

The integrated analysis of single-cell RNA sequencing and bulk transcriptome sequencing revealed prognostic genes and pathological/clinical features associated with RNA processing factors in muscle-invasive bladder cancer

Lina Ji

Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Third Hospital of Shanxi Medical University

Yishuang Guo

First Hospital of Shanxi Medical University

Yuedong Hao

First Hospital of Shanxi Medical University

Liang Wei

`weiliang0340@126.com`

First Hospital of Shanxi Medical University

Research Article

Keywords: Muscle-invasive bladder cancer, RNA processing factors, Prognostic genes, Pathological features, Fibroblasts, Multiple omics analysis

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10866581/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

The integrated analysis of single-cell RNA sequencing and bulk transcriptome sequencing revealed prognostic genes and pathological/clinical features associated with RNA processing factors in muscle-invasive bladder cancer

Lina Ji¹, Yishuang Guo², Yuedong Hao², Liang Wei^{2*}

¹(1) Cancer Center, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Tongji Shanxi Hospital, Third Hospital of Shanxi Medical University, Taiyuan, 030032, China;

(2) Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430030, China

²Department of Urology, First Hospital of Shanxi Medical University, Taiyuan, 030001, China

*** Correspondence:**

Liang Wei

Department of Urology, First Hospital of Shanxi Medical University, Taiyuan, 030001, China

weiliang0340@126.com

ORCID: 0009-0000-3673-5998

Abstract

Background: RNA processing factors are associated with pathological features and prognosis in muscle-invasive bladder cancer (MIBC). We integrated bulk transcriptomic, single-cell, pathological, and clinical data to identify RNA-processing-related prognostic factors in MIBC.

Methods: TCGA-BLCA data and literature-derived RNA processing factor-associated genes were analyzed. Differential expression and regression analyses were used to identify prognostic genes and construct a gene-based prognostic model. Pathological and clinical prognostic features were identified through correlation and regression analyses and used to build corresponding models. SHAP analysis integrated the prognostic factors and ranked their importance. Single-cell RNA sequencing data were used to identify key cell populations and assess fibroblast differentiation by pseudotime analysis.

Results: AHNAK, KHDC4, TMTC1, and RPP21 were identified as prognostic genes, and the gene-based model showed poorer survival in the high-risk group. Resnet1306, resnet1323, and resnet1743 were identified as prognostic pathological features, and their model also stratified patients by survival. Age and M stage were identified as prognostic clinical factors. In the integrated SHAP model, AHNAK had the highest importance, followed by age and resnet1306. Fibroblasts were identified as key cells, and AHNAK showed relatively high activity during fibroblast differentiation.

Conclusions: Integrated multi-omics analysis identified candidate molecular, pathological, and clinical prognostic factors for MIBC and highlighted AHNAK and fibroblasts as potentially important components of disease progression.

Keywords: Muscle-invasive bladder cancer, RNA processing factors, Prognostic genes, Pathological features, Fibroblasts, Multiple omics analysis

1. Background

Bladder cancer (BC) is a malignant tumor originating from the urothelium of the bladder and ranks among the most aggressive malignancies of the urinary system [1]. According to GLOBOCAN 2022 data, BC is the tenth most commonly diagnosed cancer worldwide [2]. Clinically, the most common histopathological type of BC is urothelial carcinoma, accounting for approximately 95% of all cases. Based on the depth of invasion into the bladder wall, urothelial carcinoma is classified into non-muscle-invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC), with NMIBC comprising about 75% of cases and MIBC accounting for roughly 25%, including cases that have already developed metastasis. Overall, MIBC has a poor prognosis and can even be life-threatening [3]. Currently, the standard treatment for MIBC typically involves neoadjuvant chemotherapy combined with radical cystectomy; however, up to 50% of patients still experience disease recurrence or metastasis. For patients with unresectable MIBC or metastatic MIBC, the primary treatment option is platinum-based combination chemotherapy, yet the survival outcomes remain dismal, with a median overall

1 survival (OS) of approximately 14 months [4]. Although immune checkpoint
2 inhibitors and antibody-drug conjugates have emerged as promising new
3 therapeutic options that can extend OS in some patients, their overall efficacy
4 remains unsatisfactory [5]. Therefore, identifying novel therapeutic targets and
5 developing innovative treatments for MIBC remain urgent priorities.

6 RNA-processing factors are a class of evolutionarily conserved
7 ribonucleoprotein complex components and associated proteins that
8 participate in post-transcriptional regulatory processes such as precursor
9 mRNA splicing in eukaryotes. By regulating splicing, RNA homeostasis, and
10 other processes, they can influence chromosome segregation and genome
11 stability [6]. Currently, studies have shown that certain RNA-processing factors
12 are closely associated with the prognosis and immune infiltration in bladder
13 cancer and could serve as potential molecular targets [7]. RNA processing
14 factors can also contribute to tumor initiation and progression by regulating
15 aberrant alternative splicing events in MIBC. Furthermore, a splicing
16 factor-alternative splicing regulatory network associated with prognosis has
17 been constructed, providing crucial evidence for the prognostic evaluation and
18 mechanistic studies of MIBC [8]. Although RNA-processing factors have been
19 implicated in MIBC, their underlying mechanisms remain insufficiently
20 characterized.

21 scRNA-seq is a technical approach that enables the analysis of transcriptomic
22 data at the single-cell level. It can be used to characterize cellular
23 heterogeneity, identify cell types, and explore intercellular communication
24 patterns and interaction mechanisms [9]. Through scRNA-seq, it is possible to
25 gain insights into the prognosis and response to immunotherapy in patients
26 with MIBC [10]. For example, one study using scRNA-seq identified a
27 fibroblast gene signature that promotes T-cell infiltration into MIBC [11].
28 However, to date, no studies have explored the role of RNA-processing factors
29 in the context of MIBC pathogenesis using scRNA-seq.

This study was based on the TCGA-BLCA dataset, which provided gene expression, pathological, and clinical data on MIBC. Additionally, genes related to RNA processing factors were retrieved from relevant literature. Subsequently, these data were processed using multiple approaches to identify prognostic genes, prognostic pathological features, and prognostic clinical features associated with MIBC prognosis. Prognostic models were then constructed separately for each of these factors, and the relationship between each prognostic factor and MIBC risk was explored. Furthermore, at the single-cell level, key cell types influencing MIBC were analyzed, and pseudotime analysis was employed to investigate the expression levels of prognostic genes during the differentiation of these key cell types. Overall, this study provides new insights into the molecular mechanisms by which RNA processing factors influence MIBC and identifies potential therapeutic targets for MIBC treatment.

2. Methods

2.1.Data collection

The transcriptome dataset TCGA-BLCA [12] related to muscle-invasive bladder cancer (MIBC) was downloaded from The Cancer Genome Atlas (TCGA, <https://portal.gdc.cancer.gov/>) database, including gene expression data from 256 tumor samples (229 with survival information) and 19 control samples, as well as pathological section data and clinical data of MIBC. Subsequently, the single-cell RNA sequencing (scRNA-seq) dataset GSE135337 [13] (platform GPL24676) for MIBC was obtained from the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) database, containing 1 MIBC sample and 1 control tissue sample. Finally, 929 RNA processing factor-associated genes were retrieved from the literature [14] (**ESM 1**).

2.2.Differential expression analysis

In TCGA-BLCA, to acquire differentially expressed genes (DEGs) between MIBC and control samples, “DESeq2” package (v 1.40.2) [15] was applied (MIBC vs control, $|\log_2\text{FoldChange (FC)}| > 1$, adjusted P (adj.P) < 0.05). Based

on the $\log_2FC$ values, all DEGs and the top 10 DEGs with up-regulation and down-regulation were presented by a volcano plot generated via “ggplot2” package (v 3.5.2) [15]. The expression levels of the top 10 up-regulated and down-regulated DEGs in the MIBC samples were displayed by a heatmap created via “pheatmap” package (v 1.0.12) [16].

2.3.Identification and functions of candidate genes

DEGs and RNA processing factor-associated genes were intersected to obtain candidate genes using the “ggvenn” package (v 0.1.10) [17]. Subsequently, using “clusterProfiler” package (v 4.15.0.3) [18], Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed on candidate genes to investigate the related functions ($P < 0.05$). According to count values, the top 10 significantly enriched GO functions and the top 15 significantly enriched KEGG pathways were displayed via “ggplot2” package (v 3.5.2). Furthermore, to explore the interactions of candidate genes at the protein level, a protein-protein interaction (PPI) network for candidate genes was constructed by Search Tool for the Retrieval of Interaction Gene/Proteins (STRING, <https://string-db.org/>) database (interaction score > 0.4). The network was visualized via “Cytoscape” software (v 3.10.4) [19].

2.4.Identification of prognostic genes and construction of prognostic models

In the MIBC samples with survival information of TCGA-BLCA, univariate Cox regression analysis ($P < 0.05$, hazard ratio (HR) $\neq 1$) with proportional hazards (PH) assumption test ($P > 0.05$) was implemented on candidate genes to select genes related to MIBC prognosis using “survival” package (v 3.7.0) [20]. After that, in the same samples, MIBC-related prognostic genes were further acquired by the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm conducted via “glmnet” package (v 4.1.8) [21] (10-fold cross-validation). Genes with the minimum lambda value and non-zero coefficients were identified as prognostic genes.

Additionally, in the MIBC samples with survival information of TCGA-BLCA, the MIBC samples were divided into a training set and a validation set in a ratio of 7:3, with 161 and 68 samples, respectively. Then, a prognostic model was

constructed in the training set. Based on the coefficients (coef) of prognostic genes obtained from LASSO algorithm and their expression levels (exp), the risk scores of MIBC patients were calculated with the following formula: [Risk score = $\sum(\text{coef}_i \times \text{exp}_i)$]. Based on the optimal cutoff value of the risk score, patients were divided into high-risk group (HRG) and low-risk group (LRG). Subsequently, the Kaplan-Meier (K-M) curves between the two risk groups were created via “survival” package (v 3.7.0) and compared via Log-rank test ($P < 0.05$). The accuracy of prognostic model was determined by receiver operating characteristic (ROC) curves for 1-, 3-, and 5-year survival using “timeROC” package (v 0.4) [22], and area under the curve (AUC) values were calculated with $\text{AUC} > 0.6$. After that, a prognostic model was constructed in the validation set using the same method and threshold to validate the accuracy of the above-mentioned model.

2.5. Clinical feature analysis

To investigate the correlation between clinical features and MIBC prognosis, the differences in age, gender, M stage, N stage, stage, and T stage between the HRG and LRG were analyzed ($P < 0.05$). Next, univariate Cox regression analysis ($P < 0.05$, $\text{HR} \neq 1$) with PH assumption test ($P > 0.05$) and multivariate Cox regression analysis ($P < 0.05$, $\text{HR} \neq 1$) with PH assumption test ($P > 0.05$) were sequentially performed via “survival” package (v 3.7.0) to acquire independent prognostic factors related to MIBC from the risk scores and clinical features. Thereafter, a nomogram model was constructed to predict the survival of MIBC patients for 1, 3, and 5 years utilizing “rms” package (v 6.8.1) [23]. The calibration curves generated via “rms” package (v 6.8.1) were utilized to verify the accuracy of the nomogram model.

2.6. Drug sensitivity analysis

To explore the differences in response to chemotherapy and targeted drugs among MIBC patients, the sensitivity to multiple common chemotherapeutic and targeted agents was predicted in TCGA-BLCA using the “pRRophetic” package (v 0.5) [24], combined with drug response data provided by Genomics of Drug Sensitivity in Cancer (GDSC, <https://www.cancerrxgene.org>) database. The predictive indicator used was the half-maximal inhibitory concentration (IC_{50}). Wilcoxon test was then performed to compare the IC_{50} values of each

1 drug between the HRG and LRG ($P < 0.05$). The IC_{50} differences for the top 10
2 drugs ranked by P-value between the 2 groups were visualized using “ggplot2”
3 package (v 3.5.2). Moreover, “psych” package (v 2.4.6.26) [25] was used to
4 analyze the correlations between the top 10 drugs and risk scores, as well as
5 between prognostic genes and the first-ranked drug ($|correlation (cor)| > 0.3$, P
6 < 0.05).

7 **2.7. Quality control analysis and identification of key cells**

8 In GSE135337, “Seurat” package (v 5.1.0) [26] was employed to process the
9 data and conduct all analyses. Initially, the genes expressed in more than 3
10 cells were retained. Then, quality control analysis was performed on the data
11 with $200 < nFeature_RNA$ (the number of expressed genes in each cell) $<$
12 $3,000$, $nCount_RNA$ (total gene expression counts in each cell) $< 10,000$, and
13 $percent.mt$ (the proportion of mitochondrial genes) $< 10\%$. Then, the filtered
14 data were standardized via LogNormalize function. Following vst method of
15 FindVariableFeatures function, the top 2,000 highly variable genes were
16 extracted. Next, the data were further scaled via ScaleData function, and
17 Principal Components Analysis (PCA) was performed. The P values of
18 principal components (PCs) were calculated via JackStraw function. A scree
19 plot was created to display the variance inflection values of different PCs. PCs
20 with $P < 0.05$ and at the inflection point of the scree plot were used for
21 clustering analysis via FindNeighbors and FindClusters functions with
22 $resolution = 0.6$. By identifying specifically highly expressed genes in cell
23 clusters via FindAllMarkers function, cell clusters were annotated based on
24 marker genes reported in the published literature [27] in combination with
25 “SingleR” package (v 2.2.0) [28]. Among them, the expression levels of
26 prognostic genes in cells were visualized by Uniform Manifold Approximation
27 and Projection (UMAP) plots, and cells expressing multiple prognostic genes
28 were identified as key cells.

29 **2.8. Pseudotime analysis, cell communication, and cell cycle analysis**

30 In GSE135337, using the aforementioned method, the key cells were
31 subjected to secondary dimensionality reduction and clustering. Subsequently,
32 the differentiation trajectory of the key cells and the expression levels of
33 prognostic genes were analyzed using the “Monocle” package (v 2.28.0) [29].

Moreover, “CellChat” package (v 1.6.1) [30] was applied to analyze the interactions between cells and calculate the potential ligand-receptor interactions ($P \leq 0.05$, $\log_2 \text{mean}(\text{molecule 1}, \text{molecule 2}) \geq 0.1$).

Furthermore, to investigate the cycle changes of key cells, the CellCycleScoring function from the “Seurat” package (v 5.1.0) was used to analyze the cycle stage of each cell.

2.9.Acquisition of pathological features for MIBC

To identify the prognostic pathological features of MIBC patients, whole-slide images (WSIs) of MIBC patients were acquired from TCGA-BLCA, and the samples were divided into a training set and a validation set at a ratio of 7:3.

Then, WSIs of all patients were standardized, and each image was segmented into non-overlapping patches with a pixel size of 512×512 . Otsu’s thresholding method was then applied for tissue detection and filtering to eliminate background and low-information regions, and only patches with a tissue content exceeding 80% were retained. Finally, the Macenko method was used to perform color normalization on all retained patches to reduce staining variations across different samples.

Subsequently, the pathological images were processed using the OpenCV software, and the `cv2.cvtColor()` method was adopted to convert the images from the blue-green-red (BGR) mode to the red-green-blue (RGB) mode, followed by normalization using the ImageNet standard normalization method (mean = [0.485, 0.456, 0.406], standard deviation = [0.229, 0.224, 0.225]). The images were resized to 224×224 pixels and input into a pre-trained ResNet-50 model based on the PyTorch framework for feature extraction. The output of the layer4 module (before the final classification layer) was preserved, generating a deep feature map with a size of $2048 \times 7 \times 7$. Global max pooling was applied to this feature map, resulting in a 2048-dimensional feature vector generated for each image.

In the subsequent cellular-level pathological analysis, the images were imported into the CellProfiler software to extract cellular-level features. The RGB images were converted into grayscale staining separation channels via the UnmixColors module, and hematoxylin and eosin components were separated based on optical density deconvolution. The IdentifyPrimaryObjects

module was utilized, and adaptive thresholding and size filtering (minimum diameter: 10 pixels, maximum diameter: 50 pixels) were implemented to ensure accurate and robust nuclear identification. The complete cellular regions were then defined by outward propagation from the nuclear boundaries using the IdentifySecondaryObjects module. After segmentation, intensity features, texture features, morphological features, and pixel distribution features were obtained through the MeasureObjectIntensity module, MeasureTexture module, MeasureObjectSizeShape module, and MeasureObjectIntensityDistribution module, respectively. After the removal of non-essential features, the quantitative image features were retained for subsequent analysis.

2.10. Acquisition of prognostic pathological features and prognostic pathological models

In the training set, the scores of RNA processing factor-related genes were calculated via the single-sample Gene Set Enrichment Analysis (ssGSEA) method, and correlation analysis was performed between these scores and pathological features. Pathological features meeting the criteria of $|\text{cor}| > 0.1$ and $P < 0.05$ were defined as candidate pathological features. Subsequently, combined with the methods described in Section 2.4, univariate Cox regression analysis and the LASSO algorithm were conducted on the candidate pathological features to identify prognostic pathological features associated with MIBC. Likewise, in accordance with the prognostic model construction methods presented in Section 2.4, prognostic pathological models were separately established in the training set and validation set based on the prognostic pathological features, and the value of these prognostic pathological features for the prognosis of MIBC was investigated.

2.11. Acquisition of prognostic clinical features and prognostic clinical models

To investigate the predictive performance of clinical data for risk stratification in MIBC patients, clinical data were divided into training and validation sets at a ratio of 7: 3. First, in the training set, differential clinical characteristics between MIBC patients and control samples were identified using t-tests ($P < 0.05$). Subsequently, prognostic clinical features associated with MIBC prognosis

1 were extracted from these differential clinical characteristics using the method
2 for identifying independent prognostic factors described in Section 2.5.

3 Next, in the validation set, MIBC patients were stratified into a high age group
4 and a low age group at the age cutoff of 60 years, and into M0 and M1 groups
5 according to M stage. Based on the regression coefficients derived from
6 multivariate Cox regression analysis, risk scores for each group were
7 calculated using the formula (risk score = coefficient $\times$ age + coefficient $\times$ M).
8 Differences in risk scores between groups were examined ($P < 0.05$) to
9 validate the association between prognostic clinical features and MIBC risk.

10 **2.12. SHAP model construction**

11 To comprehensively analyze the predictive performance of different risk
12 models for MIBC, an XGBoost model was constructed using a fusion strategy
13 across all models with 10-fold cross-validation. Optimal parameters, including
14 learning rate, maximum depth, gamma, and n-estimators, were obtained using
15 the “XGBoost” package (v 2.1.1.1) [31]. Subsequently, Platt scaling was
16 applied for calibration to achieve the highest classification accuracy, and the
17 average of the calibrated accuracies of each model was used as the prediction
18 rate of the multimodal model. Finally, the top 10 features of the fusion model
19 were selected to construct a SHAP model, and the contribution of each feature
20 to model prediction was visualized using the “shapviz” package (v 0.9.7) [32].

21 **2.13. Statistical analysis**

22 All analyses were performed utilizing R software (v 4.2.2). The Wilcoxon test
23 and t-test were utilized to compare the differences between two groups. * $P <$
24 0.05, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, ns, not significant.

25 **3. Results**

26 **3.1.Acquisition and functions of 69 candidate genes**

27 Compared with the control group, 4,563 DEGs (2,518 upregulated and 2,045
28 downregulated) were obtained in the MIBC group (**Figure 1a, b**). DEGs were
29 then intersected with RNA processing factor-associated genes, and 69
30 candidate genes were obtained (**Figure 1c**). Notably, candidate genes were
31 significantly enriched in 410 GO functions which contained 288 biological
32 processes (BPs) including mRNA processing, 63 cellular components (CCs)

1 Intersection of DEGs and RNA processing factor-associated genes. (d) GO
2 enrichment analysis. (e) KEGG enrichment analysis. (f) PPI network of
3 candidate genes

4 **3.2. The 4 prognostic genes and prognostic models related to MIBC**

5 Univariate Cox regression analysis indicated that 9 genes were significantly
6 associated with MIBC prognosis, namely LIN28A, AHNAK, TMTC1, SAGE1,
7 AHNAK2, RBPMS, PABPN1, KHDC4, and RPP21 ($P < 0.05$, $HR \neq 1$, **Figure**
8 **2a**). The subsequent PH assumption test showed that all nine genes met the
9 PH assumption and were subjected to LASSO analysis ($P > 0.05$, **ESM 3**).
10 When the error of the LASSO model was at its minimum, $\lambda = 0.10583$,
11 and 4 genes were further selected at this point (**Figure 2b**). Ultimately, these 4
12 genes, namely AHNAK, KHDC4, TMTC1, and RPP21, were identified as
13 prognostic genes, with corresponding coefficients of 0.07891, -0.15574,
14 0.03471, and -0.00573, respectively.

15 After that, a prognostic model was constructed in the training set. Based on the
16 optimal intercept value of -0.0386 ($\text{minprop} = 0.1$), MIBC patients were divided
17 into the HRG ($n = 57$) and the LRG ($n = 104$) (**Figure 2c**). The K-M curves
18 showed that the HRG had a significantly lower survival rate ($P < 0.0001$,
19 **Figure 2d**). The ROC curves verified the accuracy of the prognostic model
20 with $AUC > 0.6$ (**Figure 2e**). These results demonstrated that the prognostic
21 genes had the ability to assess the prognosis of patients with MIBC. Similarly,
22 the same results were obtained in the validation set (**Figure 2f-h**). The patients
23 in the validation set were divided into the HRG ($n = 13$) and LRG ($n = 55$).

24 In summary, the prognostic model constructed by prognostic genes could
25 accurately predict the survival of patients with MIBC.

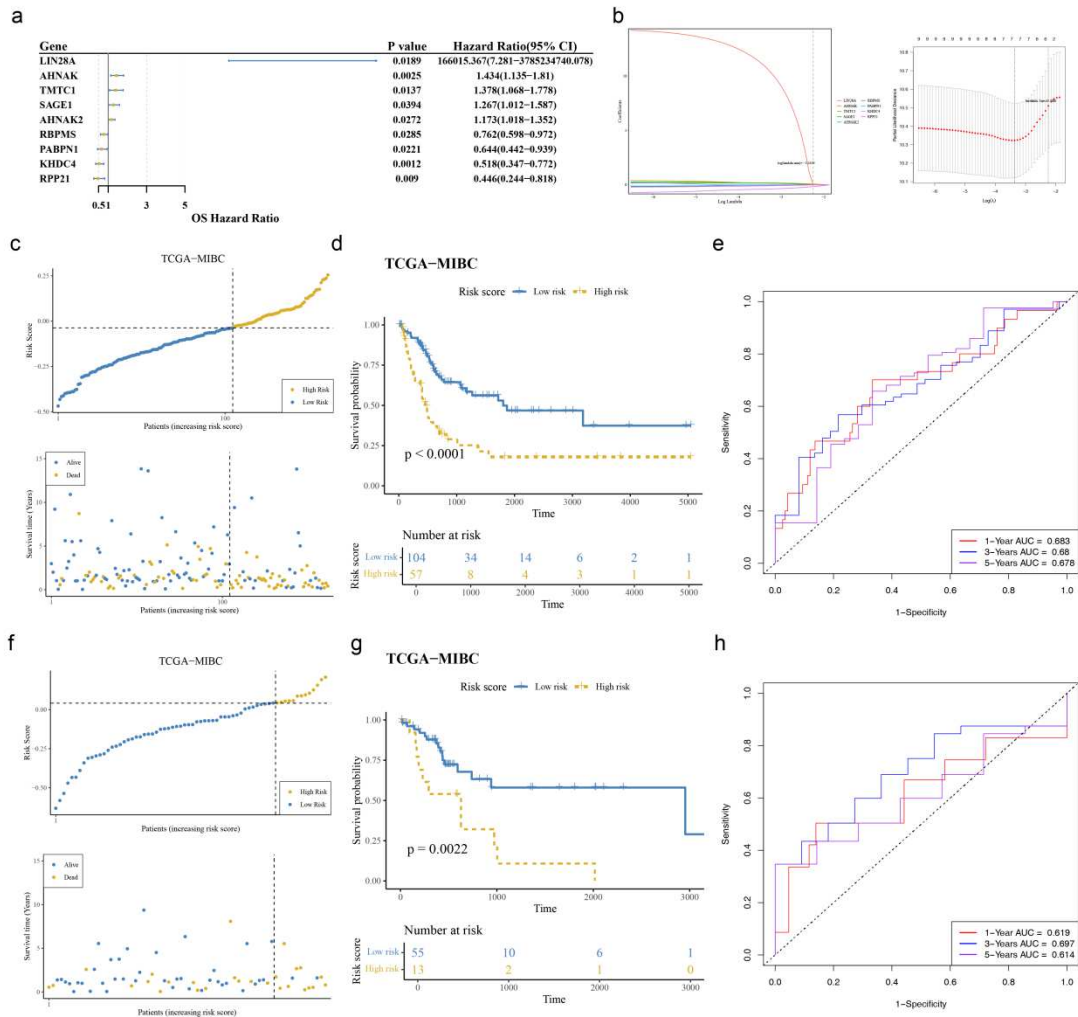


Fig. 2 Identification of prognostic genes and construction of the prognostic gene model. (a) Univariate Cox regression analysis of candidate genes. (b) LASSO regression analysis. (c) Risk score distribution in the training set. (d) Kaplan-Meier survival curves in the training set. (e) ROC curves in the training set. (f-h) Validation of the prognostic model in the validation set

3.3. The nomogram model constructed based on risk score and stage

With the exception of age and gender, the remaining four clinical features exhibited significant differences between the HRG and the LRG, including T stage, M stage, N stage, and stage (**Figure 3a**). Subsequently, univariate Cox regression analysis ($P < 0.05$, $HR \neq 1$) with PH assumption test ($P > 0.05$, **Figure 3b**) exhibited that stage, M stage, N stage, and risk score were identified as candidate independent prognostic factors. Stage and risk score were further identified as independent prognostic factors by multivariate Cox

1 regression analysis ($P < 0.05$, $HR \neq 1$, **Figure 3c**) with PH assumption test ($P >$
2 0.05 , **ESM 4**). Afterwards, the nomogram model showed that as the total score
3 increased, the probability of death in MIBC patients also gradually increased,
4 and both the risk score and stage played important roles (**Figure 3d**). The
5 slope of each calibration curve was close to 1, indicating favorable predictive
6 performance of the model (**Figure 3e**).
7 Overall, stage and risk score related to MIBC prognosis could accurately
8 predict the survival rate of patients.

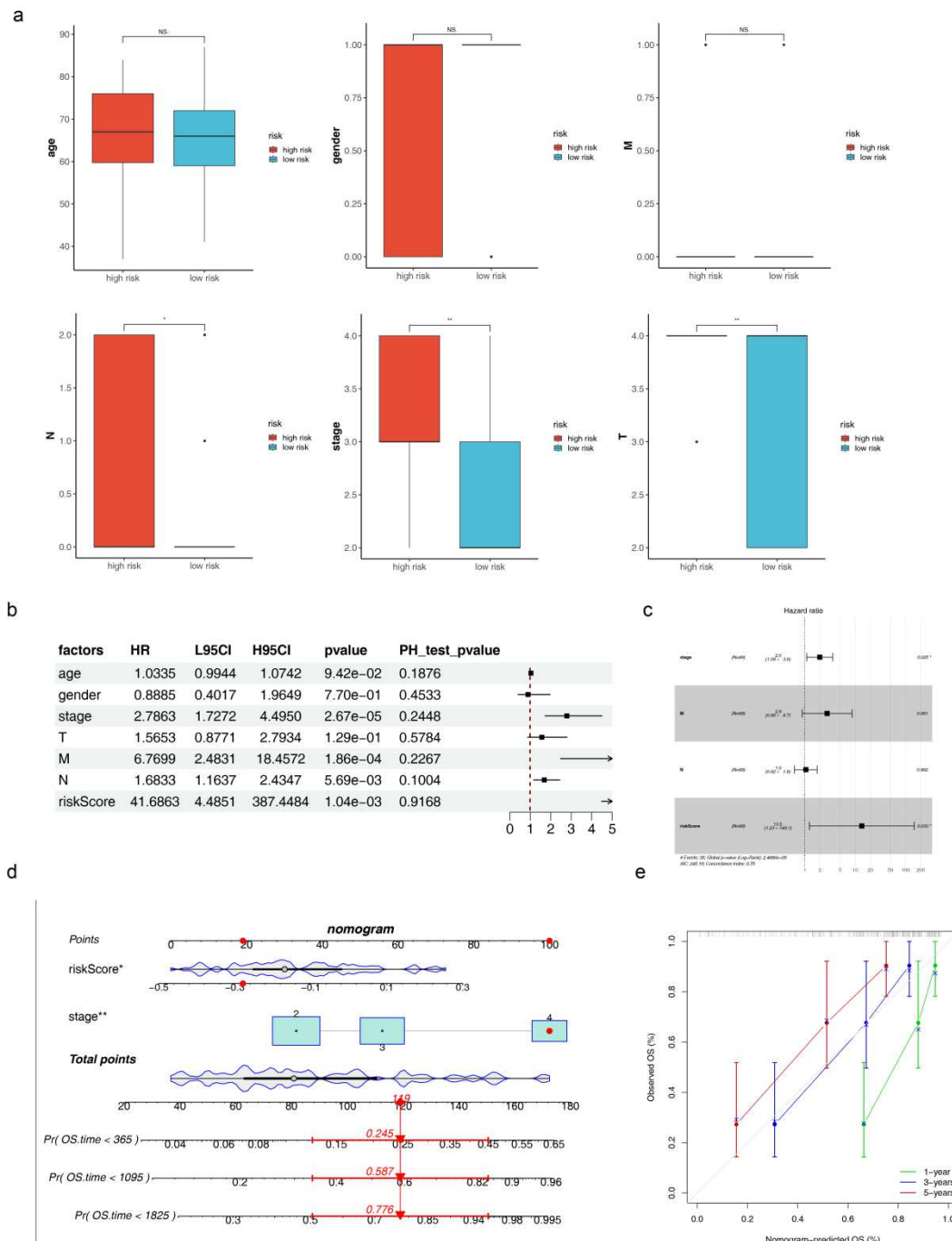
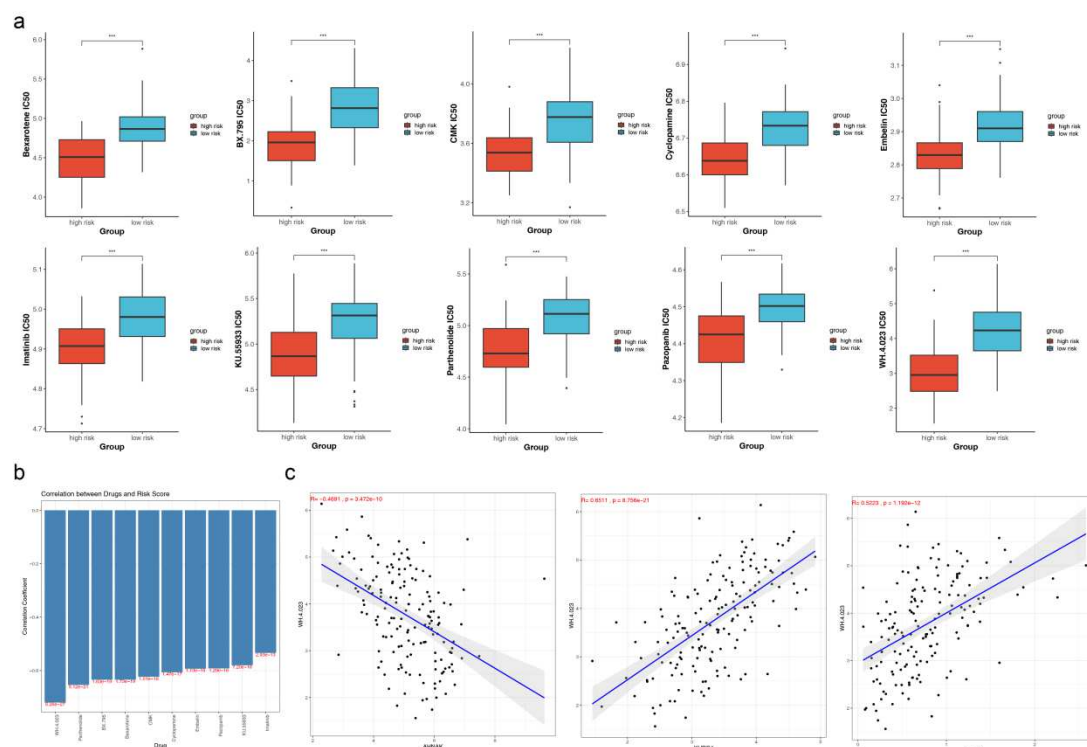


Fig. 3 Construction and evaluation of the nomogram model based on risk score and stage. (a) Differences in clinical features between HRG and LRG. (b) Univariate Cox regression analysis. (c) Multivariate Cox regression analysis. (d) Nomogram model for predicting survival. (e) Calibration curves

3.4. The differences in drug sensitivity between HRG and LRG

Drug sensitivity results showed that HRG and LRG exhibited significant

1 differences in 83 drugs ($P < 0.05$, **ESM 5**). The top 10 drugs with the most
 2 significant differences were WH.4.023, bexarotene, BX.795, embelin,
 3 cyclopamine, parthenolide, pazopanib, KU.55933, CMK, and imatinib, all of
 4 which showed significantly lower IC_{50} values in the HRG (**Figure 4a**). Of them,
 5 the 10 drugs were all significantly negatively correlated with the risk score (cor
 6 < -0.4 , $P < 0.0001$, **Figure 4b**). The drug with the strongest correlation to risk
 7 scores was WH.4.023, followed by parthenolide and BX.795. For WH.4.023,
 8 which showed the most significant difference, KHDC4 ($cor = 0.6511$, $P <$
 9 0.0001) and RPP21 ($cor = 0.5223$, $P < 0.0001$) were significantly positively
 10 correlated with WH.4.023, whereas AHNAK was significantly negatively
 11 correlated with this drug ($cor = -0.4691$, $P < 0.0001$, **Figure 4c**). These results
 12 demonstrated the associations among drugs, prognostic genes, and risk in
 13 MIBC patients.



14

15 **Fig. 4** Drug sensitivity analysis between HRG and LRG. (a) IC_{50} differences of
 16 the top 10 drugs. (b) Correlations between the top 10 drugs and risk score. (c)
 17 Correlations between prognostic genes and WH.4.023

1 **3.5.Acquisition of 6 types of cells and key cells**

2 Following quality control analysis of GSE135337, 11,166 cells and 18,240
3 genes were filtered down to 10,094 cells and 14,542 genes (**Figure 5a**). Next,
4 the top 2,000 highly variable genes were extracted, marking the top 10 highly
5 variable genes (**Figure 5b**). After performing PCA, the top 10 PCs were all
6 found to be statistically significant ($P < 0.05$, **Figure 5c**). After the 10th PC, the
7 changes tended to level off in the scree plot (**Figure 5d**). Thus, the top 10 PCs
8 were utilized for clustering analysis, and cells were clustered into 14 cell
9 clusters (**Figure 5e**). These cell clusters were annotated as six cell types,
10 comprising B cells, endothelial cells, fibroblasts, macrophages, neutrophils,
11 and T cells (**Figure 5f, g**). Of them, except for KHDC4, all the prognostic genes
12 were highly expressed in fibroblasts (**Figure 5h**). Therefore, fibroblasts were
13 defined as the key cells.

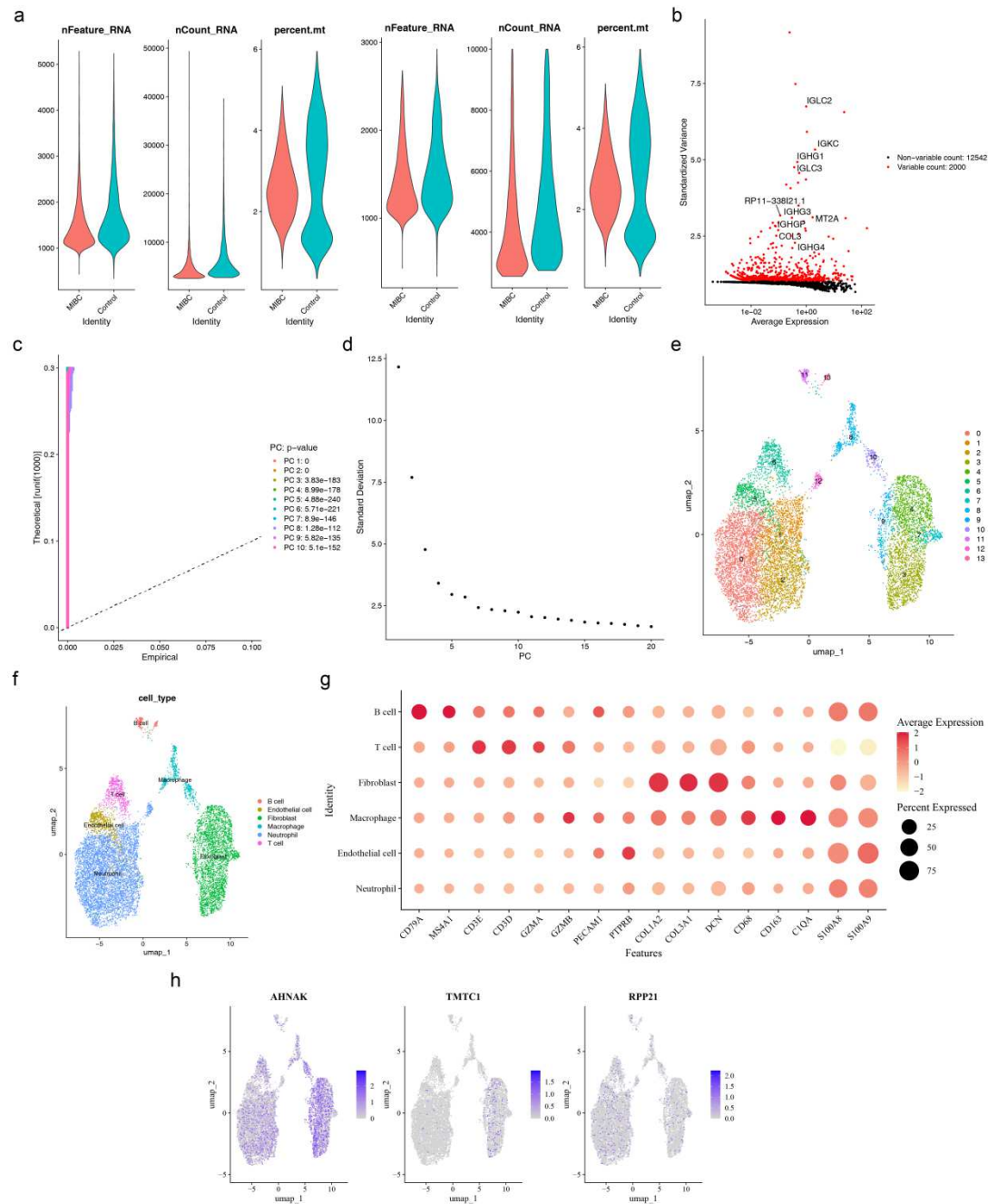


Fig. 5 Single-cell analysis and identification of key cells. (a) Quality control analysis. (b) Highly variable genes. (c) Principal component significance. (d) Scree plot. (e) Cell clusters. (f–g) Cell type annotation. (h) Expression of prognostic genes across cell types

3.6. The differentiation trajectory and related functions of fibroblasts

The fibroblasts were subjected to secondary dimensionality reduction and clustering, resulting in the formation of 3 cell subclusters (**Figure 6a**). During

1 the differentiation of fibroblasts from dark blue to light blue, the differentiation
2 process was divided into 3 states. Cells in MIBC samples were mostly
3 distributed at the terminal stage, whereas cells in the control group were
4 mainly located at the initial stage, indicating that the terminal stage of cell
5 differentiation was associated with MIBC progression (**Figure 6b**). AHNAK
6 was predominantly expressed at the initial stage, and its expression level
7 gradually decreased during differentiation and became stable starting from the
8 middle stage. RPP21 and TMTC1 were mainly distributed in the middle and
9 late stages, and their expression levels remained stable throughout the
10 differentiation process (**Figure 6c**). These findings indicated that AHNAK
11 might influence fibroblasts, and RPP21 and TMTC1 might be specifically
12 involved in fibroblast functions.

13 In addition, in control samples, fibroblasts exhibited close interactions with
14 multiple cell types, including T cells, neutrophils, endothelial cells, and
15 macrophages, and showed strong interactions with T cells and other cells
16 (**Figure 6d, e**). The COL1A2-CD44 ligand-receptor pair played an important
17 role in the process by which fibroblasts acted on macrophages (**Figure 6f**). In
18 MIBC samples, the number and strength of interactions between fibroblasts
19 and T cells, neutrophils, and other cells were reduced, and interactions among
20 other cell types were also decreased, such as those between endothelial cells
21 and T cells (**Figure 6d, e**). These results indicated that intercellular
22 interactions gradually decreased with the progression of MIBC. Multiple
23 ligand-receptor pairs displayed high activity when fibroblasts acted on other
24 cells; for example, MDK-SDC4 and MDK-NCL ligand-receptor pairs exerted
25 critical functions in the interactions between fibroblasts and endothelial cells
26 (**Figure 6g**). Subsequently, focusing on the cell cycle of fibroblasts, we found
27 that most cells were in the G1 phase, suggesting that the cells were in the
28 preparatory stage of the proliferation cycle and that cell proliferation might be
29 inhibited (**Figure 6h**). Collectively, these results revealed the association
30 between fibroblasts and MIBC progression from different perspectives.

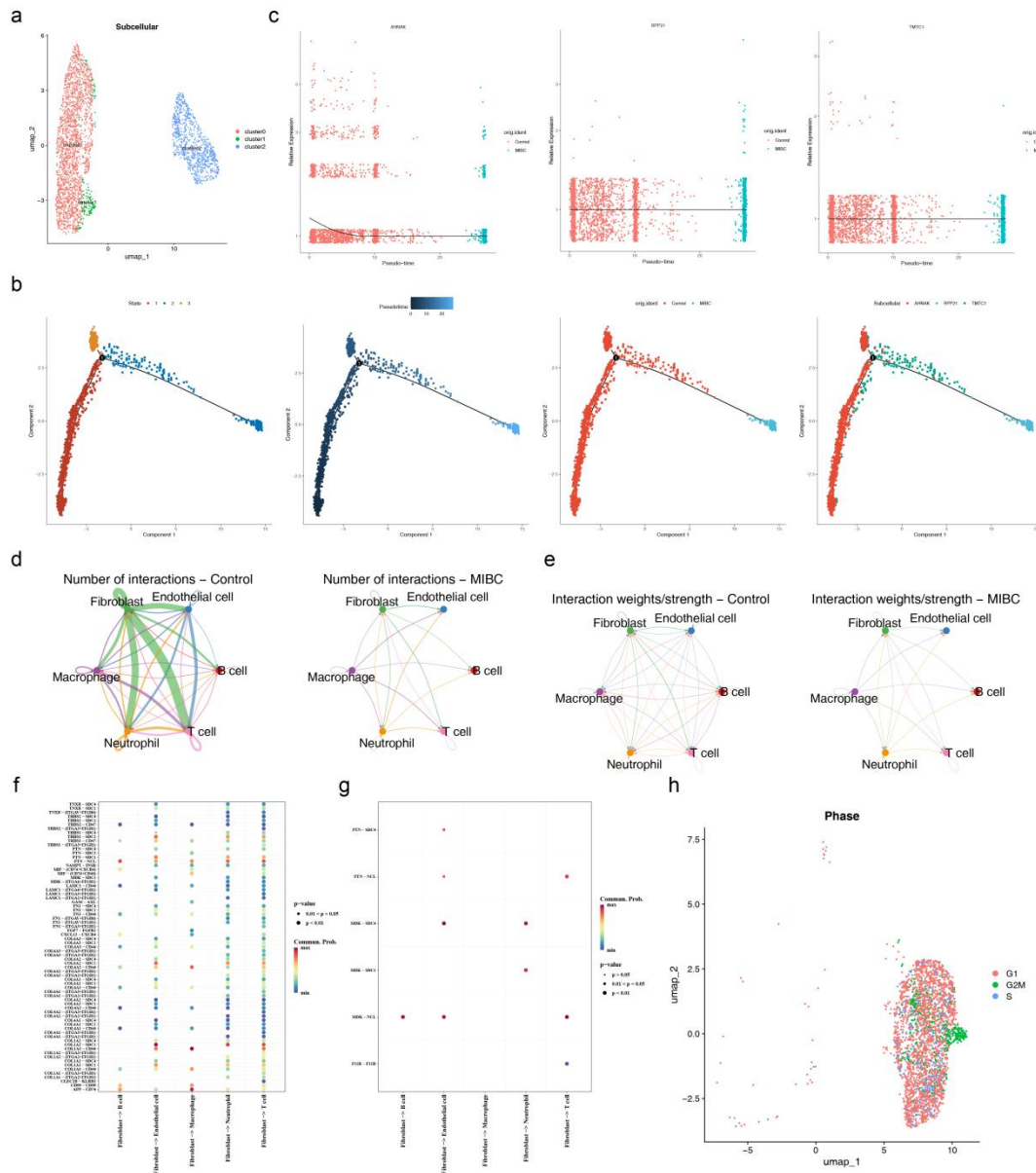


Fig. 6 Fibroblast differentiation trajectory, cell communication, and cell cycle analysis. (a) Fibroblast subclusters. (b) Pseudotime trajectory and cell states. (c) Expression dynamics of prognostic genes along pseudotime. (d–e) Cell communication networks and interaction strength. (f–g) Ligand-receptor interactions involving fibroblasts. (h) Cell cycle distribution of fibroblasts

3.7.Determination of 3 prognostic pathological features and prognostic pathological models for MIBC

In the TCGA-BLCA, the pathological sections of MIBC patients were collected. After processing these pathological sections, a total of 2,048 quantitative

1 image features were finally retained for analysis. Subsequently, a correlation
2 analysis was performed between these pathological features and the RNA
3 processing factor-associated gene scores, and 103 pathological features
4 satisfying $|\text{cor}| > 0.1$ and $P < 0.05$ were identified (**ESM 6**). Univariate Cox
5 regression analysis ($P < 0.05$, $\text{HR} \neq 1$) with PH assumption test ($P > 0.05$) was
6 conducted on these 103 pathological features, and 3 pathological features
7 were found to be associated with the prognosis of MIBC (**Figure 7a, ESM 7**).
8 The LASSO algorithm also identified the same 3 pathological features as being
9 associated with the prognosis of MIBC ($\text{lambda} = 0.00351$, **Figure 7b**).
10 Therefore, resnet1306, resnet1323, and resnet1743 were finally selected as
11 prognostic pathological features, with corresponding LASSO coefficients of
12 0.585, 0.713, and 0.371, respectively.

13 A prognostic pathological model was then constructed in the training cohort.
14 Based on the optimal cutoff value of 4.09, patients were stratified into HRG
15 and LRG, consisting of 81 and 68 individuals, respectively (**Figure 7c**).
16 Patients in the HRG exhibited significantly poorer survival ($P < 0.0001$, **Figure**
17 **7d**). The AUC values of the ROC curves all exceeded 0.6, indicating that the
18 aforementioned pathological model was reliable (**Figure 7e**). Identical results
19 were obtained in the validation cohort, which included 45 patients in the HRG
20 and 18 patients in the LRG (**Figure 7f-h**).

21 In summary, the above results indicated that resnet1306, resnet1323, and
22 resnet1743 had the ability to assess the prognosis of patients with MIBC.

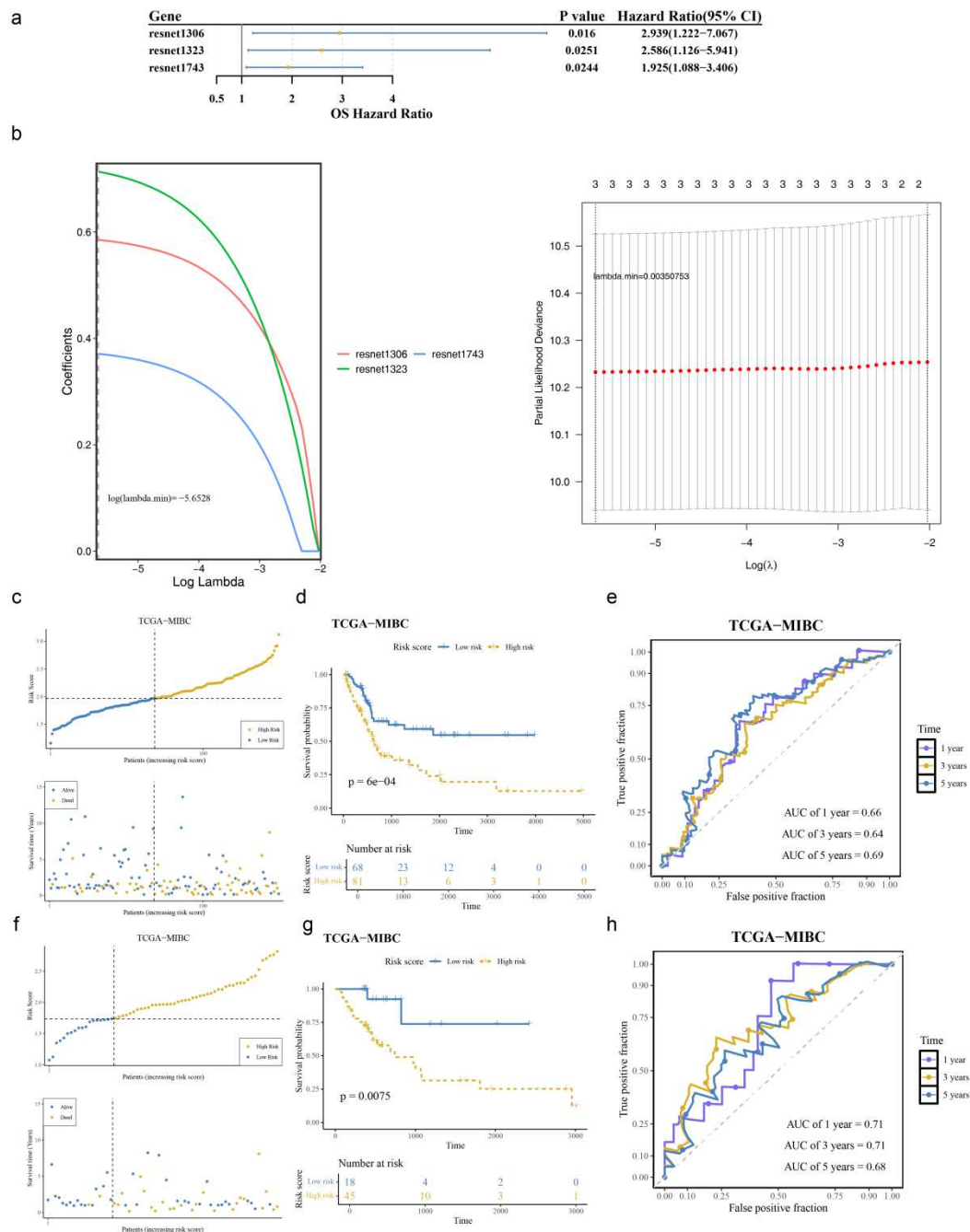
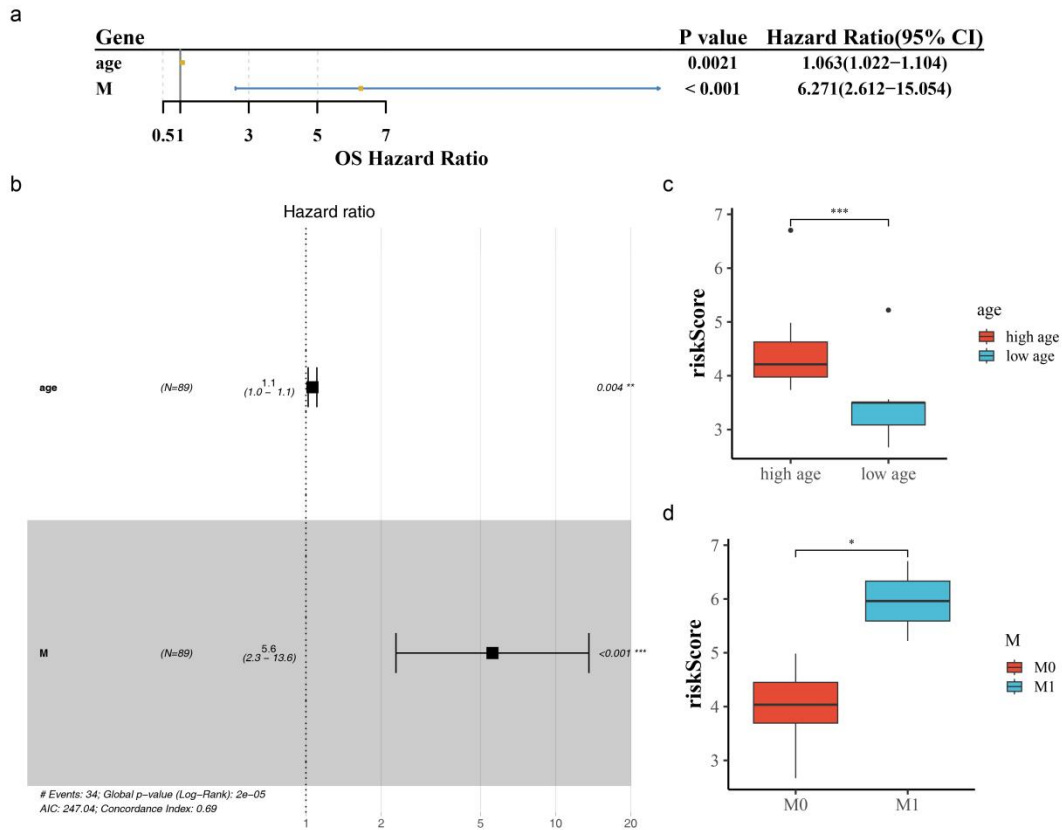


Fig. 7 Identification of prognostic pathological features and construction of prognostic pathological models. (a) Univariate Cox regression analysis of candidate pathological features. (b) LASSO regression analysis. (c–e) Prognostic pathological model in the training cohort. (f–h) Prognostic pathological model in the validation cohort

3.8.Acquisition of independent prognostic features: age, M stage

Clinical data of MIBC patients were collected from TCGA-BLCA. In the training

cohort, significant differences were observed between the MIBC and control groups in age, demographic.days to birth, diagnoses.age at diagnosis, M stage, T stage, and diagnoses.days to last follow up ($P < 0.05$, **ESM 8**). Univariate Cox regression analysis ($P < 0.05$, $HR \neq 1$) with the PH assumption test ($P > 0.05$) was performed on these 6 clinical features, and age and M stage were identified to be associated with the prognosis of MIBC (**Figure 8a**, **ESM 9**). Subsequently, multivariate Cox regression analysis ($P < 0.05$, $HR \neq 1$) with the PH assumption test ($P > 0.05$) defined age and M stage as prognostic clinical factors, with corresponding coefficients of 0.0593 and 1.7194, respectively (**Figure 8b**, **ESM 9**). The risk score for each sample was calculated accordingly. In the validation cohort, the risk score of the low-age group was significantly lower than that of the high-age group ($P < 0.001$, **Figure 8c**), and the risk score of the M1 stage group was significantly higher than that of the M0 stage group ($P < 0.05$, **Figure 8d**), indicating that M stage and age were significantly correlated with the risk of MIBC.



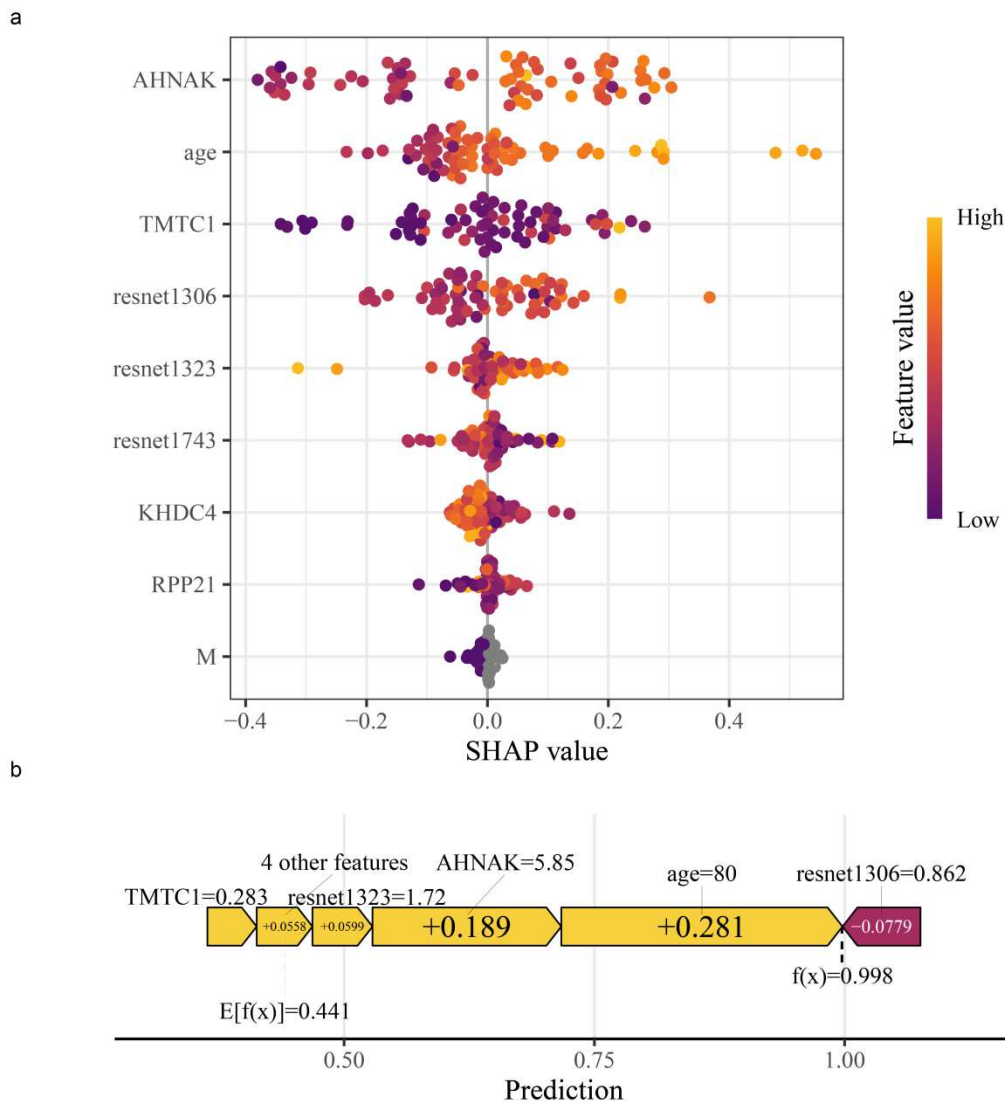
16

17 **Fig. 8** Identification and validation of prognostic clinical features. (a) Univariate
 18 Cox regression analysis. (b) Multivariate Cox regression analysis. (c) Risk

1 score comparison between age groups. (d) Risk score comparison between M
2 stage groups

3 3.9.The significance of different prognostic factors

4 The aforementioned prognostic genes, prognostic pathological features, and
5 prognostic clinical features were integrated to construct a SHAP model. This
6 model revealed that AHNAK exhibited the highest importance, followed by age,
7 resnet1306, TMTC1, and resnet1323 (**Figure 9a, b**). These results indicated
8 that AHNAK might exert a more significant effect on the prognosis of MIBC.



10 **Fig. 9** SHAP model for evaluating the importance of different prognostic
11 factors. (a–b) Feature importance and contribution patterns of integrated

1 prognostic factors

2 **4. Discussion**

3 Effective systemic treatments and novel approaches for MIBC remain limited,
4 especially metastatic MIBC [33]. Previous studies have shown that RNA
5 processing factors can serve as prognostic biomarkers for MIBC [8],
6 suggesting their potential research and development value. However, it
7 remains unclear how RNA processing factors influence the prognosis of MIBC,
8 and further investigation into the underlying mechanisms is needed. Based on
9 data from public databases, this study identified AHNAK, KHDC4, TMTC1, and
10 RPP21 as prognostic genes. The prognostic model constructed based on
11 these genes indicated that patients in the high-risk group had significantly
12 lower survival rates. By analyzing pathological sections of MIBC, we defined
13 resnet1306, resnet1323, and resnet1743 as prognostic pathological features.
14 The prognostic pathological model built using these features also
15 demonstrated that patients in the high-risk group had significantly lower
16 survival rates. Further analysis of clinical data confirmed that age and M stage
17 were clinical features associated with MIBC prognosis. Based on the
18 integration of the above three models, SHAP analysis ranked AHNAK as the
19 most important factor, followed by age and resnet1306. At the single-cell level,
20 fibroblasts were identified as key cells; among them, AHNAK showed relatively
21 high activity during fibroblast differentiation, further highlighting the critical role
22 of AHNAK in fibroblasts and suggesting that AHNAK may influence MIBC by
23 modulating fibroblast function. Overall, this study provides multifaceted data
24 support for elucidating the mechanism of action of RNA processing factors in
25 MIBC and identifying novel prognostic targets. In addition, our findings suggest
26 that the prognostic value of RNA-processing-related genes may not exist
27 independently, but rather reflect tumor microenvironment remodeling together
28 with pathological image features and fibroblast-associated states in MIBC.
29 Previous studies have shown that alternative splicing events in MIBC are

1 closely associated not only with patient prognosis, but also with immune
2 infiltration patterns and stromal characteristics[8]. Meanwhile, studies of
3 neoadjuvant immunotherapy have further demonstrated that the
4 immunosuppressive status of the tumor microenvironment can substantially
5 influence the response of MIBC patients to immune checkpoint inhibitors[33].
6 Furthermore, recent evidence indicates that the RNA-binding protein RBPM5
7 can regulate migration, invasion, and metastasis-related processes in bladder
8 cancer, suggesting a potential link between RNA-processing dysregulation and
9 tumor microenvironment remodeling. Combined with our findings, we
10 speculate that the ResNet-derived pathological features identified in this study
11 may partially capture spatial histological patterns associated with stromal
12 activation, immune cell redistribution, and CAF remodeling [34]. Therefore,
13 RNA-processing factors, pathological image features, and CAF-associated
14 microenvironmental changes may constitute an interconnected biological
15 network that collectively contributes to MIBC progression and prognosis.
16 This study identified four prognostic genes—AHNAK, KHDC4, TMTC1, and
17 RPP21—using a series of analytical approaches. AHNAK (AHNAK
18 nucleoprotein): In cancer tissues, AHNAK not only negatively regulates cell
19 proliferation but also exerts tumor-suppressive effects by enhancing TGF- β
20 signaling. It plays a critical role in the migration and invasion of various cancer
21 cells [35]. High expression of AHNAK has been reported to correlate with
22 poorer overall survival and disease-free survival in patients with
23 non-muscle-invasive bladder cancer (NMIBC), suggesting that AHNAK could
24 serve as a potential biomarker for predicting response to immunotherapy and
25 prognosis [36]. Moreover, AHNAK is a key scaffold protein that undergoes
26 extensive lysine lactylation. This widespread modification is closely linked to
27 RNA-processing pathways, including RNA splicing, transport, and mRNA
28 stability, and collectively contributes to the functional remodeling of endothelial
29 cells induced by bacterial infection [37]. While associations between
30 RNA-processing factors and NMIBC have been reported, our study
31 constructed an MIBC-related prognostic model based on AHNAK. Our findings

1 demonstrate that AHNAK is significantly associated with MIBC risk and serves
2 as a prognostic marker, which further underscores the importance of
3 RNA-processing factors in the progression of bladder cancer, particularly
4 MIBC. In summary, we hypothesize that AHNAK may influence the
5 progression and prognosis of MIBC by modulating RNA-processing-related
6 pathways. Emerging research highlights AHNAK's role as a large scaffold
7 protein in maintaining cytoskeletal integrity and mediating signal transduction,
8 including EMT and calcium signaling [38]. Although dysregulated RNA
9 modification is associated with the bladder cancer immune microenvironment
10 and tumor advancement [7], there is currently a lack of direct experimental
11 data regarding AHNAK-mediated RNA processing in MIBC. Given its dual
12 functionality in oncogenesis, it is crucial to employ knockdown/overexpression
13 strategies along with RNA immunoprecipitation sequencing or splicing profiling
14 to validate whether AHNAK serves as a cofactor for RNA-binding proteins,
15 thereby promoting MIBC malignant transformation via specific splicing events.

16 KHDC4 (KH domain-containing 4): KHDC4 is an RNA-binding protein that
17 primarily regulates gene expression by participating in the splicing and
18 processing of precursor mRNA. It is a key factor in maintaining the normal
19 physiological functions and protein synthesis of cells—especially those that
20 proliferate rapidly [39]. Furthermore, KHDC4 has been reported to be
21 associated with the progression of advanced prostate cancer. It is regulated by
22 E2F4 and acts in synergy with TRAF2 to promote the proliferation, migration,
23 and invasion of tumor cells, thereby facilitating the progression of the disease
24 to advanced and metastatic stages [40].

25 TMTC1 (transmembrane O-mannosyltransferase targeting cadherins 1):
26 TMTC1 is an endoplasmic reticulum (ER)-localized adaptor protein containing
27 tetratricopeptide repeat (TPR) motifs. Through its TPR domain, TMTC1
28 interacts with the ER calcium pump (SERCA) and acts as a negative regulator,
29 modulating ER calcium uptake and thereby maintaining intracellular calcium
30 homeostasis [41]. A recent pan-cancer analysis revealed that TMTC1
31 expression varies dramatically across different malignancies. In BC, TMTC1
32 expression is associated with tumor progression and correlates with a poorer
33 prognosis in the natural disease course. However, intriguingly, among BC

1 patients treated with the PD-L1 inhibitor Atezolizumab, high TMTC1
2 expression was associated with prolonged overall survival. This paradox
3 suggests that TMTC1 may serve as a promising biomarker for predicting the
4 response of BC patients to immune checkpoint inhibitor (ICI) therapy [42].

5 RPP21 (ribonuclease P subunit p21): RPP21 is a key protein subunit of the
6 ribonuclease P (RNase P) complex, primarily responsible for the 5'-end
7 processing of precursor tRNA and essential for tRNA maturation [43]. Notably,
8 studies have shown that RPP21, as a key factor in homologous recombination
9 repair (HDR), plays a dual role in cancer. On the one hand, loss of RPP21
10 function leads to impaired repair of DNA double-strand breaks and increased
11 genomic instability, thereby promoting tumor initiation. On the other hand,
12 given that RPP21 is often subject to gene amplification in various cancers, its
13 overexpression may enhance the ability of cancer cells to repair DNA damage
14 induced by therapy, thus contributing to resistance against radiotherapy and
15 chemotherapy [44].

16 This study identified three prognostic pathological features—resnet1306,
17 resnet1323, and resnet1743—via a series of analytical approaches.
18 Histopathological image assessment plays a critical role in cancer diagnosis
19 and subtype classification [45]. Residual networks (ResNet), as a type of deep
20 neural network, enable high-level feature learning without the need for manual
21 feature engineering, making them particularly well-suited for extending
22 diagnostic tasks to larger datasets. Currently, research on applying deep
23 neural networks to develop diagnostic models is still in its exploratory phase,
24 offering tremendous potential for further investigation [45]. Integrating
25 RNA-processing factors into pathological features can help pathologists
26 compute risk scores from histological images during routine diagnostic
27 procedures. Moreover, scoring systems built on RNA-processing factors can
28 serve as effective diagnostic tools, guiding oncologists in implementing
29 stratified, targeted therapies for patients [46, 47]. Therefore, this study
30 employed deep learning networks to extract case-specific features from MIBC
31 tissue sections. By correlating these features with gene scores related to
32 RNA-processing factors and employing various regression analysis methods,
33 we defined resnet1306, resnet1323, and resnet1743 as prognostic

1 pathological features and constructed corresponding prognostic models.
2 Based on risk scores, patients were stratified into high-risk and low-risk groups.
3 Those in the high-risk group exhibited significantly poorer prognoses,
4 demonstrating that resnet1306, resnet1323, and resnet1743 are associated
5 with MIBC prognosis [48]. Subsequent SHAP modeling revealed that
6 resnet1306 was deemed more important and may have a greater impact on
7 MIBC prognosis. It is important to note that these ResNet features are
8 essentially high-dimensional pathological phenotypes rather than directly
9 interpretable morphological metrics. As highlighted by the CHIEF pathology
10 foundation model [45], deep learning often captures subtle "dark matter"
11 regions that require mapping to specific pathological semantics. Therefore,
12 future studies should validate these features against interpretable
13 characteristics, such as the invasive tumor margin, stromal percentage,
14 necrotic regions, and lymphocytic infiltration. The Digital Spatial Profiling (DSP)
15 study by Eysers et al. provides a robust spatial framework for this purpose [49].
16 Correlating our high-dimensional features with their spatially defined
17 transcriptomic profiles will help elucidate the underlying molecular
18 mechanisms, bridging the gap between abstract AI predictions and concrete
19 biological insights.

20 Based on prognostic clinical features and model analysis, age and M stage
21 were identified as significant clinical prognostic factors. Previous studies have
22 shown that older age is associated with a lower likelihood of receiving radical
23 treatment among non-metastatic MIBC patients. Furthermore, patients who do
24 not receive such treatment have significantly poorer overall survival [50]. The
25 prognostic model developed in this study indicates that the older the age, the
26 greater the risk of MIBC. Therefore, while age alone should not preclude
27 radical treatment, the SHAP model underscores its significant importance in
28 MIBC and warrants due consideration. Although the prognostic relevance of M
29 stage in MIBC requires further validation, this study reveals that patients at the
30 M1 stage exhibit higher risk scores. This suggests that as the M stage
31 advances, the risk of MIBC escalates significantly. Consequently, both M stage
32 and age demonstrate substantial diagnostic potential for predicting MIBC
33 prognosis.

1 This study analyzed the expression levels of prognostic genes at the
2 single-cell level and identified fibroblasts as key cells. Previous studies have
3 established that distinct fibroblast clusters are significantly associated with
4 MIBC prognosis [51]. Cancer-associated fibroblasts (CAFs) drive MIBC
5 progression by modulating the microenvironment and mediating immune
6 suppression. Thus, they serve as vital prognostic biomarkers and therapeutic
7 targets [52]. Combined with pseudotime analysis results, AHNAK exhibited
8 relatively high expression during fibroblast differentiation, suggesting that
9 AHNAK might affect MIBC by influencing fibroblasts. At the transcriptomic level,
10 the SHAP model also highlighted AHNAK as the most critical factor for MIBC.
11 Furthermore, integrating findings from cell communication analyses, it was
12 observed that the frequency and intensity of intercellular interactions gradually
13 decline as the disease progresses; however, fibroblasts continue to maintain
14 strong interactions with T cells, endothelial cells, and other cell types. In line
15 with literature reports indicating that cancer-associated fibroblasts can mediate
16 T-cell dysfunction, ultimately contributing to tumor immune evasion [53], it
17 becomes evident that the perturbations in intercellular communication
18 mediated by fibroblasts can also impact MIBC progression. Taken together,
19 AHNAK may promote MIBC progression by targeting fibroblasts and thereby
20 altering their interactions with other cell types.

21 Our identification of fibroblasts as the primary carriers of AHNAK dysregulation
22 aligns with the emerging consensus that cancer-associated fibroblasts (CAFs)
23 are master regulators of the muscle-invasive bladder cancer (MIBC)
24 microenvironment. Recent studies have highlighted the spatial and functional
25 heterogeneity of CAFs, where specific subpopulations create a physical and
26 biochemical barrier against immune infiltration [52]. Mechanistically, the high
27 expression of AHNAK in fibroblasts may drive the rewiring of cell-cell
28 communication networks. We hypothesize that AHNAK potentiates the
29 interaction between fibroblasts and immune/endothelial cells through specific
30 ligand-receptor axes, such as COL1A2-CD44 and MDK-SDC4/NCL. The
31 COL1A2-CD44 interaction likely contributes to the dense extracellular matrix
32 (ECM) deposition, physically excluding T cells from tumor islets—a hallmark of
33 immune evasion [51]. Concurrently, the MDK-SDC4/NCL axis may aberrantly

1 activate receptor tyrosine kinase (RTK) signaling in endothelial cells and
2 macrophages, promoting abnormal angiogenesis and immunosuppressive
3 polarization. This disruption of normal intercellular crosstalk mirrors the
4 findings in recent literature, where CAF-derived factors induce T cell
5 dysfunction. Notably, a 2025 study demonstrated that CAF-secreted AEBP1
6 drives T cell exhaustion via the AKT/PD-L1 pathway [53]. We propose that
7 AHNAK operates within this immunosuppressive framework, potentially
8 stabilizing the CAF phenotype that actively suppresses cytotoxic T lymphocyte
9 (CTL) activity. Furthermore, the spatial distribution of AHNAK-high fibroblasts
10 may resemble the S3 clusters identified in MIBC, which are associated with
11 vascular mimicry and poor prognosis [51]. Collectively, these findings suggest
12 that AHNAK acts as a linchpin, maintaining the CAFs in an activated state that
13 systematically reprograms the tumor stroma to favor immune escape and
14 vascular remodeling.

15 **Limitations**

16 This study integrated multi-omics approaches to identify genes, pathological
17 features, and clinical characteristics associated with the prognosis of MIBC,
18 providing a theoretical foundation for the diagnosis and treatment of MIBC
19 from multiple perspectives. However, this study has so far focused solely on
20 identifying prognostic genes at the bioinformatics level. In the future, studies
21 should further explore the effects of gene knockout or overexpression on
22 MIBC-related mechanisms by combining in vitro cell experiments or in vivo
23 animal studies. In addition, the prognostic pathological features identified in
24 this study also require further in-depth investigation. This study still requires
25 validation of model robustness in independent MIBC cohorts, prospective
26 clinical samples, and multicenter pathological slide datasets. In addition,
27 further investigations using immunohistochemistry, multiplex
28 immunofluorescence, spatial transcriptomics, and functional cellular
29 experiments are needed to clarify the causal relationship among AHNAK
30 expression, fibroblast states, RNA-processing alterations, and patient
31 prognosis.

1 5. Conclusions

2 This study integrated bulk transcriptome sequencing, single-cell RNA
3 sequencing, pathological image features, and clinical data to identify
4 prognostic genes and clinicopathological features associated with RNA
5 processing factors in MIBC. AHNAK, KHDC4, TMTC1, and RPP21 were
6 identified as prognostic genes, while resnet1306, resnet1323, resnet1743, age,
7 and M stage showed prognostic relevance. These findings provide potential
8 biomarkers and a theoretical basis for prognostic assessment and targeted
9 therapy in MIBC.

10 List of abbreviations

Abbreviation	Full name
MIBC	Muscle-invasive bladder cancer
BC	Bladder cancer
NMIBC	Non-muscle-invasive bladder cancer
TCGA	The Cancer Genome Atlas
BLCA	Bladder urothelial carcinoma
GEO	Gene Expression Omnibus
scRNA-seq	Single-cell RNA sequencing
DEGs	Differentially expressed genes
FC	Fold change
adj.P	Adjusted P value
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
BP	Biological process
CC	Cellular component
MF	Molecular function
PPI	Protein-protein interaction
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins

PH	Proportional hazards
HR	Hazard ratio
LASSO	Least Absolute Shrinkage and Selection Operator
coef	Coefficient
exp	Expression level
HRG	High-risk group
LRG	Low-risk group
K-M	Kaplan-Meier
ROC	Receiver operating characteristic
AUC	Area under the curve
IC ₅₀	Half-maximal inhibitory concentration
GDSC	Genomics of Drug Sensitivity in Cancer
PCA	Principal component analysis
PCs	Principal components
UMAP	Uniform Manifold Approximation and Projection
WSIs	Whole-slide images
BGR	Blue-green-red
RGB	Red-green-blue
ssGSEA	Single-sample Gene Set Enrichment Analysis
SHAP	SHapley Additive exPlanations
CAFs	Cancer-associated fibroblasts
ECM	Extracellular matrix
CTL	Cytotoxic T lymphocyte
TGF- β	Transforming growth factor-beta
ER	Endoplasmic reticulum
TPR	Tetratricopeptide repeat
SERCA	Sarco/endoplasmic reticulum calcium ATPase
RNase P	Ribonuclease P
HDR	Homology-directed repair

ICI	Immune checkpoint inhibitor
DSP	Digital Spatial Profiling
RTK	Receptor tyrosine kinase

1 **Declarations**

2 **Ethics approval and consent to participate**

3 Not applicable.

4 **Clinical trial number**

5 Not applicable.

6 **Consent for publication**

7 Not applicable.

8 **Availability of data and materials**

9 The transcriptome, clinical, and histopathology data analyzed in this study are
10 publicly available from the TCGA-BLCA project through the Genomic Data
11 Commons Data Portal (<https://portal.gdc.cancer.gov/>). The single-cell RNA
12 sequencing dataset GSE135337 is publicly available from the Gene
13 Expression Omnibus database
14 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE135337>). The RNA
15 processing factor-associated gene set used in this study was derived from the
16 published literature cited in the manuscript and is provided in **ESM 1**. All
17 processed result tables supporting the conclusions of this article are included
18 in the article and its Supplementary Information files. No new datasets were
19 generated. Analysis scripts and additional implementation details are available
20 from the corresponding author on reasonable request.

21 **Competing interests**

22 The authors declare that they have no competing interests.

23 **Funding**

24 This research was funded by the Nature Science Foundation of Shanxi
25 Province, grant number 202103021224412.

1 **Authors' contributions**

2 Lina Ji: Conceptualization, Data curation, Validation, Visualization, Writing–
3 original draft, Writing–review & editing. Yishuang Guo: Data curation,
4 Validation, Visualization, Writing–review & editing. Yuedong Hao: Validation,
5 Writing–review & editing. Liang Wei: Conceptualization, Project administration,
6 Supervision, Writing–review & editing. This work currently described has not
7 been published, is not being considered for publication elsewhere, and its
8 publication was approved by all authors.

9 **Acknowledgements**

10 We would like to express our sincere gratitude to all individuals and
11 organizations who supported and assisted us throughout this research. In
12 conclusion, we extend our thanks to everyone who has supported and assisted
13 us along the way. Without your support, this research would not have been
14 possible.

15 **References**

- 16 1. Dobruch J, Oszczudłowski M. Bladder Cancer: Current Challenges and
17 Future Directions. *Medicina*. 2021;57(8):749.
18 <https://doi.org/10.3390/medicina57080749>.
- 19 2. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al.
20 Global cancer statistics 2022: GLOBOCAN estimates of incidence and
21 mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for*
22 *Clinicians*. 2024;74(3):229-63. <https://doi.org/10.3322/caac.21834>.
- 23 3. Teoh JY-C, Huang J, Ko WY-K, Lok V, Choi P, Ng C-F, et al. Global Trends of
24 Bladder Cancer Incidence and Mortality, and Their Associations with Tobacco
25 Use and Gross Domestic Product Per Capita. *European Urology*.
26 2020;78(6):893-906. <https://doi.org/10.1016/j.eururo.2020.09.006>.
- 27 4. Sternberg CN, de Mulder P, Schornagel JH, Theodore C, Fossa SD, van
28 Oosterom AT, et al. Seven year update of an EORTC phase III trial of high-dose
29 intensity M-VAC chemotherapy and G-CSF versus classic M-VAC in advanced

- 1 urothelial tract tumours. *European Journal of Cancer*. 2006;42(1):50-4.
2 <https://doi.org/10.1016/j.ejca.2005.08.032>.
- 3 5. Bellmunt J, de Wit R, Vaughn DJ, Fradet Y, Lee J-L, Fong L, et al.
4 Pembrolizumab as Second-Line Therapy for Advanced Urothelial Carcinoma.
5 *New England Journal of Medicine*. 2017;376(11):1015-26.
6 <https://doi.org/10.1056/NEJMoa1613683>.
- 7 6. Osadska M, Selicky T, Kretova M, Jurcik J, Sivakova B, Cipakova I, et al. The
8 Interplay of Cohesin and RNA Processing Factors: The Impact of Their
9 Alterations on Genome Stability. *International Journal of Molecular Sciences*.
10 2022;23(7):3939. <https://doi.org/10.3390/ijms23073939>.
- 11 7. Wu J, Wang X, Xu H, Tian J, Ji H, Zhu J, et al. Bioinformatics analysis of the
12 correlation between m6A RNA methylation regulators and the immune
13 infiltration and prognosis of bladder cancer. *Annals of Translational Medicine*.
14 2022;10(24):1386-. <https://doi.org/10.21037/atm-22-5993>.
- 15 8. Azhati B, Wu G, Zhan H, Liang W, Song Z, Lu L, et al. Alternative splicing
16 patterns reveal prognostic indicator in muscle-invasive bladder cancer. *World*
17 *Journal of Surgical Oncology*. 2022;20(1).
18 <https://doi.org/10.1186/s12957-022-02685-0>.
- 19 9. Cheng C, Chen W, Jin H, Chen X. A Review of Single-Cell RNA-Seq
20 Annotation, Integration, and Cell–Cell Communication. *Cells*.
21 2023;12(15):1970. <https://doi.org/10.3390/cells12151970>.
- 22 10. Wang Y, Song W, Feng C, Wu S, Qin Z, Liu T, et al. Multi-omics analysis
23 unveils the predictive value of IGF2BP3/SPHK1 signaling in cancer stem cells
24 for prognosis and immunotherapeutic response in muscle-invasive bladder
25 cancer. *Journal of Translational Medicine*. 2024;22(1).
26 <https://doi.org/10.1186/s12967-024-05685-8>.
- 27 11. Liu Z, Mao X, Xie Y, Yan Y, Wang X, Mi J, et al. Single-cell RNA sequencing
28 reveals a fibroblast gene signature that promotes T-cell infiltration in

1 muscle-invasive bladder cancer. *Communications Biology*. 2025;8(1).
2 <https://doi.org/10.1038/s42003-025-08094-9>.

3 12. National Cancer Institute. The Cancer Genome Atlas Bladder Urothelial
4 Carcinoma (TCGA-BLCA) [dataset]. Genomic Data Commons.
5 <https://portal.gdc.cancer.gov/projects/TCGA-BLCA> (accessed 26 Aug 2026).

6 13. Wang J, Gao W. Transcriptomic Characterization of Local Invasion in
7 Human Bladder Cancer by Single Cell RNA Sequencing [dataset]. *Gene*
8 *Expression Omnibus*. GSE135337. 2021.
9 <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE135337>.

10 14. Lou S, Meng F, Yin X, Zhang Y, Han B, Xue Y. Comprehensive
11 Characterization of RNA Processing Factors in Gastric Cancer Identifies a
12 Prognostic Signature for Predicting Clinical Outcomes and Therapeutic
13 Responses. *Frontiers in Immunology*. 2021;12.
14 <https://doi.org/10.3389/fimmu.2021.719628>.

15 15. Gustavsson EK, Zhang D, Reynolds RH, Garcia-Ruiz S, Ryten M.
16 ggtranscript: an R package for the visualization and interpretation of transcript
17 isoforms using ggplot2. *Bioinformatics*. 2022;38(15):3844-6.
18 <https://doi.org/10.1093/bioinformatics/btac409>.

19 16. Zhang X, Chao P, Zhang L, Xu L, Cui X, Wang S, et al. Single-cell RNA and
20 transcriptome sequencing profiles identify immune-associated key genes in the
21 development of diabetic kidney disease. *Frontiers in Immunology*. 2023;14.
22 <https://doi.org/10.3389/fimmu.2023.1030198>.

23 17. Mao X. Inhibition of cell survival and invasion by Tanshinone IIA via FTH1: A
24 key therapeutic target and biomarker in head and neck squamous cell
25 carcinoma. *Experimental and Therapeutic Medicine*. 2022.
26 <https://doi.org/10.3892/etm.2022.11449>.

27 18. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R Package for Comparing
28 Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative*
29 *Biology*. 2012;16(5):284-7. <https://doi.org/10.1089/omi.2011.0118>.

1 19. Shannon P, Markiel A, Ozier O. Cytoscape: A Software Environment for
2 Integrated Models of Biomolecular Interaction Networks. *Genome Research*.
3 2003;13(11):2498-504. <https://doi.org/10.1101/gr.1239303>.

4 20. van den Berg MEL, Castellote JM, de Pedro-Cuesta J, Mahillo-Fernandez I.
5 Survival after Spinal Cord Injury: A Systematic Review. *Journal of Neurotrauma*.
6 2010;27(8):1517-28. <https://doi.org/10.1089/neu.2009.1138>.

7 21. Yang La, Qu Q, Hao Z, Sha K, Li Z, Li S, et al. Powerful Identification of
8 Large Quantitative Trait Loci Using Genome-Wide R/glmnet-Based Regression.
9 *Journal of Heredity*. 2022;113(4):472-8. <https://doi.org/10.1093/jhered/esac006>.

10 22. Song M, Zhang Q, Song C, Liu T, Zhang X, Ruan G, et al. The advanced
11 lung cancer inflammation index is the optimal inflammatory biomarker of overall
12 survival in patients with lung cancer. *Journal of Cachexia, Sarcopenia and*
13 *Muscle*. 2022;13(5):2504-14. <https://doi.org/10.1002/jcsm.13032>.

14 23. Yu C, Zhang Y. Establishment of prognostic nomogram for elderly colorectal
15 cancer patients: a SEER database analysis. *BMC Gastroenterology*.
16 2020;20(1). <https://doi.org/10.1186/s12876-020-01464-z>.

17 24. Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of
18 clinical chemotherapeutic response from tumor gene expression levels.
19 *F1000Research*. 2014;3:228. <https://doi.org/10.12688/f1000research.5471.1>.

20 25. Revelle W. psych: Procedures for Psychological, Psychometric, and
21 Personality Research. Comprehensive R Archive Network (CRAN); 2024.

22 26. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al.
23 Integrated analysis of multimodal single-cell data. *Cell*.
24 2021;184(13):3573-87.e29. <https://doi.org/10.1016/j.cell.2021.04.048>.

25 27. Song H, Xie G, Li Y, Hu X, Yang Z, Zhao Y, et al. A single-cell atlas of
26 bladder cancer unveils dynamic cellular composition and endothelial functional
27 shifts during progression. *Discover Oncology*. 2025;16(1).
28 <https://doi.org/10.1007/s12672-025-02297-6>.

- 1 28. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based
2 analysis of lung single-cell sequencing reveals a transitional profibrotic
3 macrophage. *Nature Immunology*. 2019;20(2):163-72.
4 <https://doi.org/10.1038/s41590-018-0276-y>.
- 5 29. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The
6 dynamics and regulators of cell fate decisions are revealed by pseudotemporal
7 ordering of single cells. *Nature Biotechnology*. 2014;32(4):381-6.
8 <https://doi.org/10.1038/nbt.2859>.
- 9 30. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan C-H, et al.
10 Inference and analysis of cell-cell communication using CellChat. *Nature*
11 *Communications*. 2021;12(1). <https://doi.org/10.1038/s41467-021-21246-9>.
- 12 31. Liang D, Wang L, Zhong P, Lin J, Chen L, Chen Q, et al. Perspective: Global
13 Burden of Iodine Deficiency: Insights and Projections to 2050 Using XGBoost
14 and SHAP. *Advances in Nutrition*. 2025;16(3).
15 <https://doi.org/10.1016/j.advnut.2025.100384>.
- 16 32. Huang T-F, Luo C, Guo L-B, Liu H-Z, Li J-T, Lin Q-Z, et al. Preoperative
17 prediction of textbook outcome in intrahepatic cholangiocarcinoma by
18 interpretable machine learning: A multicenter cohort study. *World Journal of*
19 *Gastroenterology*. 2025;31(11). <https://doi.org/10.3748/wjg.v31.i11.100911>.
- 20 33. Peyrottes A, Ouzaid I, Califano G, Hermieu J-F, Xylinas E. Neoadjuvant
21 Immunotherapy for Muscle-Invasive Bladder Cancer. *Medicina*. 2021;57(8):769.
22 <https://doi.org/10.3390/medicina57080769>.
- 23 34. Yu J, Chen L, Wang G, Qian K, Weng H, Yang Z, et al. RBPMS inhibits
24 bladder cancer metastasis by downregulating MYC pathway through
25 alternative splicing of ANKRD10. *Communications Biology*. 2025;8(1).
26 <https://doi.org/10.1038/s42003-025-07842-1>.
- 27 35. Cai Y, Hu Y, Yu F, Tong W, Wang S, Sheng S, et al. AHNAK suppresses
28 ovarian cancer progression through the Wnt/ β -catenin signaling pathway.
29 *Aging*. 2021;13(20):23579-87. <https://doi.org/10.18632/aging.203473>.

36. Patwardhan MV, Kane TQ, Chiong E, Rahmat JN, Mahendran R. Loss of Glutathione-S-Transferase Theta 2 (GSTT2) Modulates the Tumor Microenvironment and Response to BCG Immunotherapy in a Murine Orthotopic Model of Bladder Cancer. *International Journal of Molecular Sciences*. 2024;25(24):13296. <https://doi.org/10.3390/ijms252413296>.
37. Wu Q, Li Z, Gong T, Zheng X, Zhou X, Peng X. *Porphyromonas gingivalis* infection induces lysine lactylation reprogramming in human umbilical vein endothelial cells. *Frontiers in Cellular and Infection Microbiology*. 2026;16. <https://doi.org/10.3389/fcimb.2026.1706727>.
38. Zhang S, Cai Z, Li H. AHNAKs roles in physiology and malignant tumors. *Frontiers in Oncology*. 2023;13. <https://doi.org/10.3389/fonc.2023.1258951>.
39. Löscher M, Schosserer M, Dausse E, Lee K, Ajuh P, Grillari-Voglauer R, et al. Inhibition of Pre-mRNA Splicing by a Synthetic Blom7 α -Interacting Small RNA. *PLoS ONE*. 2012;7(10):e47497. <https://doi.org/10.1371/journal.pone.0047497>.
40. Hu S-W, Wu C-C, Dong S-W, Tzou K-Y, Chen C-H, Wang Y-H, et al. Implications of the KHDC4-TRAF2 axis in the context of prostate cancer prognosis. *Aging*. 2025. <https://doi.org/10.18632/aging.206273>.
41. Sunryd JC, Cheon B, Graham JB, Giorda KM, Fissore RA, Hebert DN. TMTC1 and TMTC2 Are Novel Endoplasmic Reticulum Tetratricopeptide Repeat-containing Adapter Proteins Involved in Calcium Homeostasis. *Journal of Biological Chemistry*. 2014;289(23):16085-99. <https://doi.org/10.1074/jbc.M114.554071>.
42. Zhang Y, Wu D, Yu T, Liu Y, Zhao C, Xue R. Prognostic value of TMTC1 in pan-cancer analysis. *Heliyon*. 2024;10(19). <https://doi.org/10.1016/j.heliyon.2024.e38308>.
43. Jarrous N, Reiner R, Wesolowski D, Mann H, Guerrier-Takada C, Altman S. Function and subnuclear distribution of Rpp21, a protein subunit of the human

1 ribonucleoprotein ribonuclease P. RNA. 2001;7(8):1153-64.
2 <https://doi.org/10.1017/S1355838201010469>.

3 44. Abu-Zhayia ER, Khoury-Haddad H, Guttman-Raviv N, Serruya R, Jarrous
4 N, Ayoub N. A role of human RNase P subunits, Rpp29 and Rpp21, in
5 homology directed-repair of double-strand breaks. Scientific Reports.
6 2017;7(1). <https://doi.org/10.1038/s41598-017-01185-6>.

7 45. Wang X, Zhao J, Marostica E, Yuan W, Jin J, Zhang J, et al. A pathology
8 foundation model for cancer diagnosis and prognosis prediction. Nature.
9 2024;634(8035):970-8. <https://doi.org/10.1038/s41586-024-07894-z>.

10 46. Breen J, Allen K, Zucker K, Adusumilli P, Scarsbrook A, Hall G, et al.
11 Artificial intelligence in ovarian cancer histopathology: a systematic review. NPJ
12 Precision Oncology. 2023;7(1):83.
13 <https://doi.org/10.1038/s41698-023-00432-6>.

14 47. Desbois M, Udyavar AR, Ryner L, Kozlowski C, Guan Y, Dürbaum M, et al.
15 Integrated digital pathology and transcriptome analysis identifies molecular
16 mediators of T-cell exclusion in ovarian cancer. Nature Communications.
17 2020;11(1):5583. <https://doi.org/10.1038/s41467-020-19408-2>.

18 48. Zhang Y, Fang M, Wang X, Ying Z, Jiang S, Lu Y, et al. Development of a
19 PANoptosis-Related Pathomics Prognostic Model in Ovarian Cancer: A
20 Multi-Omics Study. Journal of Cellular and Molecular Medicine.
21 2025;29(22):e70958. <https://doi.org/10.1111/jcmm.70958>.

22 49. Eysers M, Irlam J, Marshall G, Smith V, Baker A, Frost L, et al. Digital spatial
23 profiling of the microenvironment of muscle invasive bladder cancer.
24 Communications Biology. 2024;7(1).
25 <https://doi.org/10.1038/s42003-024-06426-9>.

26 50. van Hoogstraten LMC, Witjes JA, Meijer RP, Ripping TM, BlaZIB study
27 group, Kiemeney LA, et al. Non-metastatic muscle-invasive bladder cancer: the
28 role of age in receiving treatment with curative intent. BJU International.
29 2022;130(6):764-775. <https://doi.org/10.1111/bju.15697>.

- 1 51. Feng C, Wang Y, Song W, Liu T, Mo H, Liu H, et al. Spatially-resolved
2 analyses of muscle invasive bladder cancer microenvironment unveil a distinct
3 fibroblast cluster associated with prognosis. *Frontiers in Immunology*.
4 2024;15:1522582. <https://doi.org/10.3389/fimmu.2024.1522582>.
- 5 52. Burley A, Rullan A, Wilkins A. A review of the biology and therapeutic
6 implications of cancer-associated fibroblasts (CAFs) in muscle-invasive
7 bladder cancer. *Frontiers in Oncology*. 2022;12.
8 <https://doi.org/10.3389/fonc.2022.1000888>.
- 9 53. Wang X, Li J, Song D, Wu Y, Liu J, Yi Z, et al. AEBP1 drives
10 fibroblast-mediated T cell dysfunction in tumors. *Nature Communications*.
11 2025;16(1). <https://doi.org/10.1038/s41467-025-63659-w>.

12

13 **Online Resource**

- 14 **ESM 1:** RNA processing factor-associated genes.
- 15 **ESM 2:** GO and KEGG enrichment results of candidate genes.
- 16 **ESM 3:** PH assumption test results of prognostic genes.
- 17 **ESM 4:** PH assumption test results for independent prognostic factor analysis.
- 18 **ESM 5:** Drug sensitivity analysis results.
- 19 **ESM 6:** Candidate pathological features correlated with RNA processing
20 factor-associated gene scores.
- 21 **ESM 7:** Prognostic pathological features identified by Cox regression analysis.
- 22 **ESM 8:** Differential clinical characteristics in the training cohort.
- 23 **ESM 9:** Prognostic clinical features and PH assumption test results.

Supplementary Files

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

- [ESM1.xlsx](#)
- [ESM7.xlsx](#)
- [ESM5.xlsx](#)
- [ESM3.xlsx](#)
- [ESM8.xlsx](#)
- [ESM4.xlsx](#)
- [ESM6.xlsx](#)
- [ESM9.xlsx](#)
- [ESM2.xlsx](#)