrath A, Tang X, Keenan DM, Dutta A. A Prognostic Signature for Lower Grade Gliomas Based on Expression of Long Non-Coding RNAs. *Mol Neurobiol.* 2019;56(7):4786-98. doi:10.1007/s12035-018-1416-y.
- [26] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, 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):15545-50. doi:10.1073/pnas.0506580102.
- [27] Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* 2019;47(W1):W556-w60. doi:10.1093/nar/gkz430.
- [28] Hakimi AA, Voss MH, Kuo F, Sanchez A, Liu M, Nixon BG, 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. doi:10.1158/2159-8290.Cd-18-0957.

- [29] Barbieri CE, Chinnaiyan AM, Lerner SP, Swanton C, Rubin MA. The Emergence of Precision Urologic Oncology: A Collaborative Review on Biomarker-driven Therapeutics. *European urology*. 2017;71(2):237-46. doi:10.1016/j.eururo.2016.08.024.
- [30] Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. *Nature*. 2012;482(7385):339-46. doi:10.1038/nature10887.
- [31] Zhou S, Treloar AE, Lupien M. Emergence of the Noncoding Cancer Genome: A Target of Genetic and Epigenetic Alterations. *Cancer discovery*. 2016;6(11):1215-29. doi:10.1158/2159-8290.Cd-16-0745.
- [32] Reis-Filho JS, Pusztai L. Gene expression profiling in breast cancer: classification, prognostication, and prediction. *Lancet*. 2011;378(9805):1812-23. doi:10.1016/s0140-6736(11)61539-0.
- [33] Jiang W, Guo Q, Wang C, Zhu Y. A nomogram based on 9-lncRNAs signature for improving prognostic prediction of clear cell renal cell carcinoma. *Cancer Cell Int*. 2019;19:208. doi:10.1186/s12935-019-0928-5.
- [34] Zhang C, Huang D, Liu A, Xu Y, Na R, Xu D. Genome-wide screening and cohorts validation identifying novel lncRNAs as prognostic biomarkers for clear cell renal cell carcinoma. *J Cell Biochem*. 2020;121(3):2559-70. doi:10.1002/jcb.29478.
- [35] Cao Z, Pan X, Yang Y, Huang Y, Shen HB. The lncLocator: a subcellular localization predictor for long non-coding RNAs based on a stacked ensemble classifier. *Bioinformatics*. 2018;34(13):2185-94. doi:10.1093/bioinformatics/bty085.
- [36] Li K, Sun D, Gou Q, Ke X, Gong Y, Zuo Y, et al. Long non-coding RNA linc00460 promotes epithelial-mesenchymal transition and cell migration in lung cancer cells. *Cancer letters*. 2018;420:80-90. doi:10.1016/j.canlet.2018.01.060.
- [37] Liu X, Wen J, Wang H, Wang Y. Long non-coding RNA LINC00460 promotes epithelial ovarian cancer progression by regulating microRNA-338-3p. *Biomed Pharmacother*. 2018;108:1022-8. doi:10.1016/j.biopha.2018.09.103.
- [38] Feng L, Yang B, Tang XD. Long noncoding RNA LINC00460 promotes carcinogenesis via sponging miR-613 in papillary thyroid carcinoma. *J Cell Physiol*. 2019;234(7):11431-9. doi:10.1002/jcp.27799.
- [39] Qu Y, Xiao H, Xiao W, Xiong Z, Hu W, Gao Y, et al. Upregulation of MIAT Regulates LOXL2 Expression by Competitively Binding MiR-29c in Clear Cell Renal Cell Carcinoma. *Cell Physiol Biochem*. 2018;48(3):1075-87. doi:10.1159/000491974.
- [40] Wang J, Zhang C, He W, Gou X. Construction and comprehensive analysis of dysregulated long non-coding RNA-associated competing endogenous RNA network in clear cell renal cell carcinoma. *J Cell Biochem*. 2018. doi:10.1002/jcb.27557.

Figures

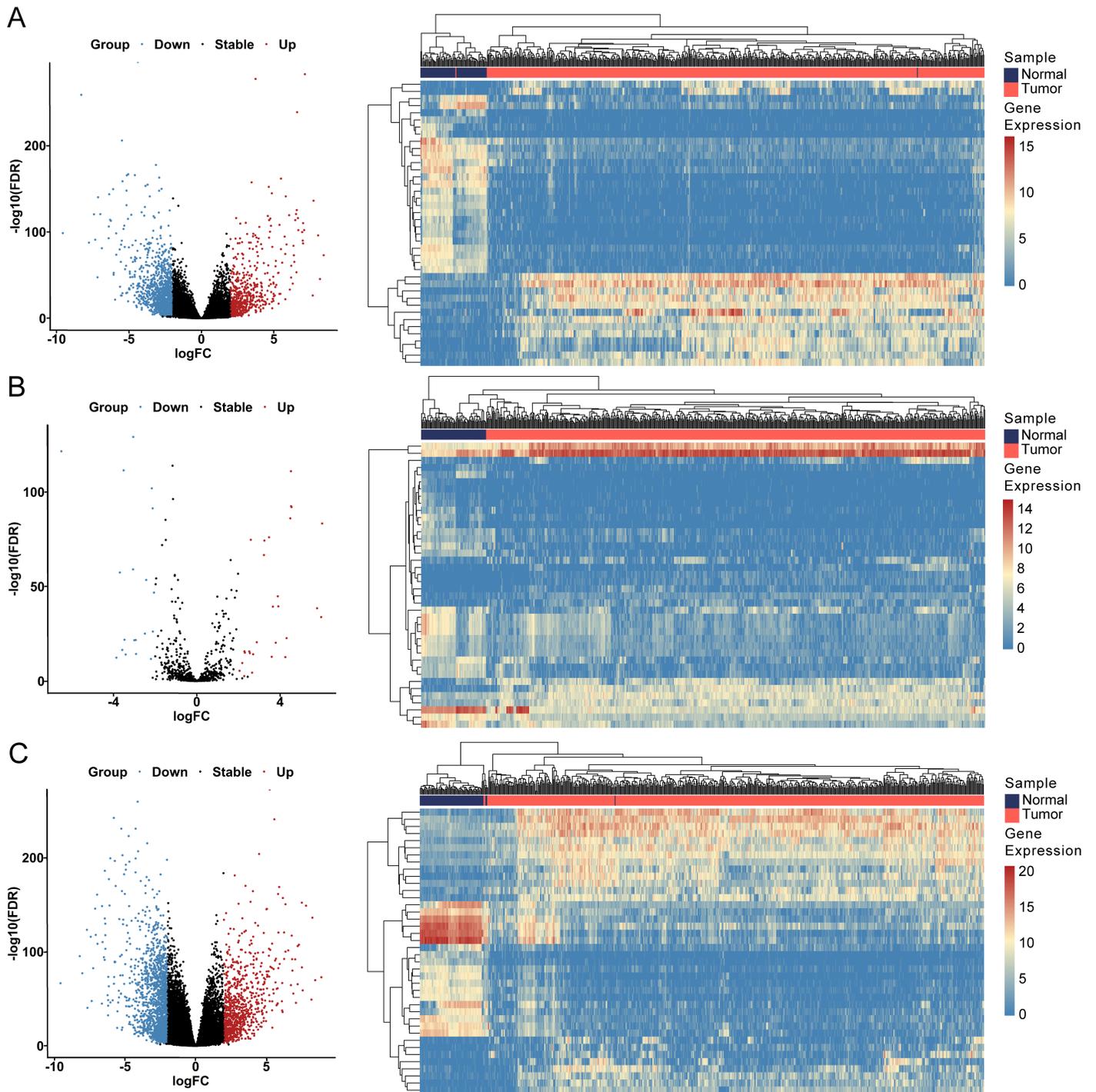


Figure 1

Identification of differentially expressed genes. Volcano plots and heatmap plots of differentially expressed lncRNAs (A), miRNAs (B), and mRNAs (C) between KIRC samples and adjacent normal samples. Red denotes upregulated genes, and blue denotes downregulated genes in both volcano plots and heatmaps. The horizontal axis of the heatmaps represents the samples, and the vertical axis of heatmaps presents the top forty DEGs.

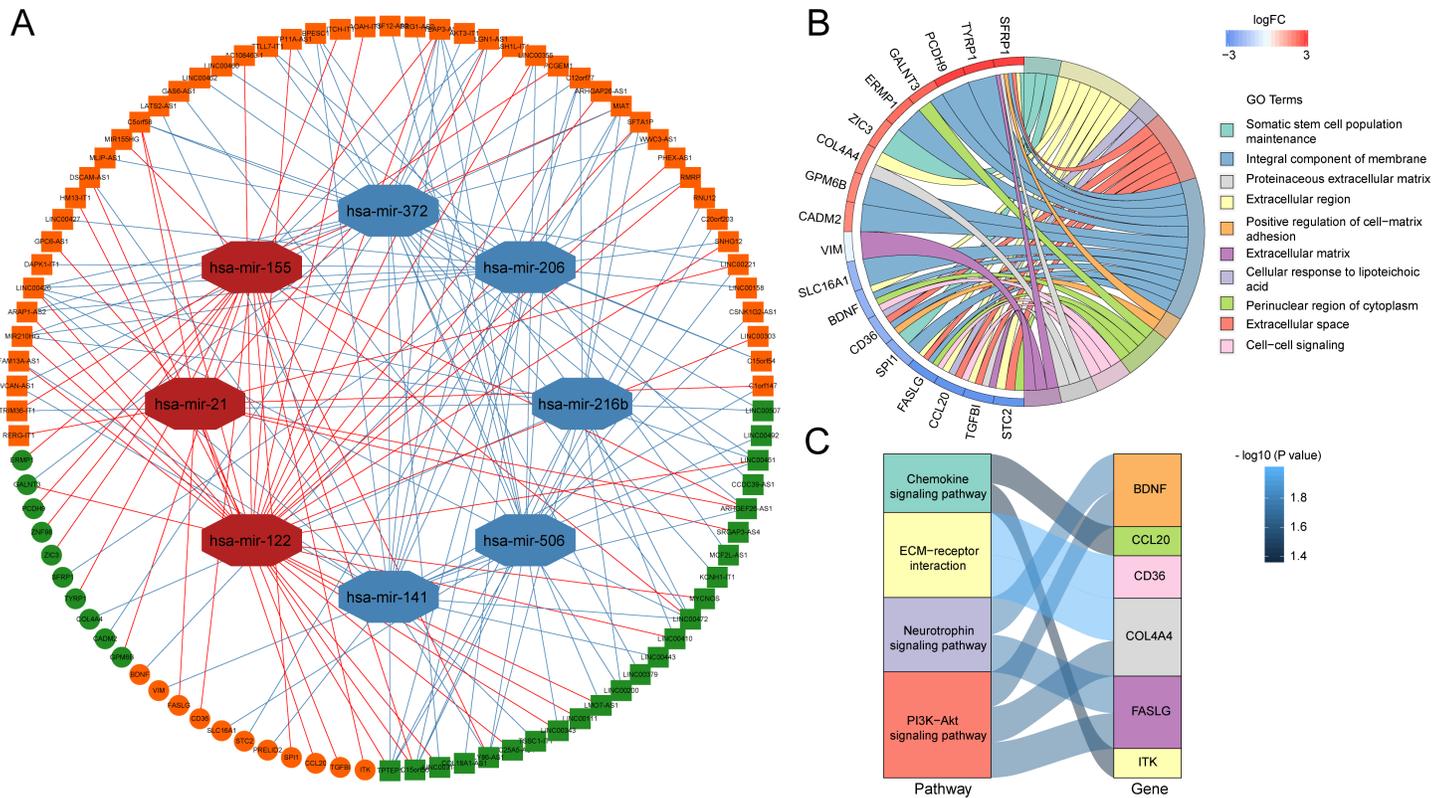


Figure 2

Construction and functional enrichment analysis of the ceRNA network. (A) The ceRNA regulatory network in KIRC. Red octagons represent upregulated miRNAs, and Blue octagons represent downregulated miRNAs. Orange circles and squares stand for upregulated mRNAs and lncRNAs, respectively. Green circles and squares present downregulated mRNAs and lncRNAs, respectively. (B) Chord plot of GO enrichment analysis for the ceRNA network. The GO terms are shown and annotated on the right of the chord diagram, while the DEmRNAs in the ceRNA network are shown on the left. Red bricks represent upregulated genes, and blue bricks represent downregulated genes. (C) Sankey diagram of KEGG pathway analysis for the ceRNA network. Rectangles on the left of the Sankey plot represent the significant pathways, while rectangles on the right represent DEmRNAs in the ceRNA network.

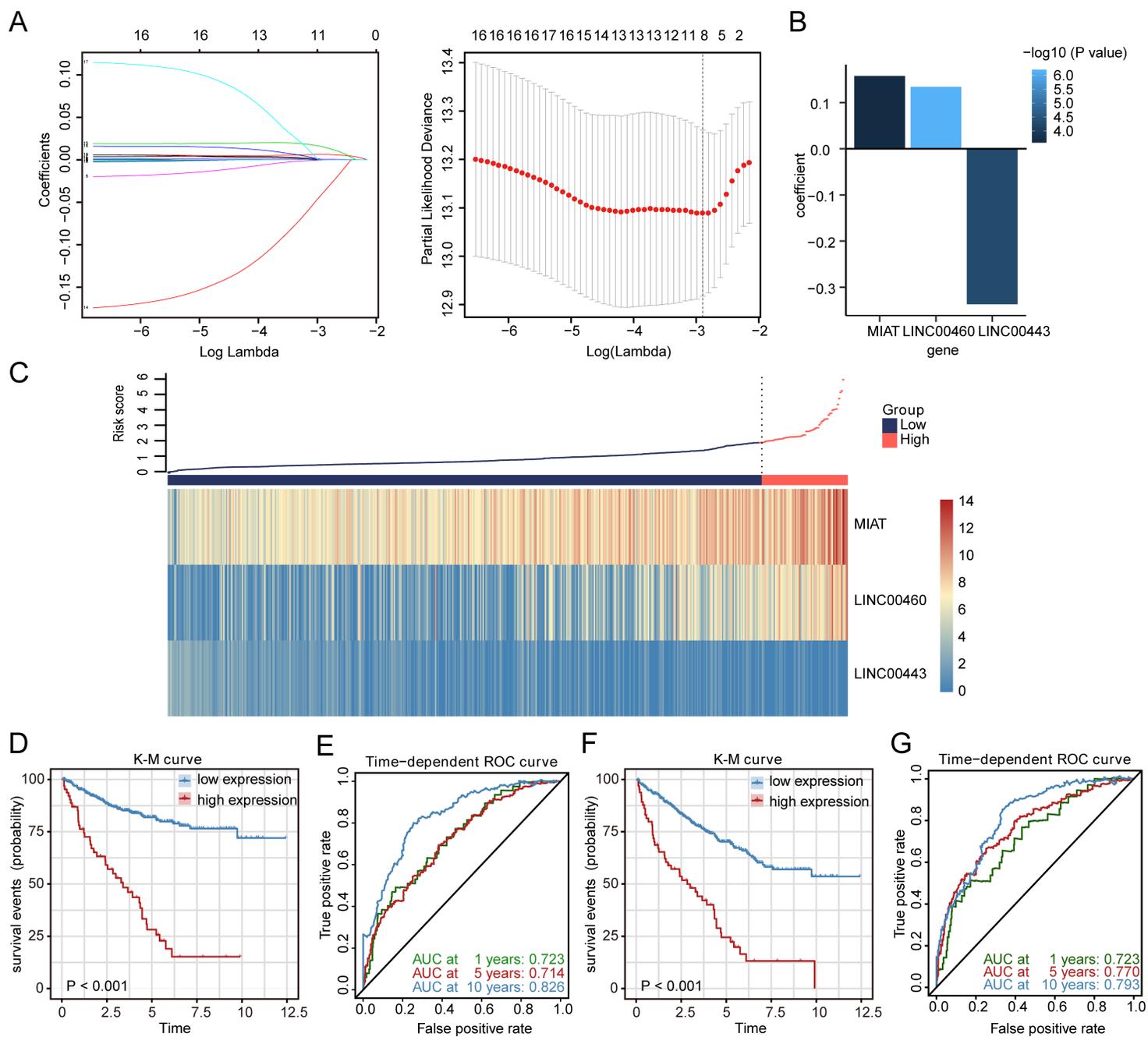


Figure 3

The lncRNA prognostic model. (A) Lasso-penalized Cox regression analysis of 17 DElncRNAs. Ten-fold cross-validation was applied to calculate the best lambda, which leads to a minimum mean cross-validated error. (B) The bar plot shows coefficients of three lncRNAs in the LPM. The color of the bars represented the P-value of the coefficients. (C) Risk score system of the LPM. The above scatter plot exhibits the risk scores of each KIRC patient with survival data, and the below heatmap displays the expression profiles of three lncRNAs in the LPM. K-M survival curves and time-dependent ROC curves of OS are shown in (D) and (E), respectively. K-M survival curves and time-dependent ROC curves of DSS are demonstrated in (F) and (G), respectively.

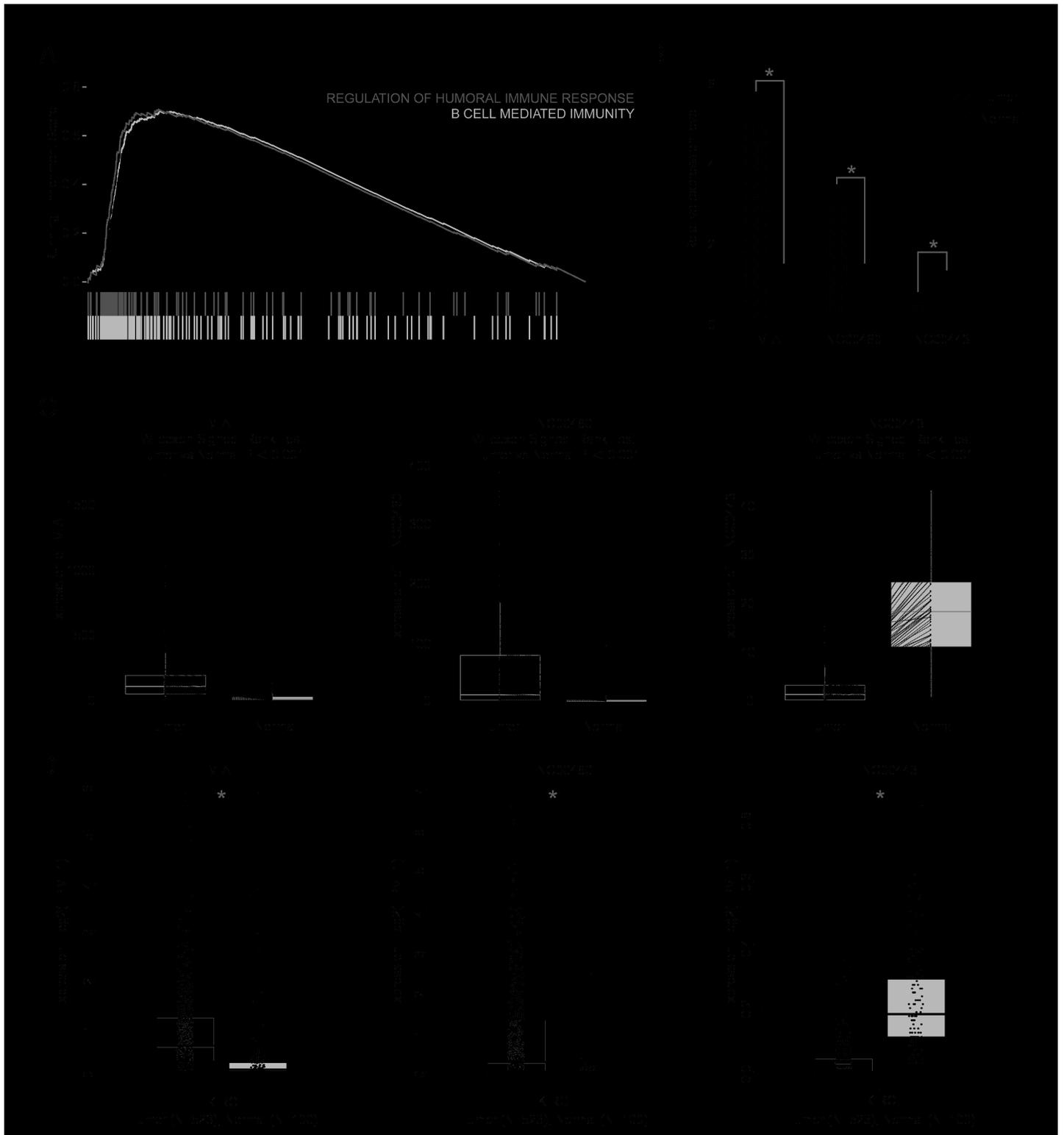


Figure 4

GSEA and validations of three lncRNAs. (A) GSEA plot shows the top three immune processes associated with high risk score. (B) The barplot exhibits the expressions of MIAT, LINC00443, and LINC00460 evaluated by RT-PCR in 21 KIRC samples and paired adjacent normal samples. * $P < 0.05$ versus control. (C) The differential expression patterns of three lncRNAs in 72 KIRC patients with paired KIRC and adjacent normal samples from TCGA. (D) The expression profiles of three lncRNAs were further

compared between the TCGA KIRC cohort (523 KIRC samples) and the GTEx cohort (100 normal kidney samples) via GEPIA2.

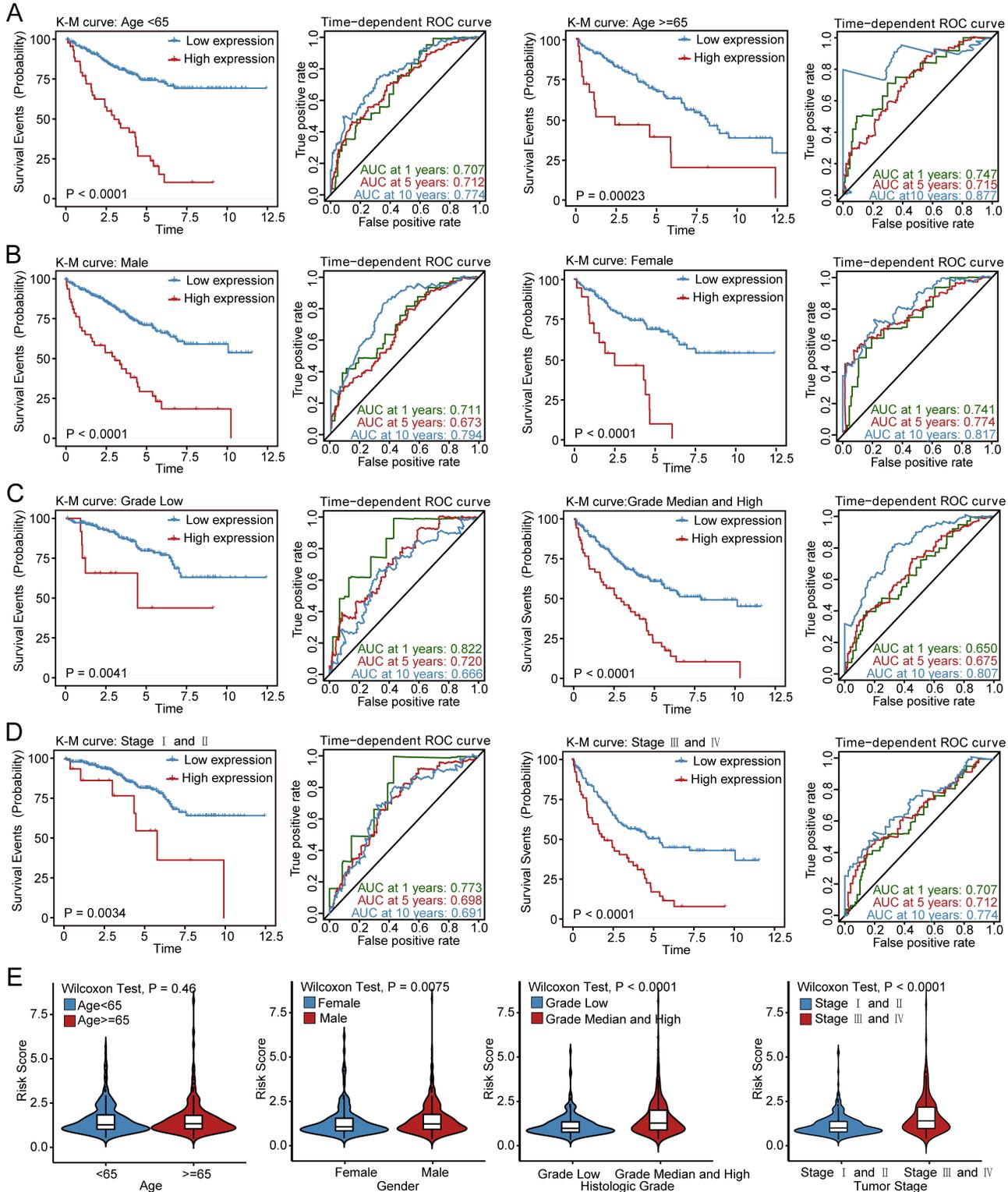


Figure 5

Stratification analysis of the LPM. Kaplan–Meier curves and time-dependent ROC curves illustrate the prognostic value of LPM based on the stratification of different clinical features. Notes: (A) Age; (B)

Gender; (C) Histological grade; (D) Tumor stage; (E) Violin plots exhibit the risk score distributions within different clinical parameters stratification.

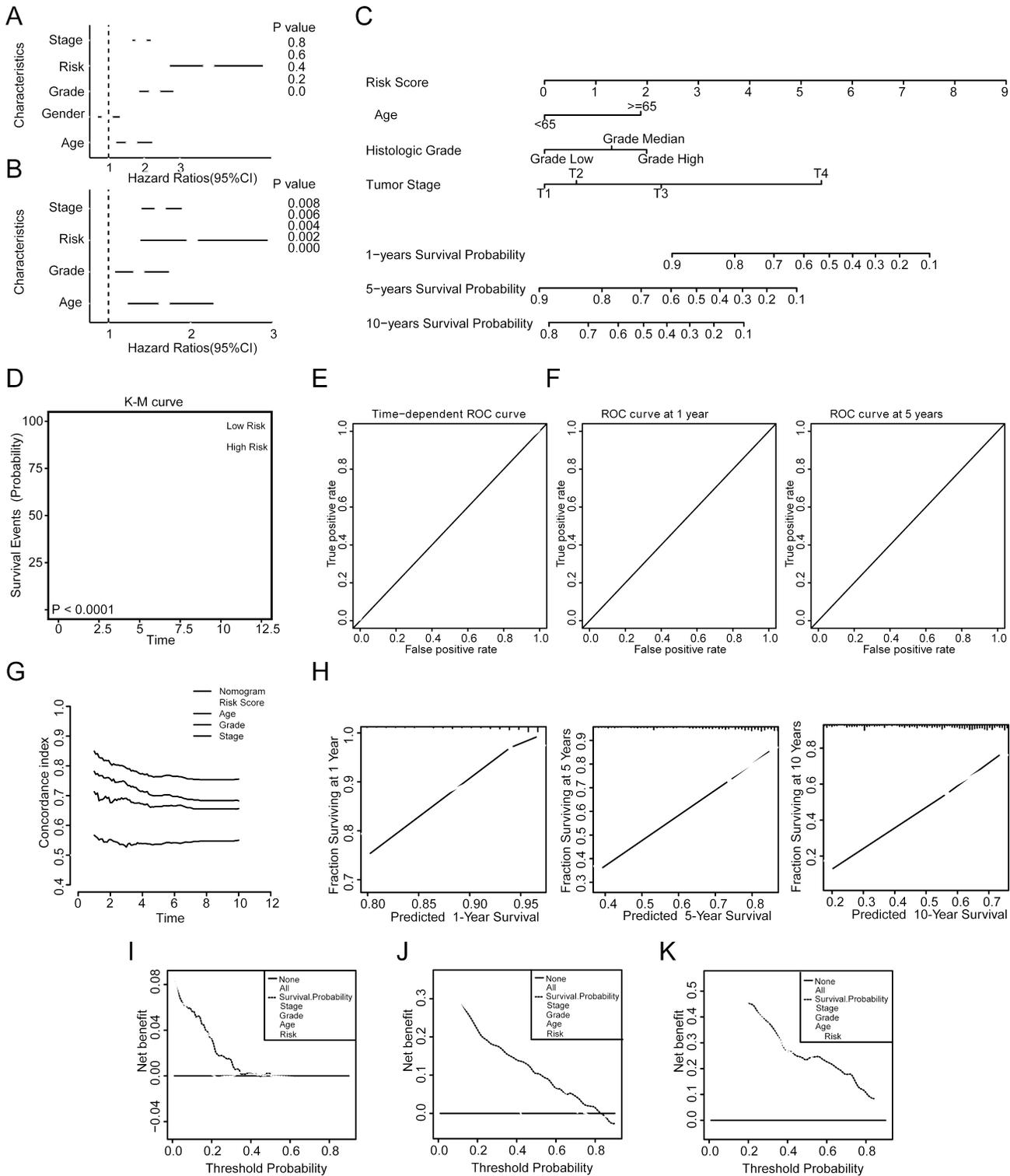


Figure 6

The LPM-based nomogram model. (A) Univariate and (B) multivariate regression analysis of the relation between the LPM and clinical features regarding prognostic value. (C) Nomogram for predicting the probability of 1-, 5-, and 10-year OS for KIRC patients. (D) Kaplan–Meier curves demonstrate the

prognostic performance of the nomogram. (E, F) Time-dependent ROC curves of the nomogram, the LPM, age, histological grade, and tumor stage. (G) The prognostic performance was compared among the nomogram, the LPM, and conventional clinical characteristics by calculating the C-index. (H) Calibration plots of the nomogram for predicting the probability of OS at 1-, 5-, and 10-years. (I, J, K) DCA plots of the nomogram, the LPM, and clinical characteristics for predicting the probability of OS at 1-, 5-, and 10-years.

Supplementary Files

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

- [SupplementalTables.docx](#)