

# Development and validation of a set of novel and robust 4-lncRNA-based nomogram predicting prostate cancer survival by bioinformatics analysis

**Peng Zhang**

Department of Urology, Weihai Central Hospital, Weihai, Shandong

**Aiyu Wang**

The Operating Room, Weihai Central Hospital, Weihai, Shandong

**Liming Dong**

Department of Urology, Weihai Central Hospital, Weihai, Shandong

**Xuefeng Zhang** (✉ [doctor13696310683@163.com](mailto:doctor13696310683@163.com))

Weihai central hospital <https://orcid.org/0000-0001-6115-9064>

---

## Research article

**Keywords:** prostate cancer, lncRNA, Prognostic signature, Survival Analysis, TCGA

**Posted Date:** August 18th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-48992/v1>

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

[Read Full License](#)

---

# Abstract

## Background and Objective:

There is significant heterogeneity between cellular composition and patient outcome in prostate cancer (PCa). Accumulating evidence shows that long noncoding RNAs (lncRNAs) possess great potential in the diagnosis and prognosis of PCa with biological and clinical significance. Therefore, this study aimed to construct an lncRNA-based signature to more accurately predict the prognosis of different PCa patients, so as to improve patient management and prognosis.

## Methods

The Cancer Genome Atlas (TCGA) database was used to download RNA-seq expression data together with the clinical information of 499 PCa tissue samples as well as 52 corresponding non-carcinoma tissue samples. Differently expressed lncRNAs (DELncRNAs) were selected based on tumor tissues and non-carcinoma samples. Through univariate and multivariate Cox regression analysis, this study constructed a 4 lncRNAs-based prognosis nomogram for the classification and prediction of survival risk in patients with PCa. The receiver operating characteristic (ROC) curve was plotted for detecting and validating our prediction model sensitivity and specificity. In addition, univariate as well as multivariate Cox regression was conducted to examine whether the constructed lncRNA signature's prediction ability was independent of additional clinicopathological variables (like age, Gleason score, N stage, T stage and M stage) among PCa cases. Possible biological functions for those prognostic lncRNAs were predicted through gene ontology (GO) together with Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis on those 4 protein-coding genes (PCGs) related to lncRNAs.

## Results

A total of 451 differently expressed lncRNAs (DELncRNAs) related to the overall survival (OS) rate for PCa cases were screened from 3838 lncRNAs in the TCGA database. Four lncRNAs (HOXB-AS3, YEATS2-AS1, LINC01679, PRRT3-AS1) were extracted after univariate as well as multivariate COX regression analysis for classifying patients into high and low-risk groups by different OS rates. As suggested by ROC analysis, our proposed model showed high sensitivity and specificity. Independent prognostic capability of the model from other clinicopathological factors was indicated through further analysis. Based on functional enrichment, those action sites for prognostic lncRNAs were mostly located in the extracellular matrix and cell membrane, and their functions are mainly associated with the adhesion, activation and transport of the components across the extracellular matrix or cell membrane.

## Conclusion

Our current study successfully identifies a novel four-lncRNA candidate, which can provide more convincing evidence for prognosis in addition to the traditional clinicopathological indicators to predict the PCa survival, and laying the foundation for offering potentially novel therapeutic treatment. Additionally, this study sheds more lights on the PCa-related molecular mechanisms.

## Introduction

Prostate cancer (PCa) represents a frequently occurring cancer in men globally [1]. A number of clinicopathological factors, such as the Gleason score, margin status, or the TNM stage, are incorporated into prior models to diagnose and detect PCa after treatment[2]. Of such factors, Gleason score exhibits the highest sensitivity and effectiveness. Nonetheless, the subjectivity and sampling error related to the assessment of Gleason score are the prominent confounders. Recently, increasing research attempts to establish the gene molecular signature for enhancing the prediction ability for PCa[3]. Nonetheless, little existing research has examined the possible roles of lncRNAs in the prediction of PCa survival risk as the new models. lncRNAs have been usually referred to as the RNA transcripts with the length of over 200 nucleotides (nt) that can not encode proteins [4]. It is increasingly suggested that, lncRNAs have exerted vital parts in numerous biological processes through regulating gene expression, or proliferation and apoptosis of cells[5]. The aberrant lncRNAs are related to tumorigenesis, which may serve as the oncogenes or tumor suppressor genes. In addition to their parts in tumor genesis and development, lncRNAs may serve as the candidate biomarkers [6]. So far, bioinformatic analysis is substantially used for molecular biological experiments or in clinic. Therefore, this work aimed to employ bioinformatic analysis to discover those differently expressed lncRNAs (DElncRNAs) between PCa and non-carcinoma tissue samples from the sequencing data obtained through The Cancer Genome Atlas (TCGA). Then, Cox regression analysis together with the survival associated risk score formula was used to develop a nomogram for prognosis based on four lncRNAs, so as to differentiate cases with favorable or dismal prognostic outcome. The receiver operating characteristic (ROC) curve was also plotted, with a higher area under curve (AUC) indicated favorable model sensitivity and specificity. In addition, univariate as well as multivariate Cox regression suggested that this 4 lncRNAs-based prognosis nomogram showed higher independence and robustness in predicting prognosis compared with additional clinicopathological variables. Besides, analyzing the functions and involved pathways for these 4 lncRNAs shed more lights on the lncRNAs prediction ability and the possible molecular mechanism.

## Materials And Methods

### The overall analysis route in this work

The general research analysis route can be observed from Fig. 1. First, those transcriptome RNA-seq expression profiles from 551 cases were obtained based on the TCGA database. Then lncRNA was analyzed for differential expression and batch survival. Univariate and multivariate COX regression analysis was performed for differentially expressed lncRNAs to establish a gene model of 4-lncRNA. Survival analysis, ROC curve analysis, and patient risk heat plot, risk curve and survival status plot

were performed respectively for our model. Target genes of lncRNAs were predicted by co-expression, and then the target genes were analyzed by functional clustering analysis. Finally, independent prognostic analysis and correlation analysis were conducted between our 4-lncRNA nomogram and other common clinical characteristics, as well as further stratified analysis and combined analysis of Gleason Score.

## Datasets

lncRNA RNA-seq data (HTSeq-Counts) including PCa and control samples and corresponding clinical data of PCa patients were extracted based on the publicly available Genomic Data Commons (GDC) data portal (<https://portal.gdc.cancer.gov/>). Altogether 551 gene expression profiling samples were collected, comprising 499 PCa and 52 non-carcinoma tissues. Samples with insufficient clinical data or those with OS < 3 months were eliminated from the present research.

## Identification of DElncRNAs between PCa and non-carcinoma tissue samples

In line with our inclusion criteria, clinicopathological data from 397 PCa cases were obtained (**Table 1**). Clinical covariates for both cases and cancers can be observed from **Table 1**. The R4.0.1 software was employed for analysis. For identifying the candidate lncRNAs in later survival analysis, the trimmed mean of M values was adopted to normalize and differentially analyze the expression profiles by edgeR package from Bioconductor [7]. The heterogeneities of lncRNA expression in PCa compared with adjacent normal tissues were presented in the form of log<sub>2</sub>FC together with related P-value. In this study, |log<sub>2</sub>FC| >1 and false discovery rate (FDR)  $q < 0.05$  were selected as the thresholds. Thereafter, the R package pheatmap function (version 1.0.12) was used to perform unsupervised hierarchical clustering according to DElncRNA expression levels.

## Survival analysis and development of the lncRNAs expression model

Firstly, RNA-seq expression data were normalized according to log<sub>2</sub>. Later, the Survival R package from CRAN (<https://rweb.stat.umn.edu/R/site-library/survival/html/00Index.html>) was used to assess the relationships of DElncRNAs with patient OS through univariate Cox proportional hazards regression. lncRNAs with p-value < 0.05 were identified to be the potential variables, which were incorporated into the stepwise multivariate Cox regression examined through Akaike Information Criterion (AIC), which help to determine the optimal exchange between model complexity and the optimal informative and explanatory effectiveness.

## Risk stratification, survival curve and concordance index (C-index)

The riskscore values for all patients were determined by the following formula (Risk score = 0.524052278 × HOXB-AS3 – 0.663887578 × LINC01679 – 0.504710478 × PRRT3-AS1 + 1.092940489 × YEATS2-AS1) on the basis of multivariate Cox regression. Using the as-prepared risk scoring system, all cases were

classified as high- or low-risk group by median riskscore value. Thereafter, heterogeneities of OS time between these two groups were examined through two-sided log-rank test. The Kaplan-Meier (K-M) method was employed to plot the OS curves for these two groups. Besides, the AUC values were determined for comparing the model sensitivity and specificity in predicting OS by the “survival ROC” of R package. The model discrimination ability was evaluated by calculating the C-index.

## **Predictive independence of our 4 lncRNAs-based nomogram for survival rate from additional clinicopathological factors**

The independence of our 4 lncRNAs-based nomogram from additional clinicopathological factors (such as age, Gleason score, and TNM classification) in predicting OS was examined by univariate as well as multivariate Cox regression. OS was used to be a dependent variable, whereas the constructed lncRNA nomogram together with additional common clinical factors were used to be the independent variables.

## **Joint analysis of the four-lncRNA signature with Gleason score**

In order to test whether our model could accurately predict the prognosis of patients with Gleason score  $\geq 8$ , a stratified analysis was carried out. To verify whether our lncRNA model could enhance Gleason score's accuracy in PCa survival risk prediction, the ROC curves of the three comparisons was drawn and the AUCs were calculated.

## **Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses**

Using the z-test and two-sided Pearson correlation coefficients, Pearson correlation coefficients were examined between those 4 lncRNAs expression and PCGs, so as to discover the possible biological processes together with related pathways involving those predictive lncRNAs. Typically, PCGs showing  $|\text{Pearson correlation coefficient}| > 0.40$  and  $P < 0.01$  were identified to show positive or negative relationship with lncRNAs. GO and KEGG analyses were conducted on the lncRNA-related PCGs via the Database for Annotation, Visualization, and Integrated Discovery at the threshold of false discovery rate (FDR)  $q < 0.05$ .

## **Results**

### **Identification of DElncRNAs between PCa and non-carcinoma tissue samples**

Based on thresholds of  $|\log_2\text{FC}| > 1.0$  and false discovery rate (FDR)  $q < 0.05$ , altogether 451 lncRNAs, among which, 307 were up-regulated whereas 144 were down-regulated, were discovered to show different expression between PCa and the non-carcinoma tissue samples, and they were applied in later

stepwise survival analysis (**Supplementary:TableS1**). The expression profiles were intuitively reflected by volcano plots (**Figure 2**). According to the DEIncRNAs profiles, the unsupervised hierarchical cluster analysis was carried out, which suggested the possibility to distinguish between PCa and non-carcinoma samples (**SupplementaryFigure S1**). To identify prognosis-related lncRNAs which are associated with patients' OS in PCa, the lncRNA expression profiles were evaluated by univariate Cox regression analysis. 36 lncRNAs related to OS were selected at the threshold of 0.05 (**Supplementary2: Table S2**), which were then applied for stepwise multivariate Cox regression analysis, and four lncRNAs therein (HOXB-AS3, LINC01679, PRRT3-AS1, YEATS2-AS1) (as shown in **Figure 3, Table2**) were ultimately screened out from the 451 lncRNAs identified before to establish a predictive model.

Expression profiles of those 4 lncRNAs were integrated with the related regression coefficients to construct a prognosis nomogram. Based on the above-mentioned analysis, the riskscore formula was the sum of 4 lncRNAs expression levels weighted by the corresponding relative regression coefficients obtained upon multivariate Cox regression, as shown below: survival risk score = (0.524052278 × HOXB-AS3 expression) + (1.092940489 × YEATS2-AS1 expression) + (-0.663887578 × LINC01679 expression) + (-0.504710478 × PRRT3-AS1 expression). Of these four lncRNAs, two had positive coefficients upon multivariate Cox regression analysis, associated with high risk because the up-regulated level indicated the reduced patient OS (HOXB-AS3 and YEATS2-AS1, Coef > 0) and the remaining two lncRNAs (LINC01679 and PRRT3-AS1, Coef < 0) were shown negative coefficients upon Cox regression analysis, which indicated that such lncRNAs were protective, because cases having up-regulate lncRNAs levels tended to show extended OS relative to patients having decreased expression (**Figure3**).

## Favorable performance for the 4-lncRNA prognostic model in the prediction of OS for PCa cases

The 493 cases were classified as high- (n=246) or low-risk (n=247) group in line with the median riskscore (0.943) obtained based on those 4 lncRNAs expression levels (also defined as the survival risk score, SRS) (**Figure 4** and **Supplementary : Table S3**). Difference in survival was determined by log-rank test. The K-M method was used for survival analysis. It was illustrated from **Figure 5** that, the KM OS curves for both groups on the basis of 4 lncRNAs showed notable difference (p=3.3e-03). Typically, the low-risk group showed significant correlation with favorable prognosis compared with high-risk group. The AUC values of time-dependent ROC curves were calculated to evaluate our constructed nomogram prognostic ability. The greater AUC value is indicative of the higher nomogram performance, and that AUC of >0.90 has excellent performance. Based on the analysis results of ROC curves, the AUC values at 1, 3 and 5 years were 0.997, 0.929 and 0.928, separately (**Figure 6**), revealing the high sensitivity and specificity of our 4 lncRNAs-based signature in the prediction of OS risk for PCa cases. Additionally, the model C-index was calculated, (C-index=0.9203, CI: 0.8482-0.9924, p-value=3.13447e-30), exhibiting good model performance.

## 4 lncRNAs-based nomogram's predictive ability was not dependent on other clinicopathological factors

For investigating the distinguishing ability of our constructed 4 lncRNAs-based signature of the survival risk for PCa cases after considering additional possible traditional prognostic factors, univariate as well as multivariate analysis was conducted for evaluating the model independent prognostic significance. The multivariate analysis results demonstrated that our 4 lncRNAs-based signature might be used to be the potent factor to independently predict PCa OS rate from other clinical factors (hazard ratio(HR)=1.014, 95% CI 1.005-1.023, p=0.003), shown in **Table 3**, and **Figure 7**, compared with conventional clinicopathological factors such as age, Gleason score and TNM classification.

## Contrast of the four-lncRNA signature with Gleason score

Hierarchical analysis showed that PCa with Gleason score>7 could be further stratified as 2 groups that had different survival by the nomogram (log-rank test, p<0.05, **Figure 8**). Combining the 4 lncRNAs-based signature and the Gleason score remarkably increased the model prognostic capacity compared with the Gleason score alone (AUC, 0.949 vs. 0.896 vs. 0.633) (**Figure 9**).

## Identification of the prognostic lncRNAs signature associated biological functional characteristic

After determining the associations between those 4 lncRNAs and the PCGs, this study selected co-expression between 1 of those 4 lncRNAs and 4080 PCGs (|Pearson correlation coefficient| > 0.4, P-value < 0.05, and q-value <0.05). Afterwards, GO and KEGG analyses were conducted for the lncRNAs-related PCGs to display the possible functions for those 4 prognostic lncRNAs. The 4080 PCGs were most significantly enriched into 28 GO terms and they were located in the extracellular matrix (ECM) and cell membrane, and their functions are mostly associated with the adhesion, activation and transport of substances across the extracellular matrix and cell membrane (such as GO: 0003779: actin binding, GO: 0022803: passive transmembrane transporter activity, GO: 0015267: channel activity, GO: 0005216: ion channel activity, GO: 0005178: integrin binding, GO: 0005539: glycosaminoglycan binding, GO: 0022836: gated channel activity, GO: 0019955: cytokine binding, GO: 0098631: cell adhesion mediator activity, GO: 0046873: metal ion transmembrane transporter activity) (**Figure 10A** and **Supplementary Figure S2 and Table S4**). Three KEGG pathways were enriched which are mainly concerned with the interaction and binding between cells or between cells and ECM, and cell proliferation (including hsa04390: Hippo signaling pathway, hsa04514: Cell adhesion molecules (CAMs), hsa04510: Focal adhesion) (**Figure 10B**, and **Supplementary: Figure S3, Figure S4 and Table S5**).

## Discussion

Cancer risk assessment tools are essential for individualized clinical diagnosis and personalized treatment. However, the traditional clinicopathological factors, risk stratification among PCa cases represented by Gleason Score face great challenges[8]. Therefore, it needs to be addressed to establish more sensitive and effective prediction model for PCa as soon as possible. Accumulating lncRNA research publications has provided us with a new perspective and avenue for diagnosing and treating diseases like cancers[4]. Evidence is mounting that a variety of spectrum of disease progression, including tumors, is often accompanied by aberrant expression of lncRNAs, which means it may be used to be the factor to independently predict prognosis for these diseases[9]. Hence, the present work thoroughly analyzed lncRNA profiles in PCa tissues together with the matched non-carcinoma tissues against TCGA database. Our use of TCGA RNA sequencing data ensured the maximum detection of lncRNAs. The 4 lncRNAs-based nomogram exhibits excellent discriminating ability, and the AUC values at 1, 3 and 5 years were 0.997, 0.929 and 0.928, separately for the whole dataset. The as-constructed lncRNAs-based signature in this study can serve as a novel tool and shed more lights on PCa diagnosis and treatment. The high-risk PCa cases may undergo systemic or adjuvant therapy, and the low-risk ones may receive active surveillance. As a result, treatments can be precisely evaluated. The four-lncRNA nomogram (HOXB-AS3, LINC01679, PRRT3-AS1 and YEATS2-AS1) was shown to be most strongly associated with patient prognosis and risk stratification upon univariate as well as multivariate COX analysis that incorporated the common clinicopathological risk factors such as age, TNM classification and Gleason score for PCa.

Gleason score represents a potent marker for evaluating PCa patient prognosis. Our multivariate COX analysis also indicated that only Gleason Score and our model were good enough to predict the overall survival risk of PCa. For PCa, Gleason score has always been adopted as an important criterion for adjuvant chemoradiotherapy or other treatments. Our study demonstrates that the four lncRNAs signature is more capable of being considered as reference factor for adjuvant therapy. Patients who had a high Gleason score may have the aggressive PCas with dismal prognosis[10]. However, Gleason score could not sufficiently accurately divide the survival risk and stage and predict prognosis of patients as a separate diagnostic indicator. Moreover, not all high-grade PCas (Gleason score  $\geq 8$ ) are at high risk and require further adjuvant therapy[11]. As revealed by the stratification analysis, the as-constructed model assisted in classifying PCa cases with Gleason score  $\geq 8$  to high- or low-risk group. Besides, our 4 lncRNAs-based model better increased the distinguishing power of Gleason score, which indicated that our constructed model helped to enhance the prediction accuracy for PCa survival risk.

Several previous studies identify lncRNAs as the excellent predictors for PCa survival. However, they were all confined to PCa biochemical recurrence patients [12–14]. In addition, in agreement with our data, Li Fan et al. found that silence of PRRT3-AS1 suppresses the proliferation of PCa cells and boost their autophagy and apoptosis[15]. It is also suggested previously that, HOXB-AS3 shows tight association with the dismal prognosis for numerous cancer types [16–18]. The above findings suggest that, HOXB-AS3 and PRRT3-AS1 can be used to be the prognostic biomarkers for numerous cancer types. Therefore, the above studies have revealed the nomogram reasonability and reliability. Moreover, such lncRNAs show possible significance in molecular targeted treatments. However, so far, little research is carried out

on LINC01679 and YEATS2-AS1. Future studies should concentrate on such lncRNAs and examine the functions within PCa. Moreover, such lncRNAs possibly display possible significance for molecular targeted therapy. A majority of lncRNAs have not been functionally annotated within PCa so far, we performed the biological function clustering and associated biological signaling pathway analysis on all four lncRNAs.

Accumulating evidence shows that lncRNAs participate in a wide range of biological processes through modulating mRNA levels epigenetically, transcriptionally, and post-transcriptionally. Nonetheless, there are still many functional annotation of lncRNAs in PCa to be further explored. This work suggested that the related biological pathway and function enrichment of those four lncRNAs by GO and KEGG. The main functional sites of our lncRNA signature are extracellular matrix and membrane, which are associated with the adhesion, activation, and transport of cellular active substances across the extracellular matrix or cell membrane. The possible molecular function analysis can shed light on future study to examine PCa incidence and development mechanisms.

## Conclusion

To sum up, some DElncRNAs between PCa and non-carcinoma tissue samples are identified in this study. Then, the 4 lncRNAs-based signature is constructed to predict the OS for PCa through bioinformatic analysis. As discovered by our results, our as-constructed prediction signature contributes to the effective classification of PCa as high- and low-risk groups. This prediction signature outstandingly improves the Gleason score performance in distinguishing PCa, and it can also be used to discriminate PCa (Gleason > 7). Our constructed prediction signature will assist the clinicians in taking individualized clinical treatment. However, more studies are needed to verify our results and to examine the predicting ability of our signature for adjuvant therapy safety and efficacy. Besides, functional study is also needed to further understand molecular mechanisms underlying PCa.

## Abbreviations

prostate cancer (PCa); long noncoding RNAs (lncRNAs); The Cancer Genome Atlas (TCGA); differently expressed lncRNAs (DElncRNAs); Receiver operating characteristic analysis (ROC); Gene ontology (GO); Kyoto Encyclopedia of Genes and Genomes (KEGG); protein-coding gene (PCG); hazard ratio (HR); overall survival (OS); area under curve (AUC); Genomic Data Commons (GDC); Akaike Information Criterion (AIC); concordance index (C-index); protein coding gene (PCG); survival risk score (SRS); confidence interval (CI); extracellular matrix (ECM); Cell adhesion molecules (CAMs)

## Declarations

### DATA AVAILABILITY

The datasets generated for this study can be found in TCGA.

## AUTHOR CONTRIBUTIONS

Peng Zhang and Xuefeng Zhang contributed to the study design, the code writing and analysis and language editing. Aiyu Wang and Liming Dong prepared the figures and tables. All authors had read and approved the final manuscript.

## CONFLICT OF INTEREST STATEMENT

All authors have declared no conflict of interest.

**Conflict of Interest:** The authors declare that they have no competing interests.

**Statement:** The study was approved by the ethics committee of Weihai Central Hospital.

## References

1. Farhood B, et al., *A systematic review of radiation-induced testicular toxicities following radiotherapy for prostate cancer*. J Cell Physiol, 2019.
2. Intasqui P, Bertolla RP, Sadi MV. Prostate cancer proteomics: clinically useful protein biomarkers and future perspectives. Expert Rev Proteomics. 2018;15(1):65–79.
3. Abou-Ouf H, et al. Validation of a 10-gene molecular signature for predicting biochemical recurrence and clinical metastasis in localized prostate cancer. J Cancer Res Clin Oncol. 2018;144(5):883–91.
4. Sun M, et al. LncRNA PART1 modulates toll-like receptor pathways to influence cell proliferation and apoptosis in prostate cancer cells. Biol Chem. 2018;399(4):387–95.
5. Quan M, Chen J, Zhang D. Exploring the secrets of long noncoding RNAs. Int J Mol Sci. 2015;16(3):5467–96.
6. Wang W, et al. An immune-related lncRNA signature for patients with anaplastic gliomas. J Neurooncol. 2018;136(2):263–71.
7. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
8. Yin Y, et al. Molecular Signature to Risk-Stratify Prostate Cancer of Intermediate Risk. Clin Cancer Res. 2017;23(1):6–8.
9. Li J, et al. LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. Gut. 2014;63(11):1700–10.
10. Egevad L, et al. Prognostic value of the Gleason score in prostate cancer. BJU Int. 2002;89(6):538–42.
11. Humphrey PA. Gleason grading and prognostic factors in carcinoma of the prostate. Mod Pathol. 2004;17(3):292–306.
12. Shao N, et al. Development and validation of lncRNAs-based nomogram for prediction of biochemical recurrence in prostate cancer by bioinformatics analysis. J Cancer. 2019;10(13):2927–

34.

13. Xu J, et al. Transcriptome analysis reveals a long non-coding RNA signature to improve biochemical recurrence prediction in prostate cancer. *Oncotarget*. 2018;9(38):24936–49.
14. Shao N, et al. Identification of seven long noncoding RNAs signature for prediction of biochemical recurrence in prostate cancer. *Asian J Androl*. 2019;21(6):618–22.
15. Fan L, Li H, Wang W. Long non-coding RNA PRRT3-AS1 silencing inhibits prostate cancer cell proliferation and promotes apoptosis and autophagy. *Exp Physiol*. 2020;105(5):793–808.
16. Zhuang XH, Liu Y, Li JL. *Overexpression of long noncoding RNA HOXB-AS3 indicates an unfavorable prognosis and promotes tumorigenesis in epithelial ovarian cancer via Wnt/beta-catenin signaling pathway*. *Biosci Rep*, 2019. 39(8).
17. Jiang W, et al., *lncRNA HOXB-AS3 exacerbates proliferation, migration, and invasion of lung cancer via activating the PI3K-AKT pathway*. *J Cell Physiol*, 2020.
18. Huang HH, et al. Long non-coding RNA HOXB-AS3 promotes myeloid cell proliferation and its higher expression is an adverse prognostic marker in patients with acute myeloid leukemia and myelodysplastic syndrome. *BMC Cancer*. 2019;19(1):617.

## Tables

1 PCa patient clinicopathological features

Features	Patients (N=397)	
	n	%
years)	<60	153 38.54
	≥60	244 61.46
n score	≤7	215 54.16
	≥8	182 45.84
stage	T2	130 32.75
	T3-4	267 67.25
stage	N0	322 81.11
	N1	75 18.89
stage	M0	311 78.34
	M1	86 21.66
race	American indian or alaska native	1 0.25
	Asian	10 2.52
	Black or african american	46 11.59
	White	329 82.87
	Not reported	11 2.77
status	Alive	388 97.73
	Dead	9 2.27

Table 2 Overview for the four prognostic lncRNAs related to PCa patient OS.

gene	coef <sup>a</sup>	exp(coef) <sup>b</sup>	se(coef) <sup>c</sup>	z	Pr(> z )
HOXB-AS3	0.524052278	1.688857522	0.178121024	2.942113552	0.003259804
LINC01679	-0.663887578	0.514845935	0.249133131	-2.664790406	0.007703632
PRRT3-AS1	-0.504710478	0.603680329	0.197208039	-2.559279435	0.010488939
YEATS2-AS1	1.092940489	2.983032765	0.413609893	2.642442815	0.008231036

<sup>a</sup>Coef: coefficient; <sup>b</sup>exp(coef):hazard ratio; <sup>c</sup>se(coef):the range of values at hazard ratio

Table 3 OS-related univariate as well as multivariate Cox regression analysis

Predictors	Univariate analysis				Multivariate analysis			
	HR	HR.95L	HR.95H	pvalue	HR	HR.95L	HR.95H	pvalue
riskScore	1.011634053	1.00650523	1.01678901	8.18E-06	1.013586794	1.004722569	1.022529225	0.002601825
age(<60Y,≥60Y)	1.063724213	0.959190535	1.179650091	0.241797732	1.099128551	0.971013403	1.244147165	0.134973704
T stage(T2,T3-4)	2.140172886	0.567879531	8.065689517	0.260992904	0.613795488	0.117322963	3.211177866	0.563178046
N stage(N0,N1)	3.50500237	0.775413395	15.84321563	0.103209029	1.409992903	0.085715392	23.19396706	0.809959377
M stage(M0,M1)	1.714435831	0.330971655	8.880791385	0.520628659	0.874796352	0.053104922	14.4105034	0.925446246
gleasonScore(<8,≥8)	2.363916745	1.084202815	5.15411167	0.030522261	2.623777794	0.897852213	7.667419884	0.077895527

\*HR-hazard ratio

## Figures

499 tumor  
52 normal

551 prostate cancer in TCGA

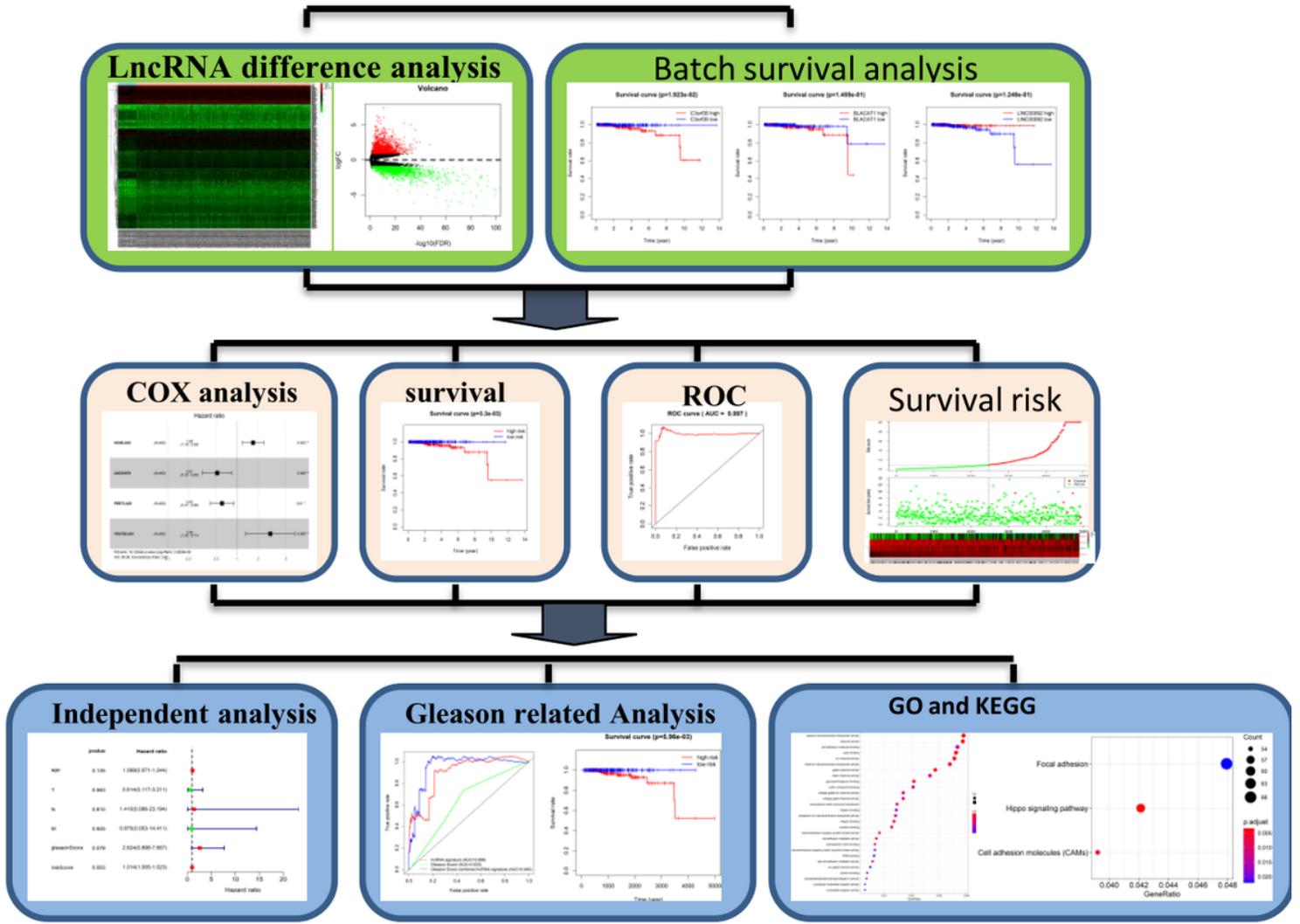


Figure 1

The overall analysis design of our study

# Volcano

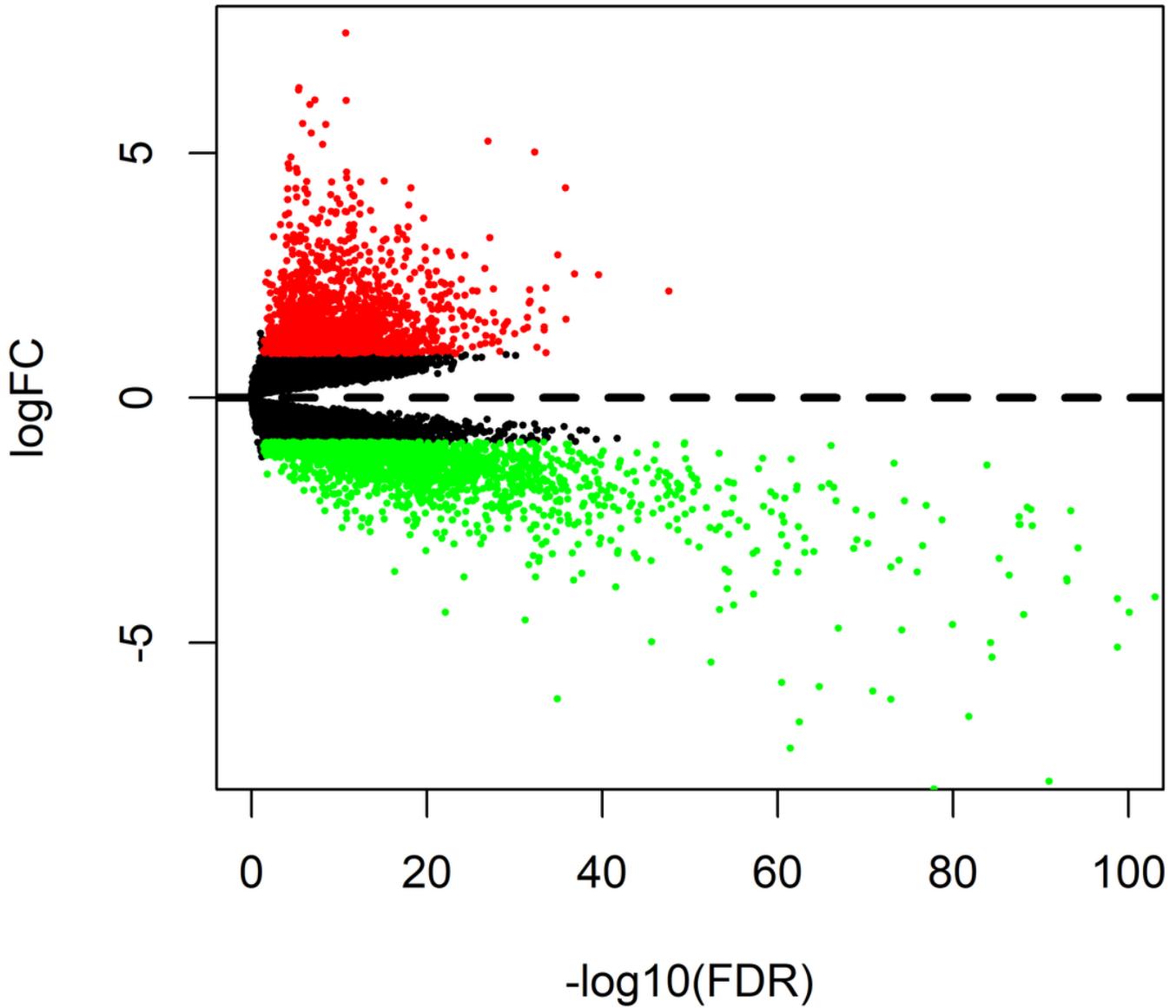
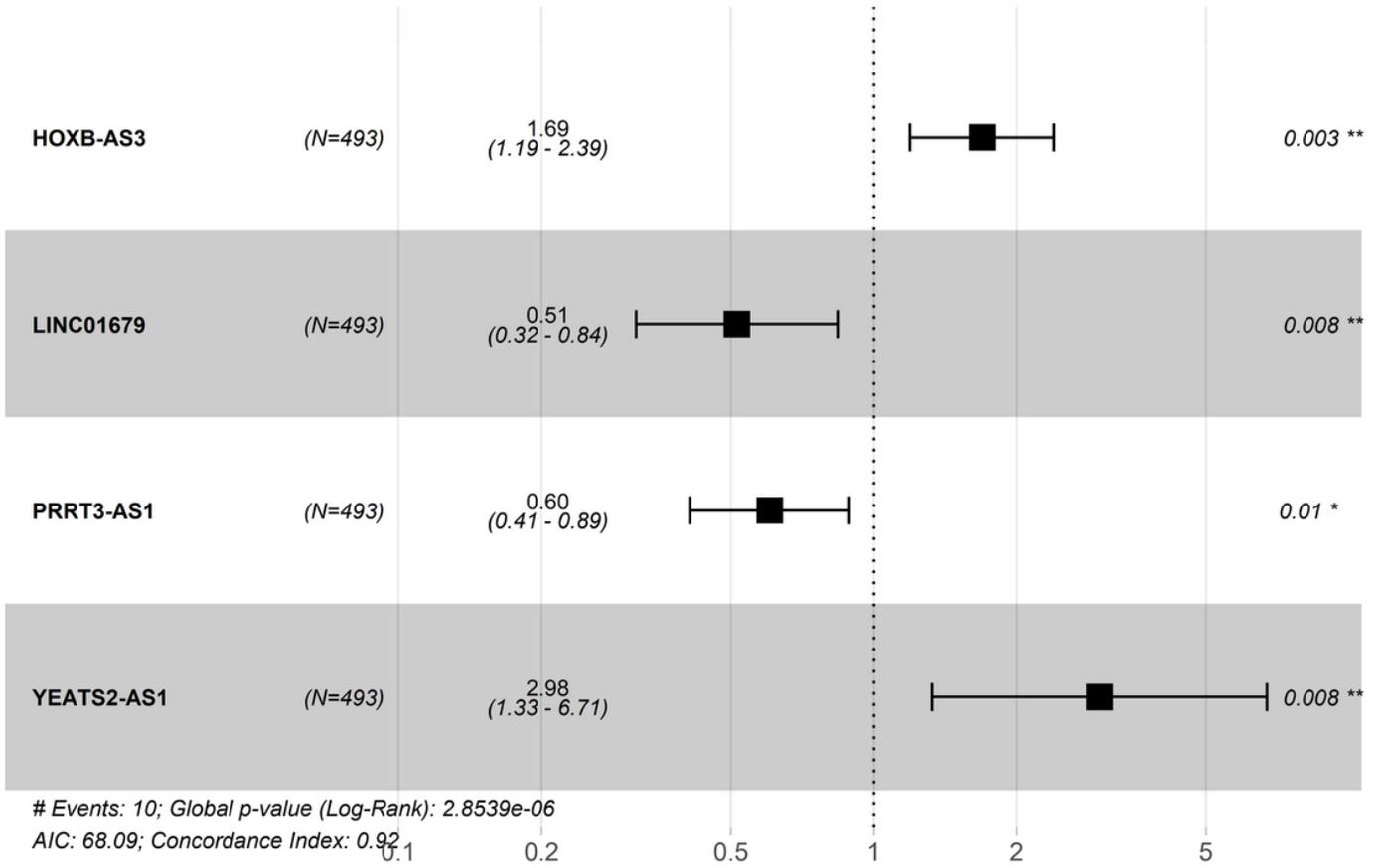


Figure 2

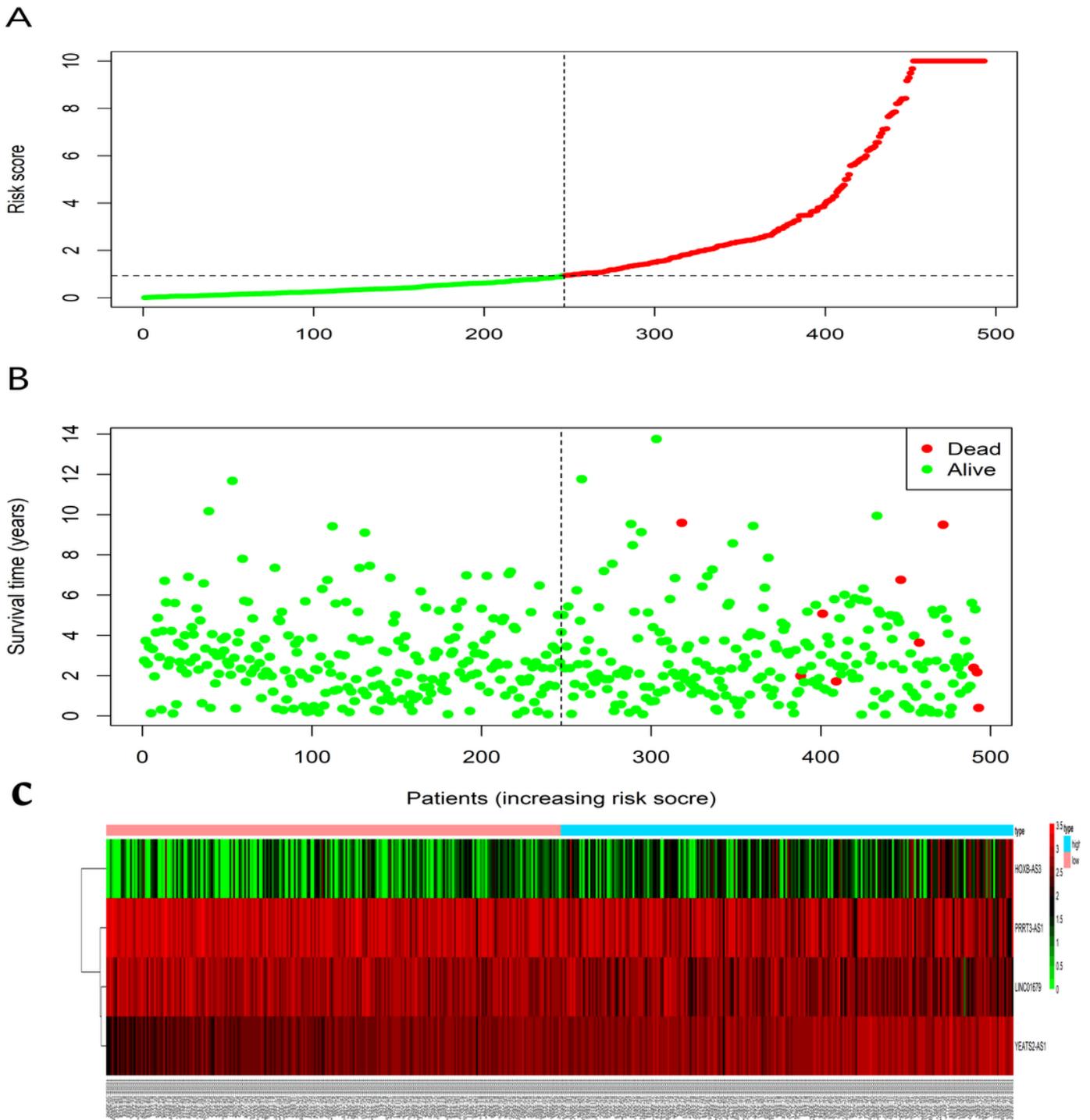
DE lncRNAs between prostate cancers and non-carcinoma tissue samples. Red dots represent differentially upregulated lncRNAs, green dots represent differentially downregulated lncRNAs and then dark dots represent no differentially expressed genes.

### Hazard ratio



**Figure 3**

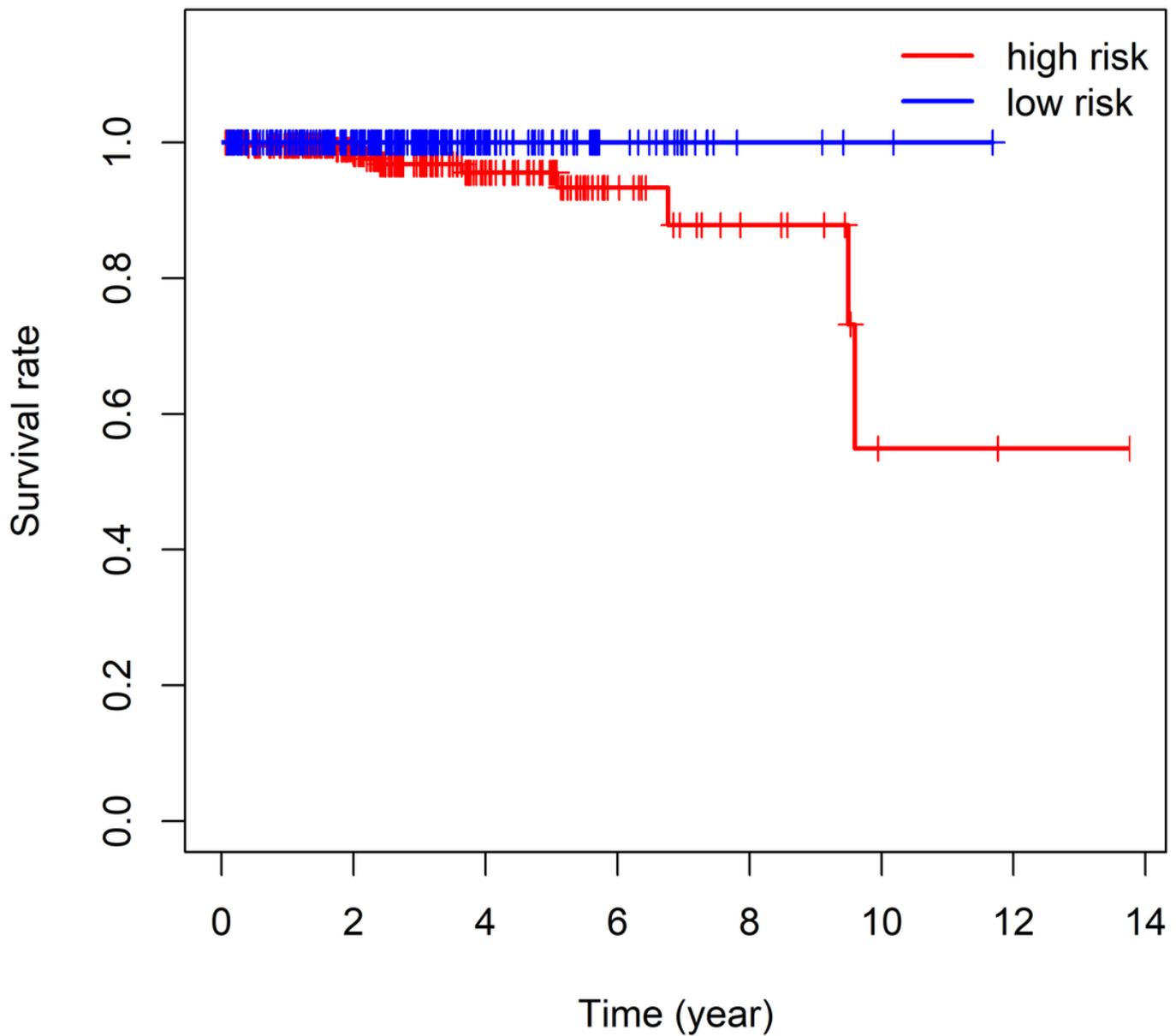
four lncRNAs Forest map formed by multivariate Cox regression analysis.



**Figure 4**

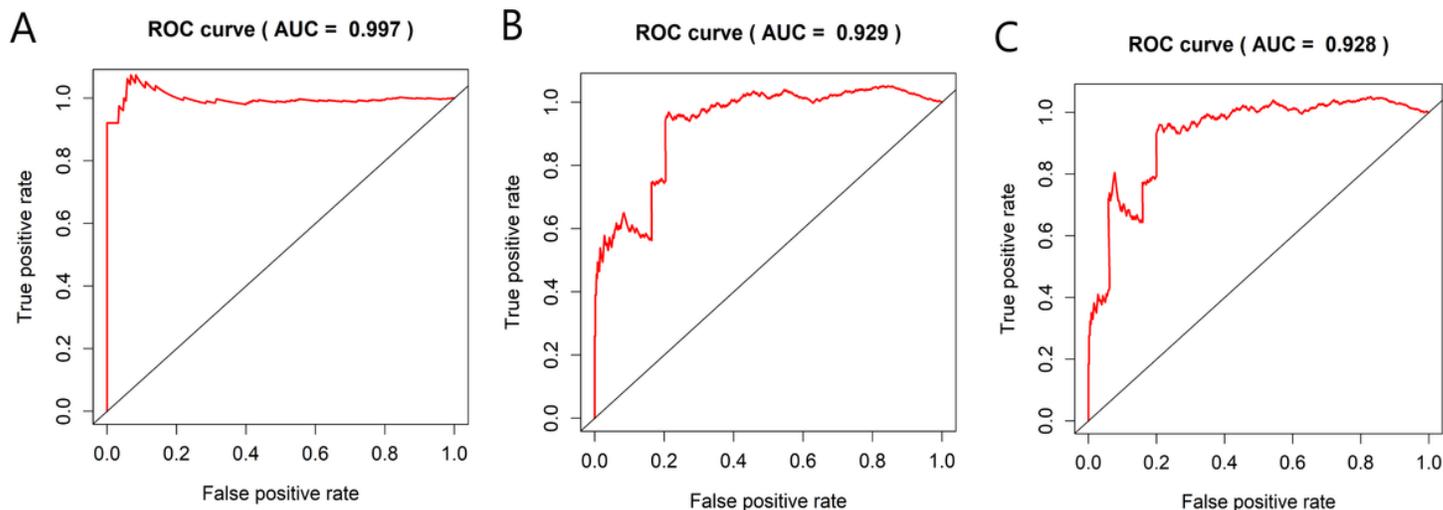
Ability of the 4 lncRNAs-based signature in predicting the prognosis for PCa cases. (A) Risk score distribution for patients. (B) PCa patient survival time. (C) Expression heat map for those four lncRNAs incorporated into the prognosis model. The vertical black dotted line stands for the optimum threshold to divide cases to high- or low-risk group.

## Survival curve (p=3.3e-03)



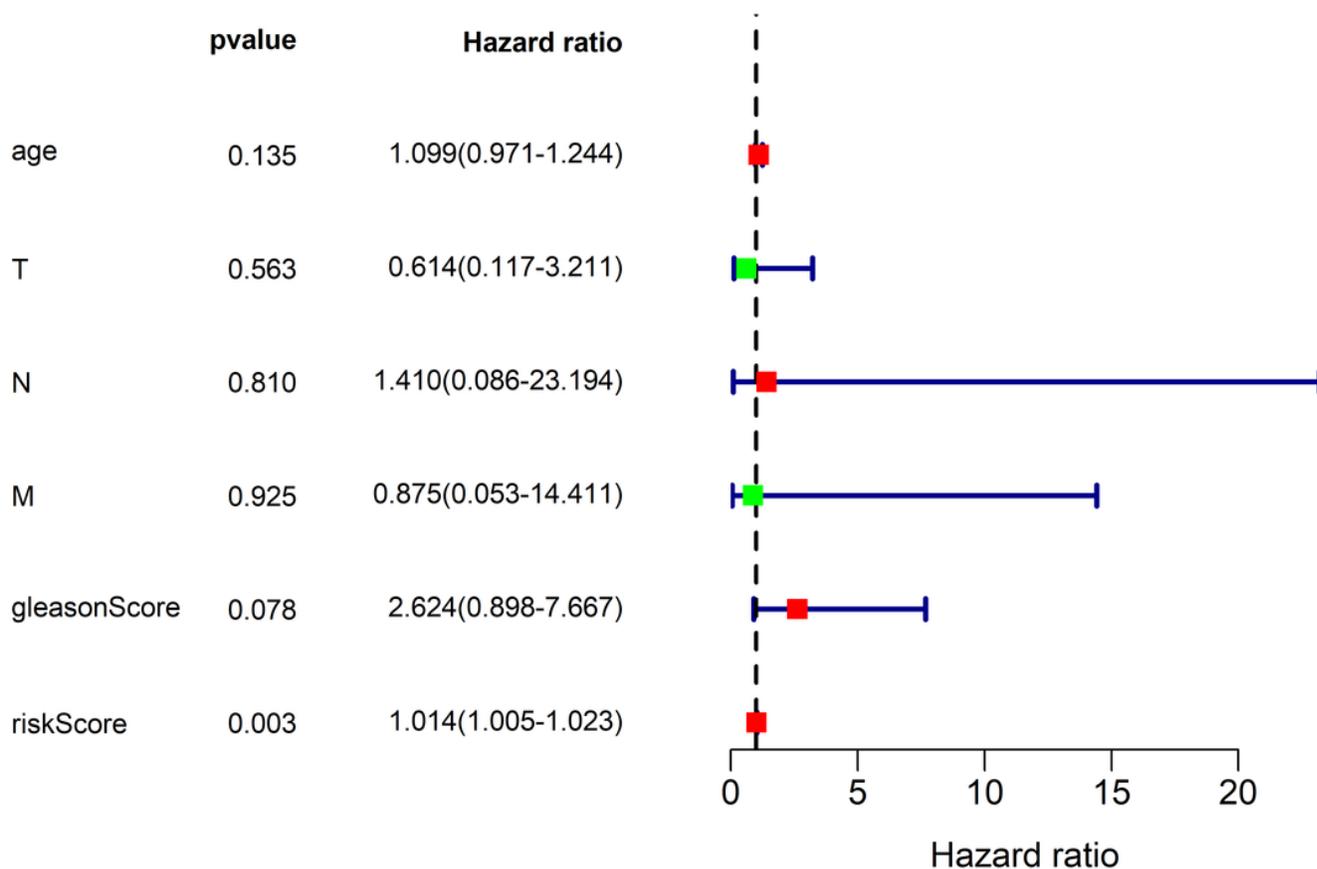
**Figure 5**

Prediction ability of the four-lncRNA nomogram evaluated based on Kaplan-Meier analysis, log rank p and C-index. OS time between low- and high-risk groups were visualized and compared by plotting the Kaplan-Meier curves.



**Figure 6**

Prognosis prediction ability of the four-lncRNA nomogram evaluated by ROC curves and dynamic AUC lines. The time-dependent ROC curves at 1 (A), 3 (B) and 5 (C) years during the follow-up period, together with dynamic AUC lines were drawn for the cases.



**Figure 7**

Multivariate analysis results demonstrated the possibility to use the 4-lncRNA signature as the potent predicting factor for PCa OS rate from other clinicopathological factors. Red represents high risk indicators and green represents low risk ones.

### Survival curve (p=2.359e-02)

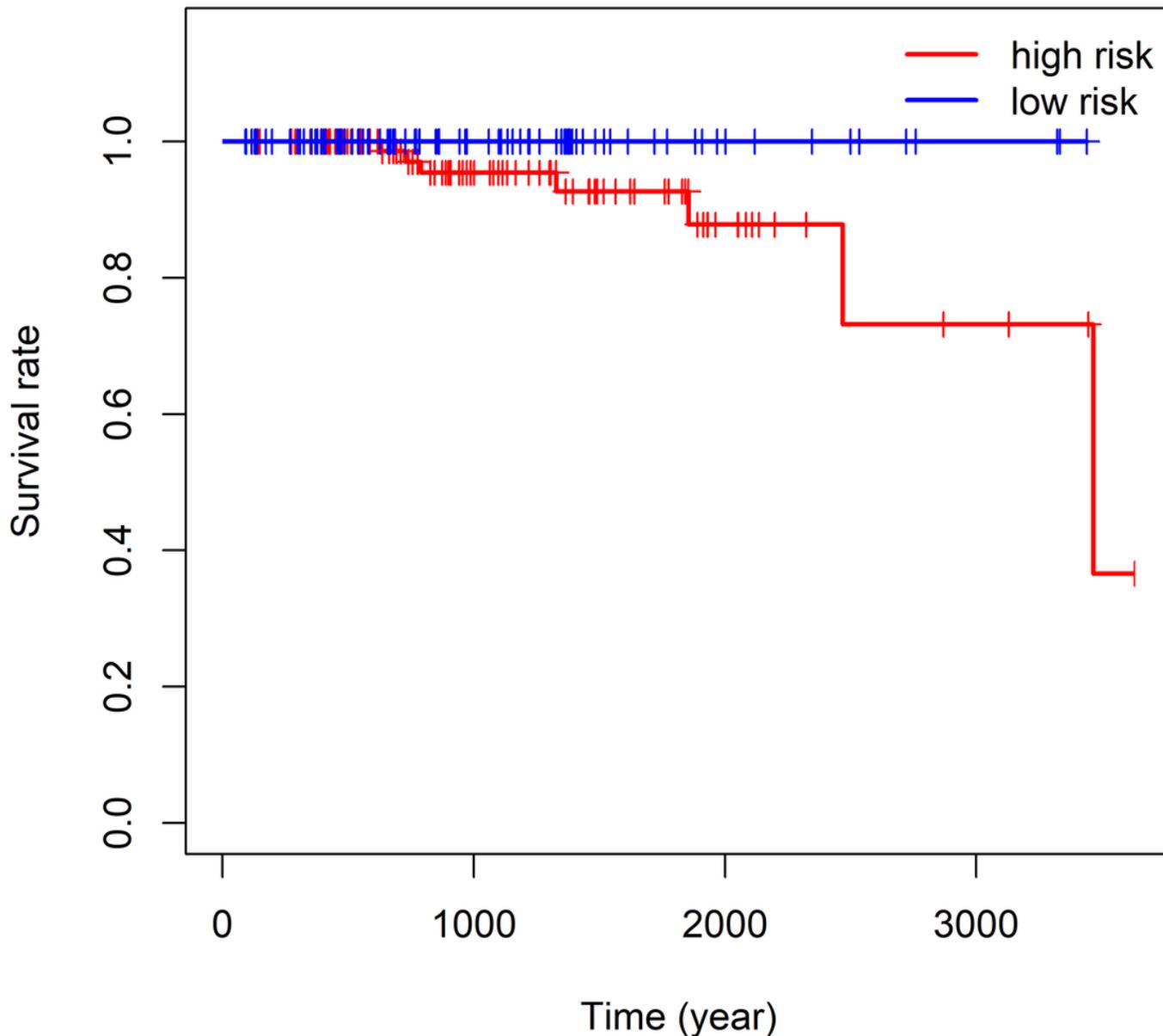
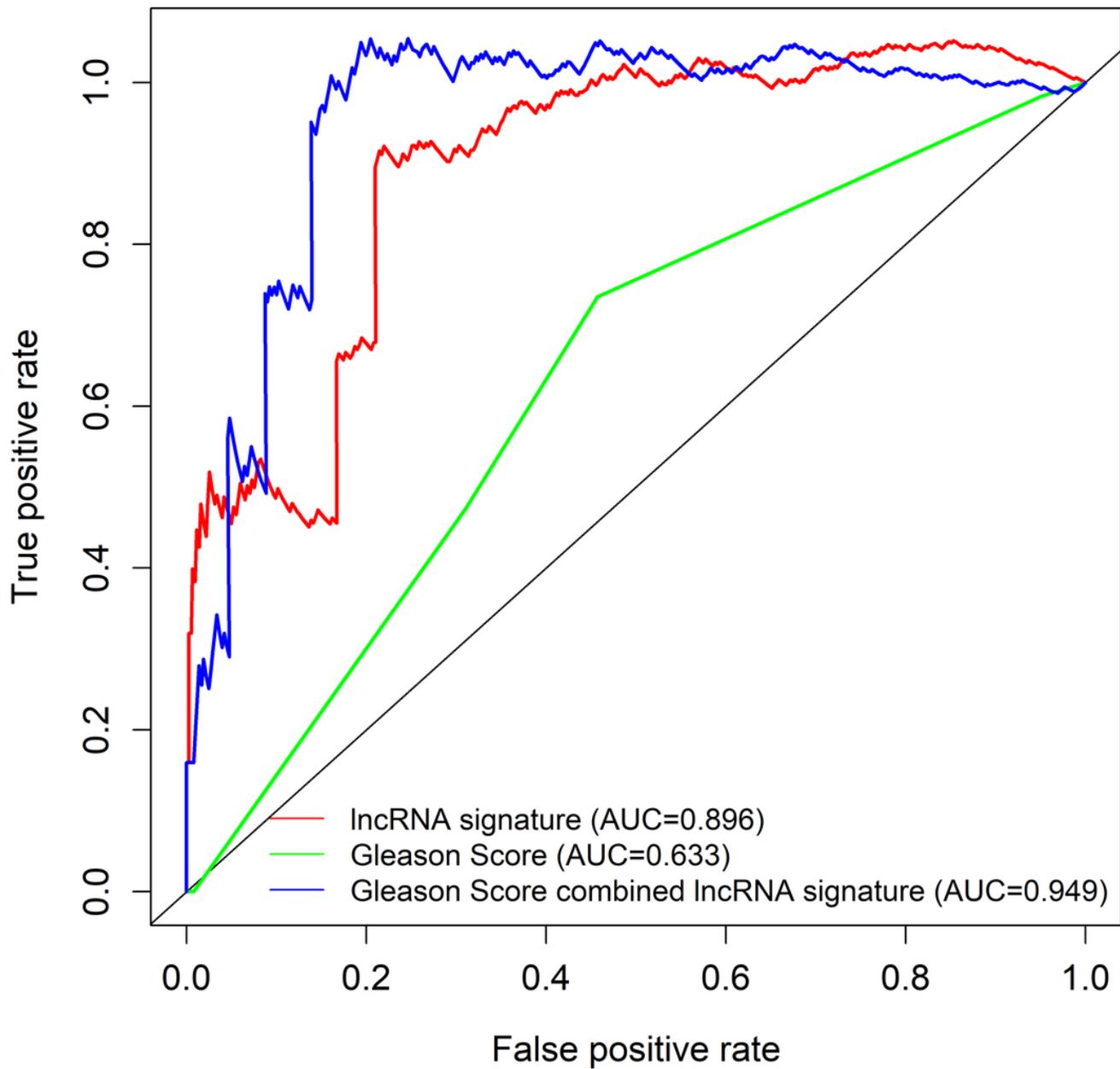


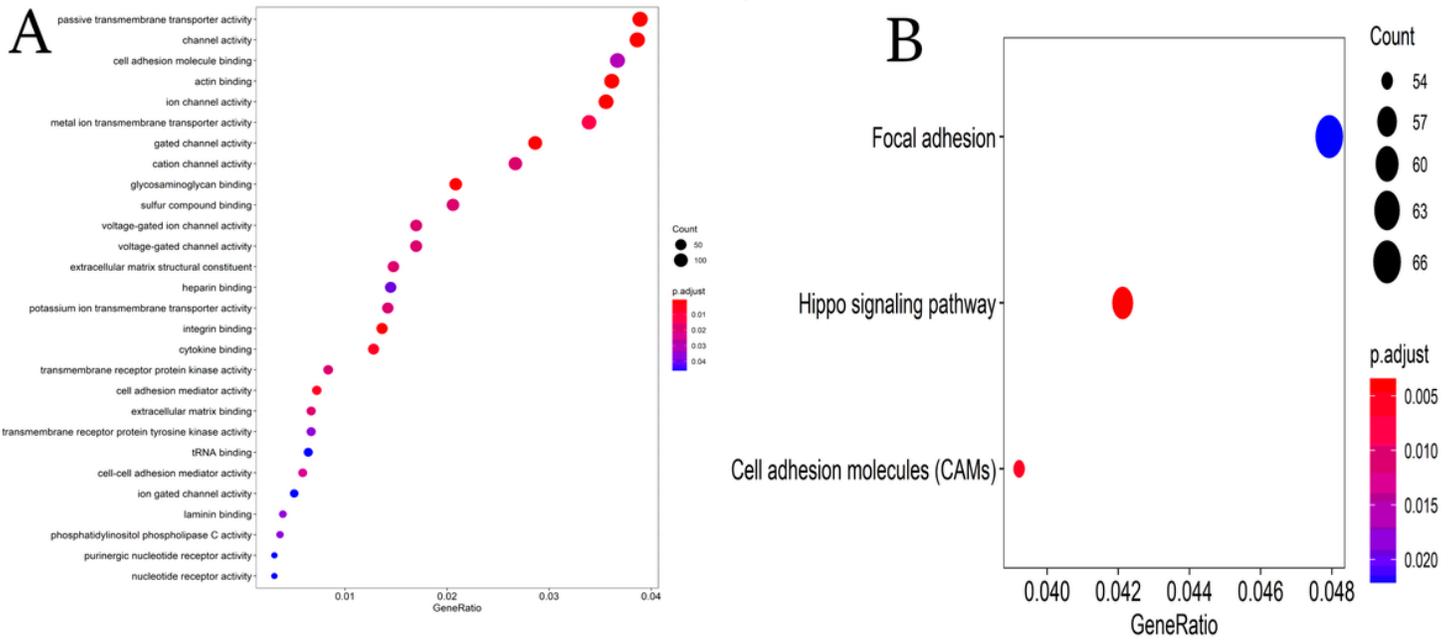
Figure 8

Kaplan-Meier survival analysis for PCa cases having the Gleason score >7. All cases were classified as high- or low-risk group using the four-lncRNAs model.



**Figure 9**

ROC curves for the prediction of the 3 years overall survival among the 4-IncRNA signature model, Gleason score and our IncRNA nomogram combined with Gleason score.



**Figure 10**

Functional enrichment analysis for the related biological processes and pathways related to the 4 lncRNAs used in that model. GO biological process enrichment results (A). KEGG signaling pathways analysis (B).

## Supplementary Files

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

- [Supplementaryfiguresandtableslegend.pdf](#)
- [TABLES1.xls](#)
- [TABLES2.xls](#)
- [TABLES3.xls](#)
- [TABLES4.xls](#)
- [TABLES5.xls](#)
- [FIGURES1.tiff](#)
- [FIGURES2.tiff](#)
- [FIGURES3A.png](#)
- [FIGURES3B.png](#)
- [FIGURES3C.png](#)
- [FIGURES4.tiff](#)