# Construction and validation of a novel eight-gene risk score to predict the progression and prognosis of bladder cancer

**Ruiliang Wang**
shanghai tenth people's hospital
**Zongtai Zheng**
shanghai tenth people's hospital
**Shiyu Mao**
Shanghai tenth's peoples hospital
**Wentao Zhang**
Shanghai tenth people's hospital
**Ji Liu**
Shanghai tenth people's hospital
**Cheng Li**
Shanghai tenth people's hospital
**Shenghua Liu** ( shenghualiu1990@163.com )
Shanghai Tenth People's Hospital, Tongji University School of Medicine https://orcid.org/0000-0003-0270-3343
**Xudong Yao**
Shanghai Tenth People's Hospital https://orcid.org/0000-0001-7234-3940

**Primary research**

**Keywords:** Bladder cancer, LASSO, WGCNA, risk score, nomogram, malignant progression, prognosis

**Posted Date:** January 27th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-120263/v2

# Abstract

**Background**: The progression from non-muscle-invasive bladder cancer (NMIBC) to muscle-invasive bladder cancer (MIBC) increases the risk of death. It is therefore important to find new relevant molecular models that will allow for effective prediction of the progression of bladder cancer (BC).

**Methods:** Using RNA-Sequence data of 49 BC patients in our center and weighted gene co-expression network analysis methods, a co-expression network of genes was developed from which three key modules with prognostic value were selected using Univariate Cox regression in The Cancer Genome Atlas Program (TCGA). Subsequently, an eight-gene risk score was established using the Least absolute shrinkage and selection operator Cox model.

**Results:** A novel eight-gene risk score was closely related to the malignant clinical features of BC and could predict the prognosis of patients in the training dataset (TCGA) and two validation sets (GSE3289 and GSE13507). Further, a nomogram for predicting the overall survival of patients was designed. The nomogram showed good calibration with clinical value through decision curve analysis. Lastly, we found that the mRNA and protein expression level of the eight genes were found to be differentially expressed in cell lines and tissue.

**Conclusions:** In our study, we established a novel eight-gene risk score which could predict the progression and prognoses of BC patients.


# 1 Introduction

Globally, bladder cancer (BC) is ranked 4$^{th}$ as the most commonly diagnosed cancer in men [1]. Among BC patients, 75% are diagnosed with non-muscle invasive bladder cancer (NMIBC) whereas 25% are diagnosed with muscle-invasive bladder cancer (MIBC). However, >30% of patients with NMIBC develop recurrence and progresses to MIBC within 5 years after diagnosis [2]. Moreover, three-quarters of MIBC patients develop distant metastases with their long-term survival rate being 15%[3].

Cystoscopy is considered the gold-standard diagnostic tool for BC [4]. However, its ability to predict disease progression had been determined [5]. Although some models have been developed to predict recurrence and progression of BC, the accuracy of such models is less than 60% [6, 7]. Although some biomarkers and risk models are available for predicting clinical outcome of patients with BC, their specificity and/or sensitivity is unsatisfactory [2, 8]. For this reason, it is important to identify valuable models or biomarkers for evaluating the prognosis and monitor the progression of BC.

In this study, a new prognostic risk model based on eight differently expressed genes between MIBC and NMIBC is established which can predict the progression of BC. The performance of the risk model was verified using gene expression data from the Gene Expression Omnibus (GEO). A prognostic nomogram based on the risk score was built to predict the prognosis of BC. Moreover, real-time quantitative (RT-

qPCR) and immunohistochemistry (IHC) were conducted to investigate the expression levels of eight genes in cell lines and BC samples.

# 2 Materials And Methods

## 2.1 Clinical samples and RNA sequencing

A total of 49 BC patients (35 NMIBC and 14 MIBC) in Shanghai Tenth People's Hospital (STPH) who underwent either transurethral resection of bladder tumor or radical cystectomy between November 2019 and April 2020, were included in the training set. The following criteria were used to enroll patients in the training set were: (1) histologically confirmed BC; (2) availability of freshly collected tissue from surgery; and (3) availability of clinical data. Informed consent was obtained and ethical approval was granted by the Ethics Committee of Shanghai Tenth People's Hospital. Total RNA was isolated from BC tissues using Trizol reagent (Invitrogen, Thermo Scientific, Shanghai, China). The cDNA libraries were constructed using a customized protocol and the Illumina Hiseq 2500 sequencer (Sangon Biotech, Shanghai, China).

## 2.2 Public data processing

Gene expression data from GSE32894 (n=224) and GSE13507 (n=165) datasets were downloaded from GEO (http://www.ncbi.nlm.nih.gov/geo) as a series of matrix file format that was earlier processed by the original authors using the MAS 5.0 algorithm. The latest TCGA data containing clinical features and follow-up information was downloaded using GDC API. The clinical data of the training sets (STPH and TCGA datasets) and validation sets (GSE32894 and GSE13507) are shown in **Table 1**. The representative immunohistochemistry (IHC) of genes in both normal human bladder tissues and tumor tissues were performed using data from the Human Protein Atlas (HPA) (http://www.proteinatlas.org/).

## 2.3 Establishment of prediction models and bioinformatics analyses

### 2.3.1 Weighted gene co-expression network analysis

The "limma" R package was adopted to screen for differentially expressed genes (DEGs) between 35 NMIBC and 14 MIBC samples in STPH. A weighted gene co-expression network analysis (WGCNA) was constructed [9] using the 'WGCNA" R package v1.68 based on the DEGs (P-value<0.05 and |log2FC|≥0.5). WGCNA was performed as follows: (1) Outlier samples were omitted to increase the reliability of the weighted gene co-expression network, (2) An appropriate β value was selected using 0.85 as the degree of independence ($R^2$), (3) A weighted adjacent matrix was transformed into a topological overlap matrix (TOM) to determine the network connectivity of the genes, (4) Genes with similar expression profiles were classified into gene modules based on the average linkage hierarchical clustering following the TOM-based dissimilarity measure, (5) All genes were represented by the expression of module eigengenes (MEs) in a given module. Modules that were highly correlated with NMIBC/MIBC subtype (|r| ≥0.3) were selected for further analyses, and (6) The genes in the selected modules (brown, turquoise, and yellow) were extracted.

### 2.3.2 Construction of a novel eight-gene risk score

Based on hub genes extracted from WGCNA, univariate Cox regression analysis was conducted to select genes associated with overall survival (OS) in the TCGA dataset [10]. Thereafter, the genes were subjected to Least absolute shrinkage and selection operator (LASSO) analysis with ten-fold cross-validation to construct a novel eight-genes risk score using “glmnet” package. Genes with minimal influence on the patient’s overall survival were removed whereas those with non-zero coefficients were selected. The risk score for each patient was calculated as follows: risk score = $Coef_1$ × expression of $gene_1$ + $Coef_2$ × expression of $gene_2$ + … + $Coef_m$ × expression of gene m. Coef denotes the corresponding coefficient of the gene. Using these gene expression data, the ‘survminer’ R package v4.6 was used to evaluate the optimal cutoff value of the risk score in each cohort. The established optimal cutoff value was used to divide patients into low- and high-risk groups accordingly.

### 2.3.3 Gene sets enrichment analysis (GSEA)

GSEA (http://www.broadinstitute.org/gsea/index.jsp) was conducted to compare pathways between patients in high- and low-risk groups. Signaling pathways with $P<0.05$ and a false discovery rate $<0.25$ were regarded as statistically significant.

### 2.3.4 Gene Set Cancer Analysis

Gene Set Cancer Analysis (GSCALite, http://bioinfo.life.hust.edu.cn/web/GSCALite/) provides a single nucleotide variation (SNV) module through the maftools [11]. Herein, we employed the SNV module to analyze and visualize the SNV of the eight genes in BC.

### 2.4 Eight-Gene Risk score validation

To evaluate the generalizability of the model, the training sets (STPH and TCGA dataset) and validation datasets (GSE3289 and GSE13507) were analyzed using similar formula and coefficients. The Kaplan-Meier curves of the validation datasets were separately drawn before analyzing the prognostic performance of the novel eight-genes risk score. Furthermore, the receiver operating characteristic (ROC) curve and Violin plot were used to verify the relationship between the risk score and clinical features of MIBC.

### 2.5 Development of the nomogram

A nomogram was developed from the factors that were significant in the final multivariate Cox regression analyses. The ROC, concordance index (C-index) and calibration plots were conducted to assess the performance of the nomogram. The nomogram, C-index and calibration plots were generated using the ‘Rms’ R package. Furthermore, a decision curve analysis (DCA) was employed to ascertain the net benefits of nomogram and other crucial prognostic factors.

### 2.6 Cell culture

The human normal bladder epithelial cell line SV-HUC-1 and BC cell line T24, UMUC3, J82, 5637, EJ obtained from Shanghai Cell Bank of the Chinese Academy of Sciences (Shanghai, China) were used for *in vitro* experiments. The SV-HUC-1 cell lines were cultured in F12K containing 10% FBS and 1% penicillin/streptomycin (P/S). BC cell lines were cultured in RPMI-1640 (Gibco; Thermo Fisher Scientific, Inc.) supplemented with 10% FBS (Gibco; Thermo Fisher Scientific, Inc.) and 1% P/S (Gibco; Thermo Fisher Scientific, Inc.). Cell culture reagents were bought from Gibco and the cells were cultured at 37˚C and 5% $CO_2$.

### 2.7 RT-qPCR

TRIzol reagent (Invitrogen, Thermo Scientific, Shanghai, China) was used for extracting RNA from the cell lines. The cDNA Synthesis SuperMix Kit (Cat No.11141ES60; Yeasen, Shanghai. China) was used to synthesize cDNA which was subjected to qPCR using qPCR SYBR Green Master Mix KIT (Cat No. 11203ES03; Yeasen, Shanghai. China). The gene expression level was normalized to the expression of GAPDH mRNA. Primer sequences used are presented in **Table SI**.

### 2.8 Statistical analysis

The Wilcoxon rank-sum test was used to identify the differential expressed genes between NMIBC and MIBC. The Student's t-test or one-way ANOVA was used to evaluate the association of the risk score with clinical characteristics. Kaplan-Meier and log-rank tests were employed to compare the OS of BC patients in the high- and low-risk groups. The prediction performance of the risk score in NMIBC/MIBC subtype was determined from ROC curves. A heatmap and correlation matrix were generated using 'Pheatmap' R package v1.0.12 and 'corrplot' R package v0.84, respectively. SPSS 22.0 (SPSS, Armonk, NY, USA), R v3.6.1 (https://www.r-project.org/) and Graphpad Prism V7 (GraphPad Software, Inc.) were used for data analysis. A two-sided P value<0.05 was considered significant.

## 3 Results

### 3.1 Data acquisition and DEGs selection

A flow diagram of the study is shown **Figure 1**. A total of 49 BC patients (35 NMIBC and 14 MIBC) from STPH were enrolled for the study. From our findings, 2725 differentially expressed genes were identified between NMIBC and MIBC.

### 3.2 WGCNA analysis based on DEGs between NMIBC and MIBC

After the hierarchical clustering analysis, one sample was removed as an outlier. A co-expression network was constructed using 48 BC samples with complete clinical T stage and NMIBC/MIBC subgroups **(Figure 2A)**. With the chosen power of β=13 (scale-free $R^2$=0.85) as the soft-thresholding **(Figure 2B)**, seven modules were obtained **(Figure 2C)**. The highest module-trait association was found between three

modules and the NMIBC/MIBC subgroups (brown, turquoise, and yellow) **(Figure 2D)**. Subsequently, 1351 genes from three key modules were selected for further analysis.

### 3.3 Construction and verification of the novel eight-genes risk score

Among the 1315 genes, 159 were statistically significant (P-values < 0.05) in the univariate Cox analysis, hence were selected for LASSO Cox regression analysis. The Lambda.1se penalty parameter was selected based on the ten-fold leave-one-out cross-validation in accordance with the minimum criteria. In this way, eight genes (*CD96, PDCL3, IP6K2, TRIM38, U2AF1L4, DDB1, KCNJ15 and CTU1*) with nonzero coefficients were obtained **(Figure 3A-C)**. The prognostic performance of the genes was assessed individually. Results showed that their prognostic values were statistically significant (**Figure S1A**). According to the aforementioned formula, risk scores for each patient were calculated. Kaplan-Meier analysis revealed that the high risk group showed significantly poorer OS than BC patients in the low risk group (P<0.001) **(Figure 3D)**. We further validated the prognostic value of the risk score in two validation sets. Interestingly, the risk scores still had the prognostic ability in GSE3289 (P=0.035) and GSE13507 (P=0.037) **(Figure 3E, F)**. Among the 411 BC samples with complete sequencing data in TCGA, mutations of the eight genes were only found in 29 independent samples in TCGA dataset. Notably, *CD96* and *DDB1* recorded the highest rates of mutations (28%) (**Figure S1B**). GSEA analysis showed that high risk score was associatd with the 'Focal adhersion', 'pathway in cancer', 'WNT signal', 'bladder tumor', 'ECM receptor' and 'adherens junctions pathway' **(Figure S2)**.

### 3.4 Association of the novel eight-gene risk score with clinical features

Patients with high risk scores showed positively association with clinicopathological features, including age (P<0.001), pathological stage (P<0.001), histological grade (P<0.01), T stage (P<0.001), M stage (P<0.01), N stage (P<0.001) and morphology (P<0.001) in TCGA dataset **(Figure 4A)**. A dot plot based on the survival status of TCGA set indicated that patients with high-risk scores had a higher mortality rate than those with low-risk scores (**Figure 4B**). The ROC revealed that the risk score could effectively differentiate NMIBC from MIBC with an AUC of 0.903 **(Figure 4C)**. The distribution of risk scores in training and validation sets are presented in detail in **Figure 4D**. It can be observed that MIBC has higher score than NMIBC (all P<0.05).

### 3.5 Construction and calibration of the nomogram

The univariate Cox analyses revealed that T stage, histological grade, age and risk score were significantly related to OS in the TCGA dataset **(Figure 5A)**. After multivariate Cox analysis, risk score, age, T stage, and pathological stage retained their prognostic value. A nomogram was constructed for predicting the 1-, 3- and 5-year OS based on the significant factors in multivariate Cox analyses **(Figure 5B)**. The AUC values of ROC curves were 0.764 (1-year ROC), 0.796 (3-year ROC) and 0.805 (5-year ROC), whereas the C-index of the nomogram was 0.75±0.02 (mean±SD) for OS **(Figure 5C)**. Besides, calibration plots proved that the prediction results of the nomogram were largely in agreement with the actual 1-, 3-

and 5-year OS **(Figure 5D)**. Furthermore, compared with clinical features, the nomogram achieved the best clnical net benefit via DCA plot (**Figure 5E**).

### 3.6 Validation of the novel eight-genes score through *in vitro* experiments and public database

The mRNA expression pattern of the eight genes in five human BC cell lines (T24, UMUC3, 5637, J82 and EJ) and one human normal bladder epithelial cell line (SV-HUC-1) was determined using RT-qPCR. Results showed that five of eight genes (*CTU1, IP6K2, KCNJ15, PDCL3* and *U2AF1L4*) were significantly different in transcriptional levels between BC and normal bladder cells **(Figure 6)**. The expression level of *IP6K2* and *KCNJ15* was lower in all cancer cell lines compared with normal cell lines. The expression of *CTU1, PDCL3* and *U2AF1L4* was varied in different BC cell lines. IHC images of normal and tumor tissues from HPA revealed that *DDB1, CD96, IP6K2, TRIM38* and *KCNJ15* had higher IHC staining intensity in normal tissues compared with tumor tissues. *PDCL3* and *U2AF1L4* showed medium to high in of IHC staining intensity in normal and tumor tissues. Neither normal tissues nor tumor tissues expressed the CUT1 protein (Figure 6B).

# 4 Discussion

Recent developments in the field of high-throughput sequencing technologies have resulted in the generation of many bioinformatics methods to find biomarkers related to the progression of malignant tumors [12, 13]. So far, several prediction models have been developed for BC [14-16]. However, these models are based on public datasets without validation set or data from their clinical centers, which limits the generalizability and reproducibility of the models. Through an in-depth exploration of the RNA-sequence data from our center and TCGA, we constructed an eight-genes risk score using WGCNA and LASSO regression analyses The results showed that the eight-genes risk score had novel performance in prognostic stratification in training (TCGA) and validation (GSE3289 and GSE13507) data sets, in which BC patients in the high risk group showed significantly poorer OS than BC patients in the low risk group. Taken together, the risk score presented an oncogenic function in malignant progression and was considered as a useful prognostic tool for prognostic prediction in BC.

GSEA analysis confirmed that high risk scores were highly enriched in ‘WNT signal’ and ‘bladder tumor’ pathways, suggesting the important role of the risk score in tumorigenesis and malignant progression. The GSCA results further indicated that these genes rarely mutated, implying that potential dysfunction in these genes may not be as a result of genetic mutations, but rather due to aberrant alterations at the transcriptional level [17]. The genes in the risk score were reported to play important roles in tumor in previous studies. For example, high expression of *CD96* in BC tissue, that was discovered in 1992, indicates high immune cells infiltration and hence killing of cancer cells [18, 19]. Damage-specific DNA binding protein 1 (DDB1), as the only oncogene in our model, is a multifunctional protein. Previous studies showed that DDB1 interacts with several chaperone that regulate DNA repair mechanisms, cell cycle and gene transcription [20]. DDB1 can directly interact with Cullin4B (CUL4B) to form the CUL4B-DDB1 complex which promotes ubiquitination and degradation of a variety of tumor

suppressor factors hence enhance progression of osteosarcoma [21]. The tripartite motif protein 38 (TRIM38), a member of the TRIM family, was found to confer protection against various cellular processes such as cell differentiation, proliferation, apoptosis, and antiviral defense. On the other hand, TRIM38 cause abnormal activation of NF-κB pathway thereby inhibiting cancer [22, 23].

We then integrated the risk score and other independent prognostic factors to establish a nomogram for clinical practice. A nomogram that comprised the age, T stage, pathological stage and risk score was drawn to predict the OS. ROC showed that the nomogram could accurately predict the 1-, 3- and 5-year OS of BC. Moreover, calibration plots and C-index confirmed the prognostic significance and predictive superiority of the nomogram. DCA revealed that compared with clinical features, the nomogram had the optimal enhanced clinical net benefit with wider threshold probabilities. All of these findings indicated that the risk score-based nomogram may be a reliable evaluation tool to perform mortality risk identification in BC patients.

The results of IHC assay in HPA and RT-qPCR revealed that hub genes were differentially expressed between normal and BC tissues or cells. Nevertheless, the mRNA expression level (RT-qPCR in cell lines) and protein intensity (IHC in tissues) may not match completely due to the following reasons: (a) the tumor microenvironment between tumor tissues and cell lines was different. The occurrence and development of tumors is as a result of complex interactions [24, 25]. (b) The regulation of mRNA and protein expression is complex, and there is no clear relationship between mRNA and protein expression levels.

Some limitations exist in the present study. The sample size was small, and more samples with prognostic information are needed to validate our findings. Further researches are needed to explore the mechanisms of the eight genes in tumor progression.

# 5 Conclusion

This study demonstrates that a novel eight-genes risk score can reliably predict the malignant progression and prognosis of BC patients. The constructed nomogram based on the eight-genes risk score could accurately predict the survival of BC patients. Dysregulated expression of the eight genes was validated in BC samples and cell lines by IHC and RT-qPCR, respectively. Further functional studies and mechanistic experiments should be conducted to better understand the clinical value of the eight hub genes in BC.

# Declarations

### Ethics declarations

The present study was approved by the Ethics Committee of Shanghai Tenth People's Hospital (Shanghai, China). The research was conducted in accordance with the World Medical Association Declaration of Helsinki. All patients provided written informed consent.

### Availability of data and materials

All datasets used during the study are available from the corresponding author upon reasonable request. Public data that support the findings of this study are openly available in TCGA (http://cancergenome.nih.gov/) and NCBI GEO dataset (https://www.ncbi.nlm.nih.gov/).

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.


### Founding

This study was funded by the Shanghai Science Committee Foundation (grant number 19411967700) and the Natural Science Foundation of China (grant number 81472389).


### Authors' contributions

R.W. and Z.Z. conceived and designed the study. X.Y. and S.L. acquired the funding. R.W., S.M. and W.Z. collected and collated the data. All authors analyzed and interpreted the data. R.W. wrote the manuscript, Z.Z., S.M. and W.Z. critically reviewed and revised the manuscript. J.L., C.L., and Z.Z. designed the tables and figures. All authors read and approved the manuscript and agreed to be accountable for all aspects of the research.


### Acknowledgements

We would like to thank other members of the Department of Urology for their assistance and invaluable support.


## References


1. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2020**. *CA Cancer J Clin* 2020, **70**(1):7-30.
2. Dyrskjot L, Reinert T, Algaba F, Christensen E, Nieboer D, Hermann GG, Mogensen K, Beukers W, Marquez M, Segersten U *et al*: **Prognostic Impact of a 12-gene Progression Score in Non-muscle-invasive Bladder Cancer: A Prospective Multicentre Validation Study**. *Eur Urol* 2017, **72**(3):461-469.
3. Abufaraj M, Dalbagni G, Daneshmand S, Horenblas S, Kamat AM, Kanzaki R, Zlotta AR, Shariat SF: **The Role of Surgery in Metastatic Bladder Cancer: A Systematic Review**. *Eur Urol* 2018, **73**(4):543-557.
4. Witjes JA, Bruins HM, Cathomas R, Comperat EM, Cowan NC, Gakis G, Hernandez V, Linares Espinos E, Lorch A, Neuzillet Y *et al*: **European Association of Urology Guidelines on Muscle-invasive and**

**Metastatic Bladder Cancer: Summary of the 2020 Guidelines**. *Eur Urol* 2020.

5. Rose TL, Lotan Y: **Advancements in optical techniques and imaging in the diagnosis and management of bladder cancer**. *Urol Oncol* 2018, **36**(3):97-102.
6. Xylinas E, Kent M, Kluth L, Pycha A, Comploj E, Svatek RS, Lotan Y, Trinh QD, Karakiewicz PI, Holmang S *et al*: **Accuracy of the EORTC risk tables and of the CUETO scoring model to predict outcomes in non-muscle-invasive urothelial carcinoma of the bladder**. *Br J Cancer* 2013, **109**(6):1460-1466.
7. Kluth LA, Black PC, Bochner BH, Catto J, Lerner SP, Stenzl A, Sylvester R, Vickers AJ, Xylinas E, Shariat SF: **Prognostic and Prediction Tools in Bladder Cancer: A Comprehensive Review of the Literature**. *Eur Urol* 2015, **68**(2):238-253.
8. Dyrskjot L, Zieger K, Real FX, Malats N, Carrato A, Hurst C, Kotwal S, Knowles M, Malmstrom PU, de la Torre M *et al*: **Gene expression signatures predict outcome in non-muscle-invasive bladder carcinoma: a multicenter validation study**. *Clin Cancer Res* 2007, **13**(12):3545-3551.
9. Shi J, Zhang P, Liu L, Min X, Xiao Y: **Weighted gene coexpression network analysis identifies a new biomarker of CENPF for prediction disease prognosis and progression in nonmuscle invasive bladder cancer**. *Mol Genet Genomic Med* 2019, **7**(11):e982.
10. Li Y, Lu S, Lan M, Peng X, Zhang Z, Lang J: **A prognostic nomogram integrating novel biomarkers identified by machine learning for cervical squamous cell carcinoma**. *J Transl Med* 2020, **18**(1):223.
11. Mayakonda A, Koeffler HP: **Maftools: Efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies**. 2016.
12. Yan X, Fu X, Guo ZX, Liu XP, Liu TZ, Li S: **Construction and validation of an eight-gene signature with great prognostic value in bladder cancer**. *J Cancer* 2020, **11**(7):1768-1779.
13. Zheng Z, Mao S, Guo Y, Zhang W, Liu J, Li C, Yao X: **N6methyladenosine RNA methylation regulators participate in malignant progression and have prognostic value in clear cell renal cell carcinoma**. *Oncol Rep* 2020, **43**(5):1591-1605.
14. Wang C, Yang Y, Yin L, Wei N, Hong T, Sun Z, Yao J, Li Z, Liu T: **Novel Potential Biomarkers Associated With Epithelial to Mesenchymal Transition and Bladder Cancer Prognosis Identified by Integrated Bioinformatic Analysis**. *Front Oncol* 2020, **10**:931.
15. Lian P, Wang Q, Zhao Y, Chen C, Sun X, Li H, Deng J, Zhang H, Ji Z, Zhang X *et al*: **An eight-long non-coding RNA signature as a candidate prognostic biomarker for bladder cancer**. *Aging (Albany NY)* 2019, **11**(17):6930-6940.
16. Pan S, Zhan Y, Chen X, Wu B, Liu B: **Identification of Biomarkers for Controlling Cancer Stem Cell Characteristics in Bladder Cancer by Network Analysis of Transcriptome Data Stemness Indices**. *Front Oncol* 2019, **9**:613.
17. Hu MM, Yang Q, Zhang J, Liu SM, Zhang Y, Lin H, Huang ZF, Wang YY, Zhang XD, Zhong B *et al*: **TRIM38 inhibits TNFalpha- and IL-1beta-triggered NF-kappaB activation by mediating lysosome-dependent degradation of TAB2/3**. *Proc Natl Acad Sci U S A* 2014, **111**(4):1509-1514.

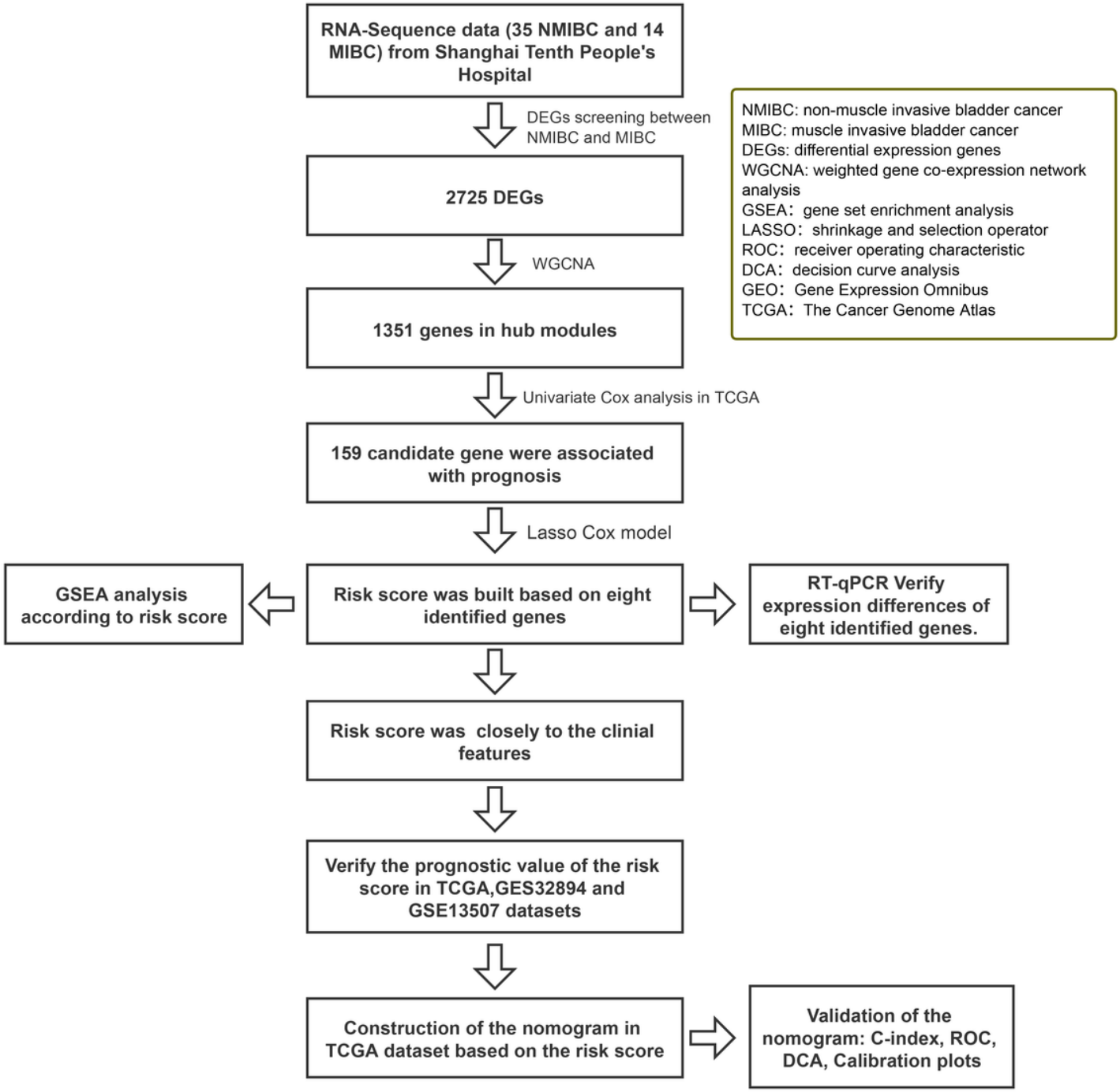
18. Georgiev H, Ravens I, Papadogianni G, Bernhardt G: **Coming of Age: CD96 Emerges as Modulator of Immune Responses**. *Front Immunol* 2018, **9**:1072.
19. Blake SJ, Dougall WC, Miles JJ, Teng MW, Smyth MJ: **Molecular Pathways: Targeting CD96 and TIGIT for Cancer Immunotherapy**. *Clin Cancer Res* 2016, **22**(21):5183-5188.
20. Srinivasan S, Chitalia V, Meyer RD, Hartsough E, Mehta M, Harrold I, Anderson N, Feng H, Smith LE, Jiang Y *et al*: **Hypoxia-induced expression of phosducin-like 3 regulates expression of VEGFR-2 and promotes angiogenesis**. *Angiogenesis* 2015, **18**(4):449-462.
21. Chen B, Feng Y, Zhang M, Cheng G, Chen B, Wang H: **Small molecule TSC01682 inhibits osteosarcoma cell growth by specifically disrupting the CUL4B-DDB1 interaction and decreasing the ubiquitination of CRL4B E3 ligase substrates**. *Am J Cancer Res* 2019, **9**(9):1857-1870.
22. Kim K, Kim JH, Kim I, Seong S, Kim N: **TRIM38 regulates NF-kappaB activation through TAB2 degradation in osteoclast and osteoblast differentiation**. *Bone* 2018, **113**:17-28.
23. Hu MM, Xie XQ, Yang Q, Liao CY, Ye W, Lin H, Shu HB: **TRIM38 Negatively Regulates TLR3/4-Mediated Innate Immune and Inflammatory Responses by Two Sequential and Distinct Mechanisms**. *J Immunol* 2015, **195**(9):4415-4425.
24. Li F, Teng H, Liu M, Liu B, Zhang D, Xu Z, Wang Y, Zhou H: **Prognostic Value of Immune-Related Genes in the Tumor Microenvironment of Bladder Cancer**. *Front Oncol* 2020, **10**:1302.
25. Zhou X, Qiu S, Nie L, Jin D, Jin K, Zheng X, Yang L, Wei Q: **Classification of Muscle-Invasive Bladder Cancer Based on Immunogenomic Profiling**. *Front Oncol* 2020, **10**:1429.

# Tables

Table 1: Characteristic of datasets used in this study

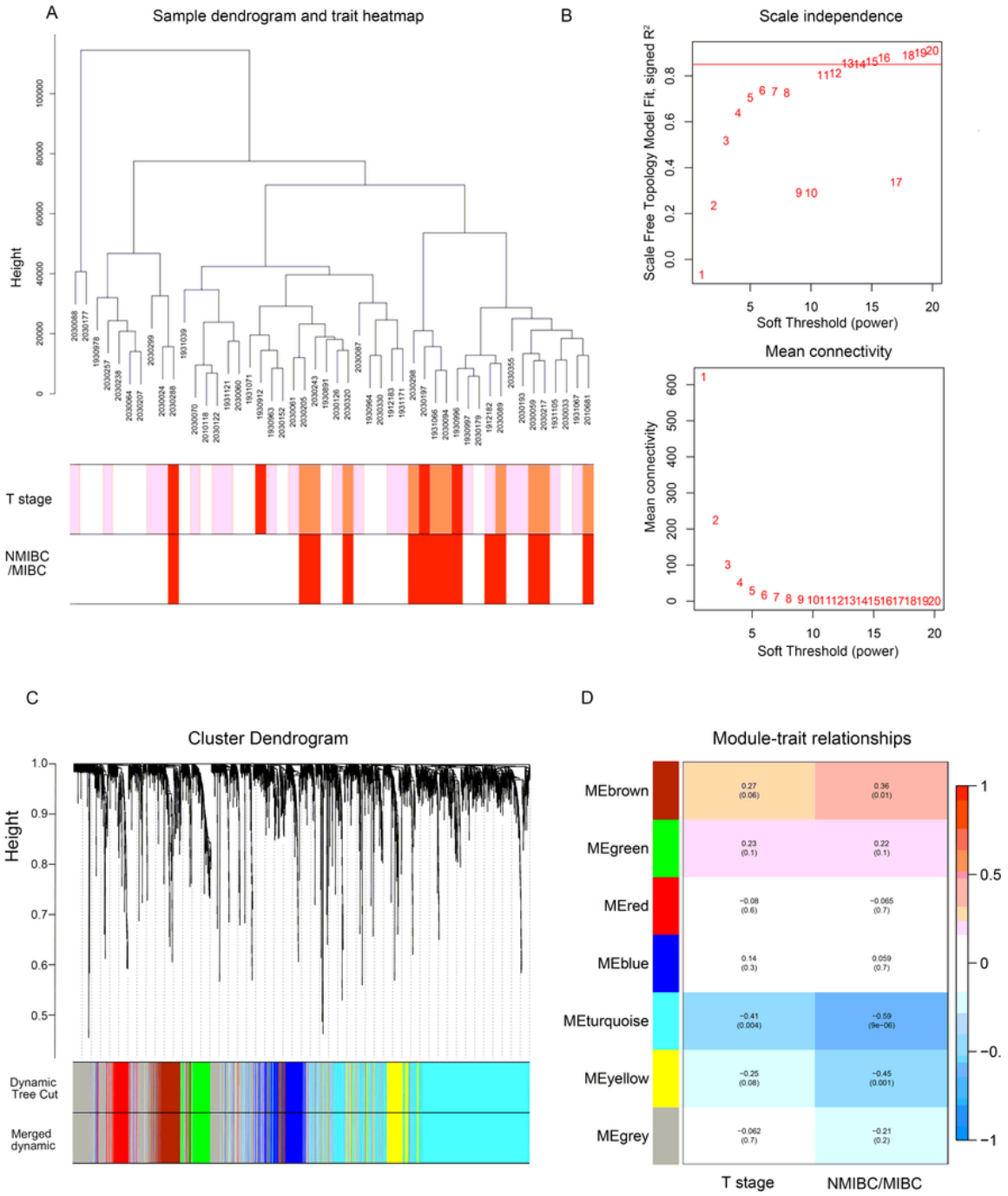
| Characteristic, n (%) | | GSE32894 | GSE13507 | TCGA | STPH |
|---|---|---|---|---|---|
| Subtype | NMIBC | 173 | 103 | 3 | 35 |
| | MIBC | 51 | 62 | 400 | 14 |
| Age (years) | ≥65 | 70 | 69 | 254 | 35 |
| | <65 | 154 | 96 | 149 | 14 |
| Gender | Male | 163 | 135 | 299 | 42 |
| | Female | 61 | 30 | 104 | 7 |
| Stage | I | - | 103 | 2 | |
| | II | - | 1 | 128 | |
| | III | - | 0 | 138 | |
| | IV | - | 1 | 132 | |
| | Unknown | - | 60 | 3 | |
| Grade | Low | - | 105 | 21 | 5 |
| | High | - | 60 | 378 | 44 |
| | Grade I | 45 | - | - | |
| | Grade II | 84 | - | - | |
| | Grade III | 93 | - | - | |
| | Unknown | 2 | - | 4 | |
| N stage | N0 | 26 | 1 | 234 | |
| | N1-3 | 20 | 104 | 128 | |
| | Unknown | 177 | 60 | 41 | |
| M stage | M0 | - | 1 | 196 | |
| | M1 | - | 104 | 11 | |
| | Unknown | - | 60 | 196 | |
| †NMIBC, non-muscle invasive bladder cancer<br>MIBC, muscle invasive bladder cancer | | | | | |

# Figures




## Figure 1

Flowchart showing the construction process of the risk score model




## Figure 2

The WGCNA network construction and key module identification. (a) Sample dendrogram and trait indicator. The clustering was a visual result of calculations based on Pearson correlation coefficients between samples. The color intensity was proportional to T stage and NMIBC/MIBC subgroup of BC. (b) The analysis of topology for soft threshold powers and β=13 (scale-free R2=0.85) was set as the soft thresholding for further adjacency calculation. (c) The original modules and merged modules are

displayed at top and bottom of the clustering dendrogram. (d) Module-trait relationships between identified modules and clinical features. The numbers represent Pearson's correlation between the clinical traits and modules. The numbers in the parentheses correspond to the p value. WGCNA: weighted gene correlation network analysis. NMIBC: non-muscle-invasive bladder cancer; MIBC: muscle-invasive bladder cancer. ME: module.

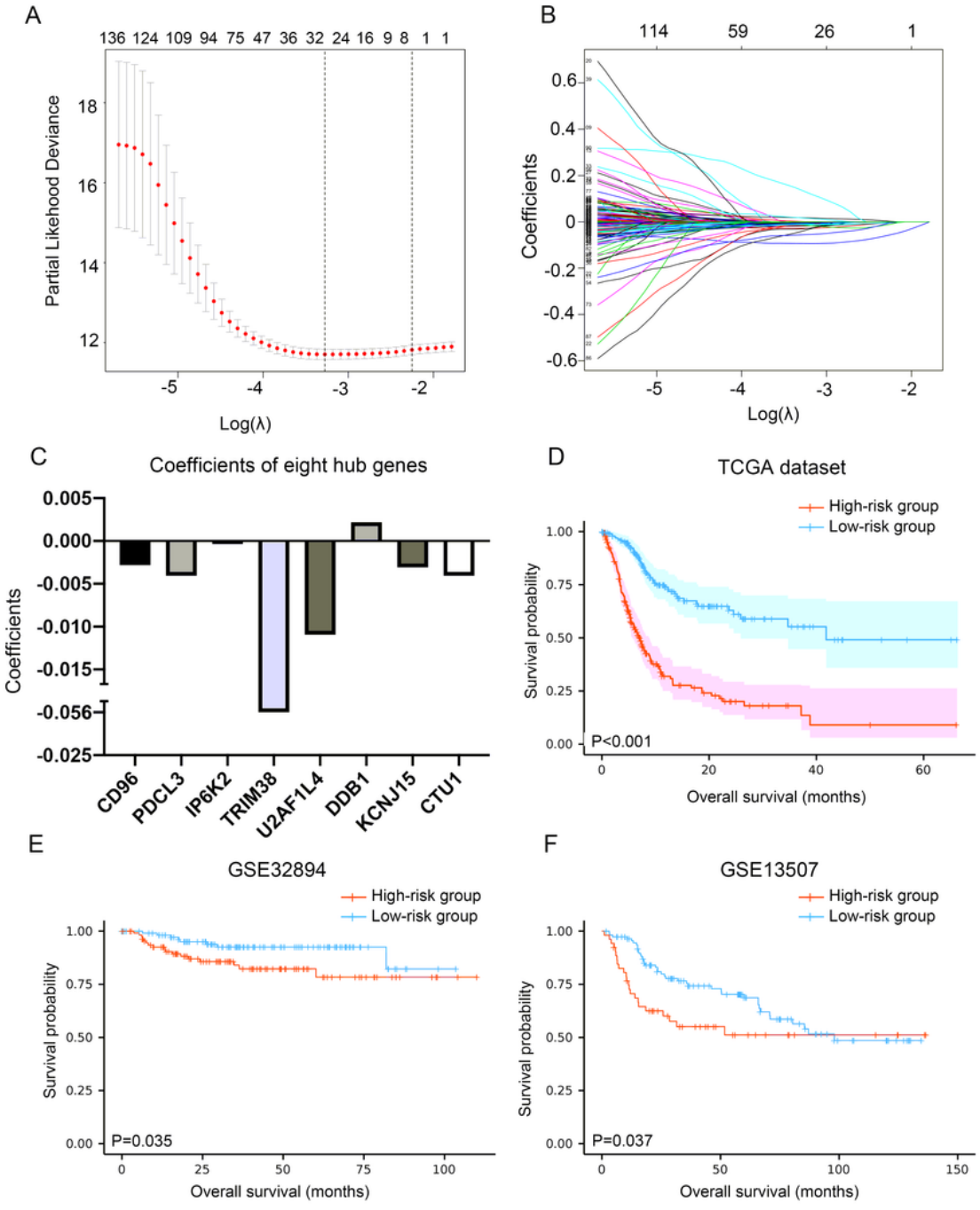
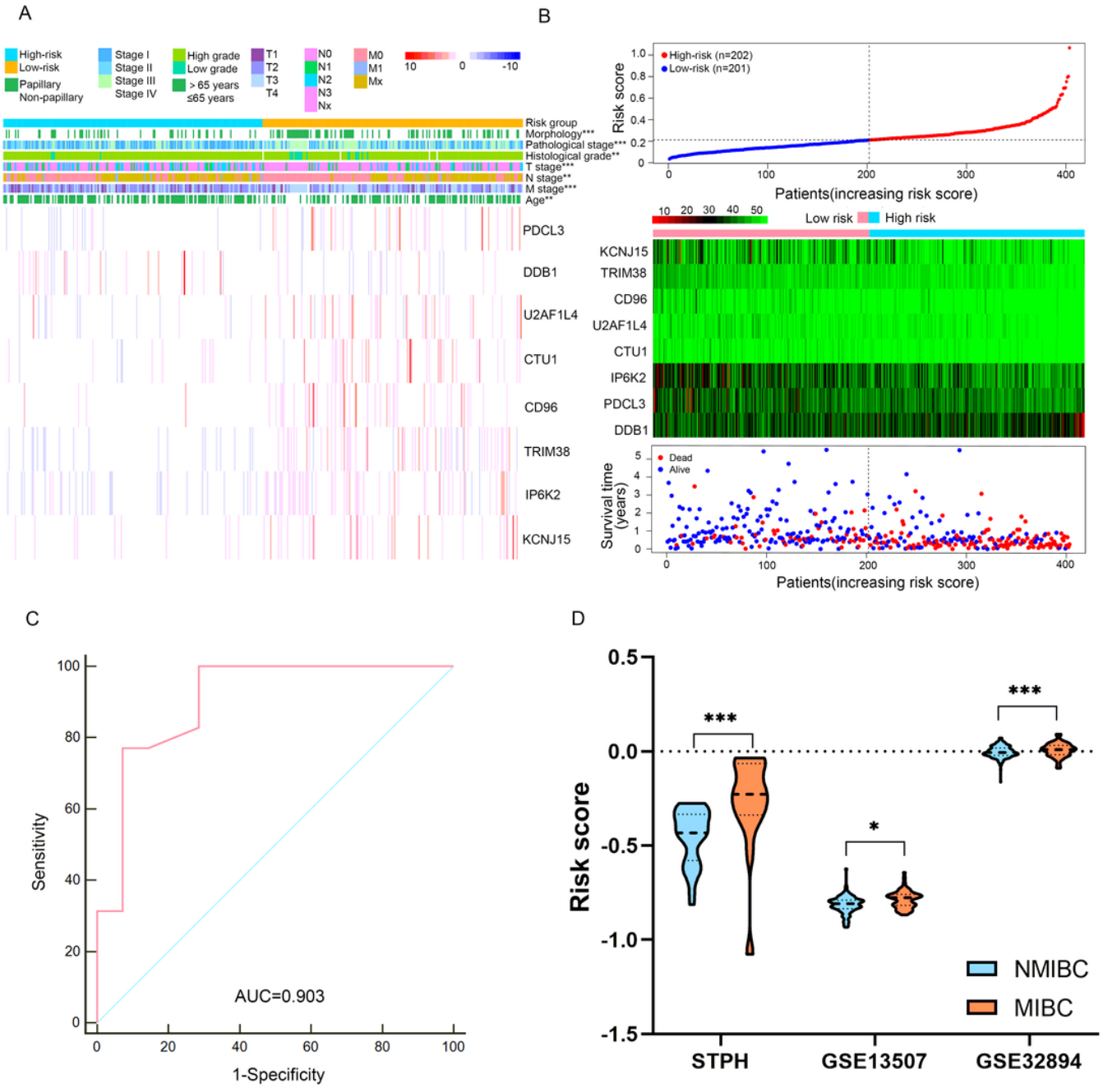



**Figure 3**

Construction and validation of the risk score. (a) Plots of the cross-validation error rates. (b) Distribution of LASSO coefficients of OS-associated genes. (c) Coefficient values of the eight genes. (d, e, f) Kaplan-Meier curves of OS for BC patients assigned to high- and low-risk groups in TCGA, GSE32894 and GSE13507. LASSO: least absolute shrinkage and selection operator; TCGA: The Cancer Genome Atlas; BC: bladder cancer; NMIBC: non-muscle-invasive bladder cancer; MIBC: muscle-invasive bladder cancer; OS: overall survival.




## Figure 4

Associations between risk score and clinical variables. (a) Clinical variables and expression levels of eight genes were compared between the low- and high-risk groups (b) Risk scores distribution, OS status

of each patient, and heatmaps of eight genes in TCGA dataset. (c) Receiver operating characteristic curve showed the discrimination power of the risk score for NMIBC and MIBC in STPH dataset. (d) The distribution of risk scores was compared between NMIBC and MIBC in STPH, GSE13507 and GSE32894 datasets. *P<0.05, **P<0.01, ***P<0.001.

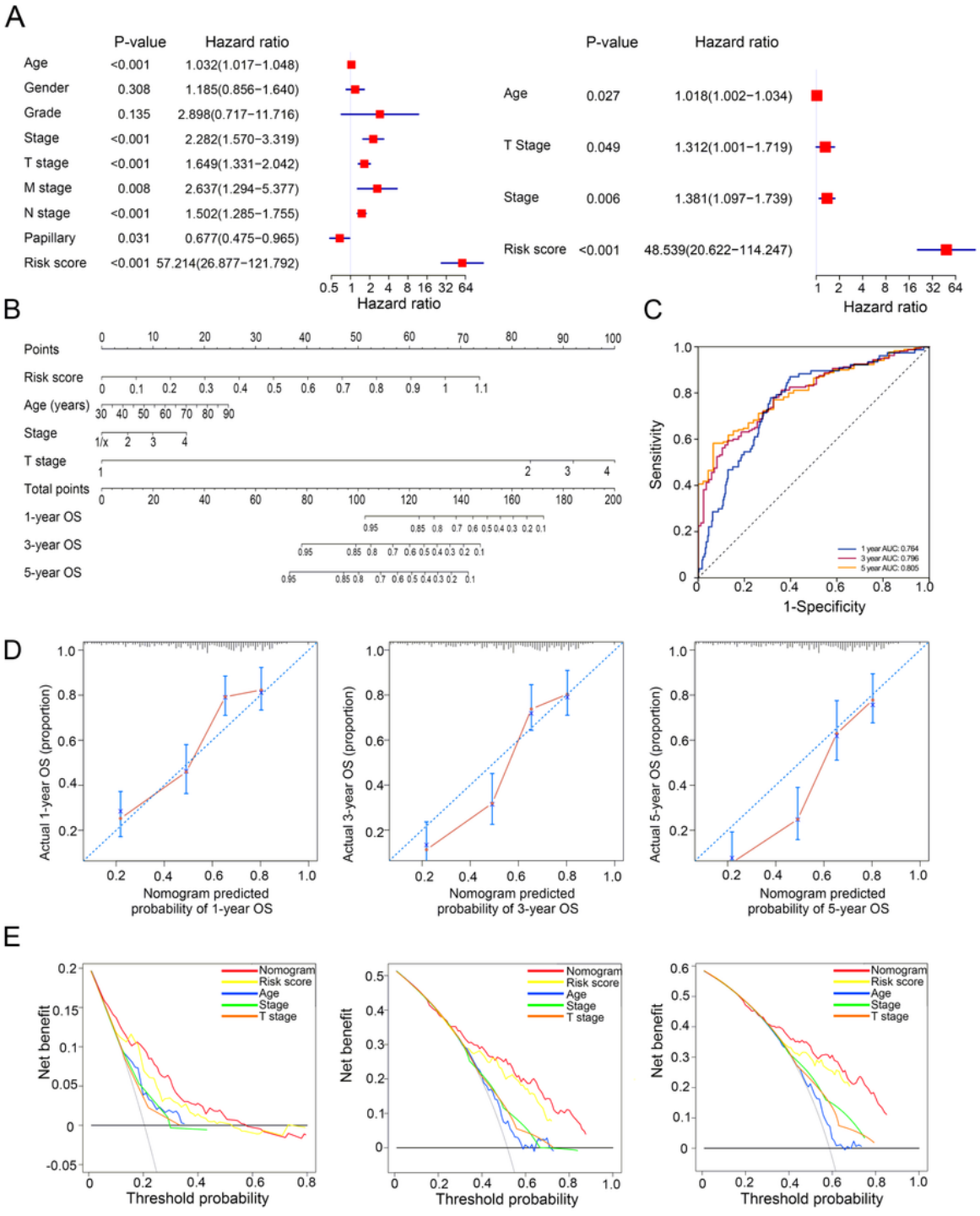



Figure 5

Nomogram construction based on the risk score in TCGA dataset. (a) Univariate and multivariate Cox analyses indicated that the risk score was significantly associated with OS. (b) Nomogram for predicting the probability of 1-, 3-, and 5-year OS. (c) ROC curve showed the predictive performance of 1-, 3-, and 5-year OS. (d) Calibration plots of the nomogram for predicting the probability of 1-, 3-, and 5-year OS. (e) DCA analysis showing the performance of the nomogram to predict the 1- 3-, and 5-year OS. OS: overall survival; ROC: Receiver operating characteristic; DCA: decision curves analyses; HR: hazard ratio; CI: confidence interval.

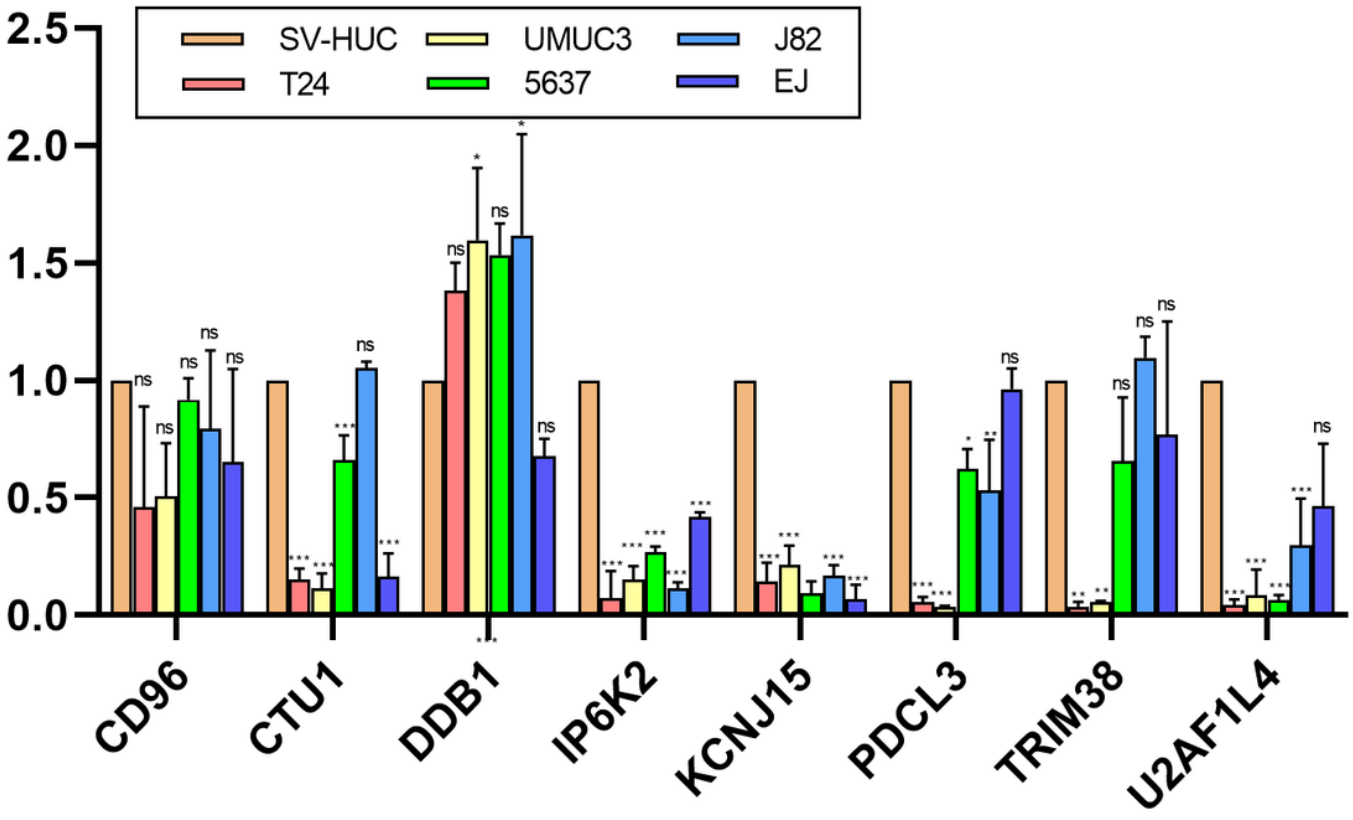



**Figure 6**

mRNA and protein level validation of eight genes. (a) RT-qPCR validates expression of eight genes in bladder cancer cell lines T24, UMUC3, 5637, EJ and J82 and normal bladder epithelial cell line SV-HUC-1. (b) IHC intensity of eight genes in normal bladder tissue and bladder cancer tissue from HPA. IHC: Immunohistochemistry; BC: bladder cancer.

# Supplementary Files

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

- SuppTable.docx
- newS1.tif

- FS2.tif