

Construction and Validation of an RNA-Binding Protein-Related Bladder Cancer Prognostic Model Based on Bioinformatics

Jun Li, Ying Xiong, Qi Zhao, Xiaowu Pi, Hao Huang, Feng Guo, Hui Pan*

The First Affiliated Hospital of Yangtze University, Jingzhou, China

Email: *panhui1963@163.com

How to cite this paper: Li, J., Xiong, Y., Zhao, Q., Pi, X.W., Huang, H., Guo, F. and Pan, H. (2022) Construction and Validation of an RNA-Binding Protein-Related Bladder Cancer Prognostic Model Based on Bioinformatics. *Yangtze Medicine*, 6, 66-75. <https://doi.org/10.4236/ym.2022.63006>

Received: June 29, 2022

Accepted: August 8, 2022

Published: August 11, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Objective: This study aimed to construct a bladder cancer prognostic model using bioinformatics to predict the survival of bladder cancer patients. **Methods:** RNA sequences and corresponding clinical data were downloaded from The Cancer Genome Atlas (TCGA) database, and the differentially expressed RNA-binding proteins (RBPs) were screened for analysis using the limma software package. Then, GO enrichment analysis and KEGG pathway analysis were performed on these differentially expressed RNA-binding proteins, and a PPI network was constructed. Finally, a risk model was constructed based on the screened central RBP, and a Kaplan-Meier survival curve was drawn to evaluate the prognostic value of central RBP and predict the prognosis of bladder Cancer (BLCA) patients with this model. Finally, the human protein atlas (HPA) online database (<http://www.proteinatlas.org/>) was used to further detect the differential expression of central hub RBP at the protein level between tumor tissue and normal tissue. **Results:** The bladder cancer prognostic model constructed with these six central RBPs had good sensitivity and specificity in predicting the prognosis of bladder cancer patients. **Conclusion:** This study explored the genes and regulatory networks of bladder cancer prognosis-related RNA-binding protein and bladder cancer, and constructed a bladder cancer prognosis model, which provides a theoretical basis for the development of new bladder cancer prognosis biomarkers in the future.

Keywords

RNA Binding Protein, Bladder Cancer, Differentially Expressed RBP, Bioinformatics

1. Introduction

Bladder cancer (BLCA) refers to malignant tumors that occur in the bladder mucosa. It is a common malignant tumor of the urinary system, and its morbidity and mortality have been on the rise in recent years [1] [2]. Early-stage bladder cancer has a better prognosis, but it is easy to recur and develop into muscle-invasive bladder cancer (muscle-invasive bladder cancer, MIBC). In addition, MIBC has strong invasive and metastatic ability and poor prognosis, with a 5-year survival rate of less than 50% [3] [4]. Therefore, there is an urgent need to identify early diagnostic biomarkers and prognostic indicators to improve treatment outcomes and survival in bladder cancer.

RNA-binding proteins (RBPs) interact with RNAs to form ribonucleoprotein complexes that regulate RNA expression and function [5]. As an important player in post-transcriptional regulation, RBP is involved in almost all post-transcriptional regulatory processes, including RNA splicing, translation, transport, localization, degradation and stabilization [6]. RBP dysregulation has been reported in a variety of cancers, affecting tumorigenesis and progression [6]. However, the role of RBP-related mechanisms in cancer development remains uncertain. Therefore, clarifying the role of RBPs in BLCA will help us better understand tumor pathogenesis and develop prognostic and response biomarkers.

Previously, most studies have focused on the correlation between a single or limited number of RBPs and BLCA. A comprehensive study of RBP function will help us fully understand its role in BLCA. Therefore, this study downloaded RNA-seq data and corresponding clinical information on BLCA from The Cancer Genome Atlas (TCGA) database to screen for RBPs differentially expressed between tumor and normal samples. Subsequently, a series of bioinformatic analysis methods were performed based on these differential RBPs, and 6 independent prognostic RBPs were finally identified, which were then used to construct a prognostic survival model. The results of this study may contribute to the establishment of an RBP-based prognostic assessment model for BLCA patients.

2. Materials and methods

1) Data processing. The RNA sequence and corresponding clinical data were downloaded from the TCGA database (<https://portal.gdc.cancer.gov/>), including 19 normal samples and 414 BLCA samples. We use the limma software package (<http://www.bioconductor.org/packages/release/bioc/html/limma.html>) for analysis. By using the criteria of false discovery rate (FDR) < 0.05 and $|\log_2 \text{fold change (FC)}| > 1$, differentially expressed RBPs were screened.

2) GO enrichment analysis and KEGG pathway. Through GO enrichment and KEGG pathway analysis (using R package DOSE, cluster profile, rich plot, ggplot2, etc.), the biological functions of differentially expressed RBP were systematically checked. Both P and FDR values < 0.05 were considered statistically significant.

3) Protein-protein interaction (PPI) network construction and module screening. Submit RBPs with different expressions to the STRING database (<http://www.string-db.org/>) to detect PPI [7]. Then use Cytoscape 3.7.0 to build and visualize the PPI network. The Molecular Complexity Detection (MCODE) plug-in is used to screen the key modules whose MCODE score and node count are both > 5 [8] from the PPI network. $P < 0.05$ was considered statistically significant.

4) Prognostic model construction. Through univariate and multivariate Cox regression, six independent prognostic-related RBPs were determined. Then, a risk-scoring model was constructed based on the expression levels and coefficients of the RBP of the six hubs. Use the following formula to calculate the risk score of each BLCA patient: risk score = $\beta_1 * \text{Exp1} + \beta_2 * \text{Exp2} + \beta_i * \text{Exp}$, where β represents independent prognostic-related RBP, and the coefficient value proficient represents the expression level of RBP independent of prognosis, I Represents the i -th center RBP.

5) Verify the prognostic value and expression level of central RBP. The prognostic value of the 6 RBPs in BLCA was evaluated by using the log-rank test to draw the Kaplan-Meier survival curve. The human protein atlas (HPA) online database (<http://www.proteinatlas.org/>) is used to study the differential expression of 6 hubs RBPs at the protein level between tumor tissues and normal tissues.

3. Results

1) Screening of differentially expressed RBP. We obtained RNA sequencing data and clinical information from the TCGA database, which contains 414 BLCA tissues and 19 normal tissues. In this study, the expression value of 1542 RBP was analyzed [9]. Using the DEseq software package meeting the $P < 0.05$ and $|\log_2 \text{FC}| > 1.0$ criteria, a total of 385 differentially expressed RBPs were identified, including 218 up-regulated and 167 down-regulated RBPs. The heat map and volcano map of these differentially expressed RBPs are shown in **Figure 1**.

2) The enrichment analysis of GO and KEGG pathways of differentially expressed RBP. In order to study the potential functions and mechanisms of the identified RBP, we divided these differentially expressed RBPs into up-regulated and down-regulated groups and conducted enrichment analysis of GO and KEGG pathways. GO enrichment analysis shows that the biological processes of up-regulated RBPs are mainly concentrated in ncRNA metabolism and processing and tRNA metabolism, while down-regulated RBPs are mainly enriched in RNA splicing and regulation and cellular amide metabolism. Cell component analysis showed that both up-regulated and down-regulated RBPs are mainly rich in cytoplasmic ribonucleoprotein and ribonucleoprotein particles. Molecular function analysis showed that up-regulated RBPs enriched catalytic activity to a large extent and acted on RNA and ribonuclease activities; at the same time, down-regulated RBPs were mainly enriched in translation factor regulatory activities and activities and RNA binding. KEGG pathway enrichment analysis

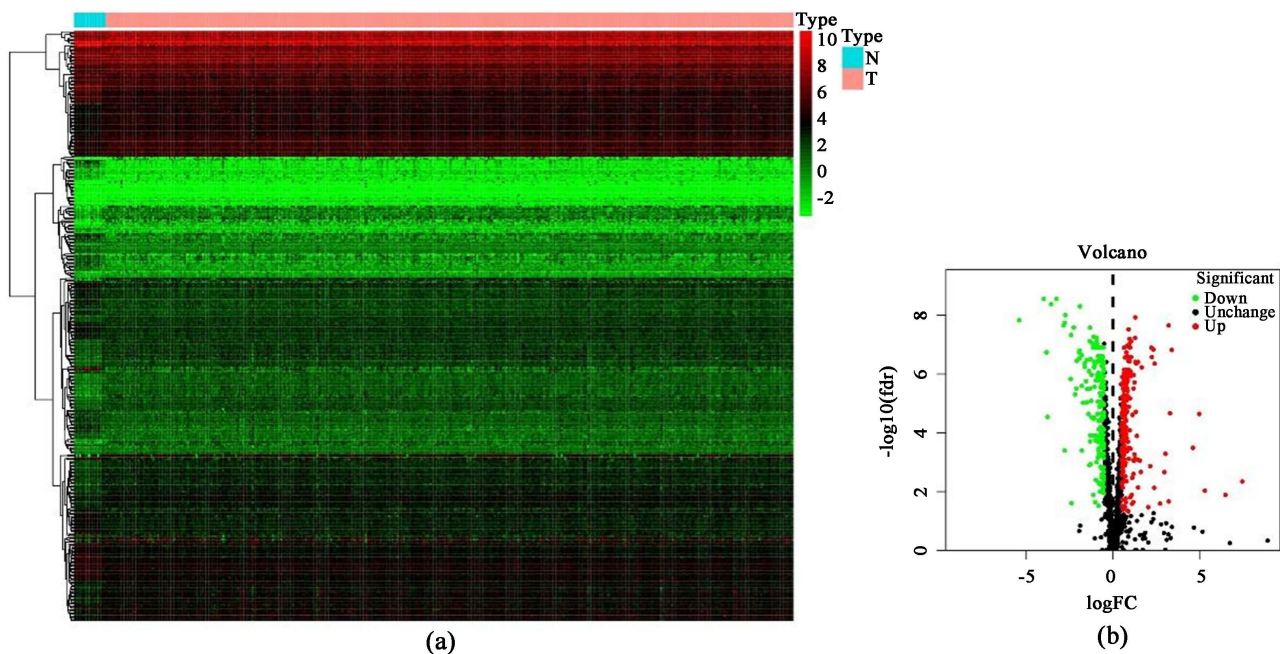


Figure 1. (a) Differential RBPs; (b) 218 up-regulated and 167 down-regulated.

showed that up-regulated RBP was significantly enriched in RNA transport, spliceosomes, and ribosomes, while down-regulated RBP was enriched in messenger RNA monitoring pathways and RNA transport and degradation.

3) PPI network construction and key module screening. In order to further explore the role of differential RBP in BLCA, based on the STRING database, Cytoscape was used to establish a PPI network with 373 nodes and 4062 edges. Subsequently, we used the MODE tool to analyze the co-expression network to identify potential key modules. The most important modules include 104 nodes and 1151 edges. KEGG pathway analysis shows that RBP in these key modules is enriched in eukaryotes, spliceosomes, mRNA monitoring pathways, RNA polymerase, cytoplasmic DNA sensing pathways, RNA transport, RNA degradation, and ribosomes.

4) Identifying RBPs Related to Prognosis. A total of 373 key differential RBPs were screened from the PPI network. To determine the association between RBP and the prognosis of BLCA patients, univariate Cox regression analysis was performed to assess the prognostic value of these key differential RBPs. These identified RBPs identified 23 central RBPs (**Figure 2(a)**). Subsequently, a multivariate Cox regression analysis was performed to further analyze these 23 RBPs, and the results showed that the 6 central RBPs are independent prognostic indicators for BLCA patients (**Figure 2(b)**).

5) Construction and analysis of prognostic-related risk scoring models. We established a prognostic-related risk scoring model based on 6 independent prognostic-related RBPs. In the model, we calculate the risk score of each BLCA patient according to the following formula: $\text{risk score} = \beta_1 \times \text{Exp1} + \beta_2 \times \text{Exp2} + \beta_i \times \text{Exp}_i$, where β is the regression coefficient and Exp is the expression level.

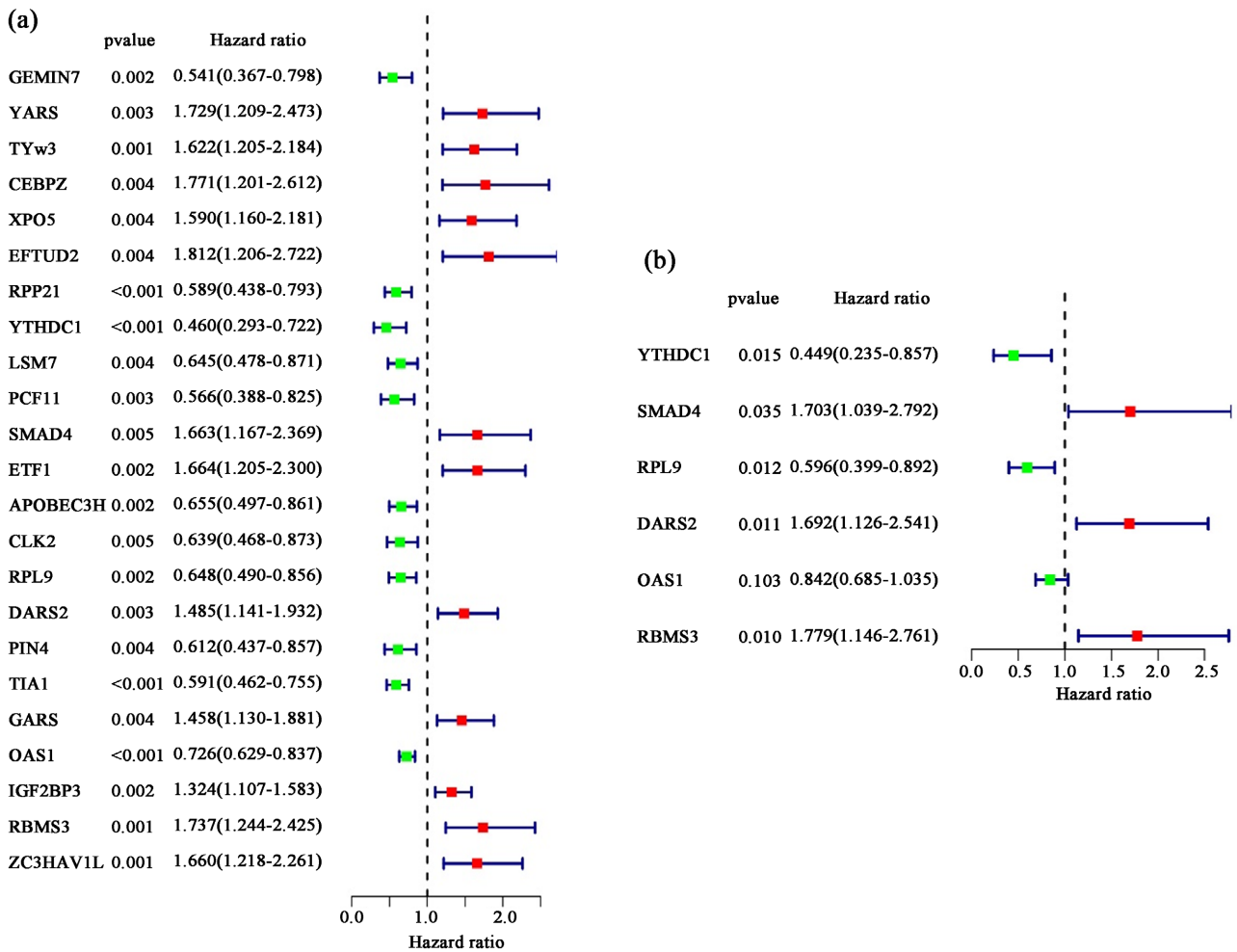


Figure 2. (a) RBP associated with BC prognosis in univariate Cox regression analysis; (b) RBP associated with BC prognosis in multivariate Cox regression analysis.

According to the median risk score in the formula, the training group is divided into low-risk and high-risk cohorts. Thereafter, according to the median of the training group and the risk score in the formula, the test group was also divided into low-risk and high-risk cohorts. The results of the training group showed that the OS of the high-risk group had poorer OS for the lower-risk group (**Figure 3**). ROC analysis showed the prognostic value of RBP in 6 centers. The area under the ROC curve (AUC) of the training group model was 0.721 (**Figure 4(a)**), indicating that it has better diagnostic capabilities. In order to assess whether the risk scoring model has the same prognostic significance in the test group, the same formula is used in the test group. It was found that the OS of the high-risk group was poorer than that of the low-risk group (**Figure 3**), and the area under the ROC curve was 0.685 (**Figure 4(b)**). Therefore, it suggests that the model has better sensitivity and specificity for prognosis.

6) Assessing the prognostic value of clinical parameters. Cox regression analysis is used to evaluate the impact of different clinical characteristics on the prognosis of BLCA patients. Univariate Cox regression analysis results showed

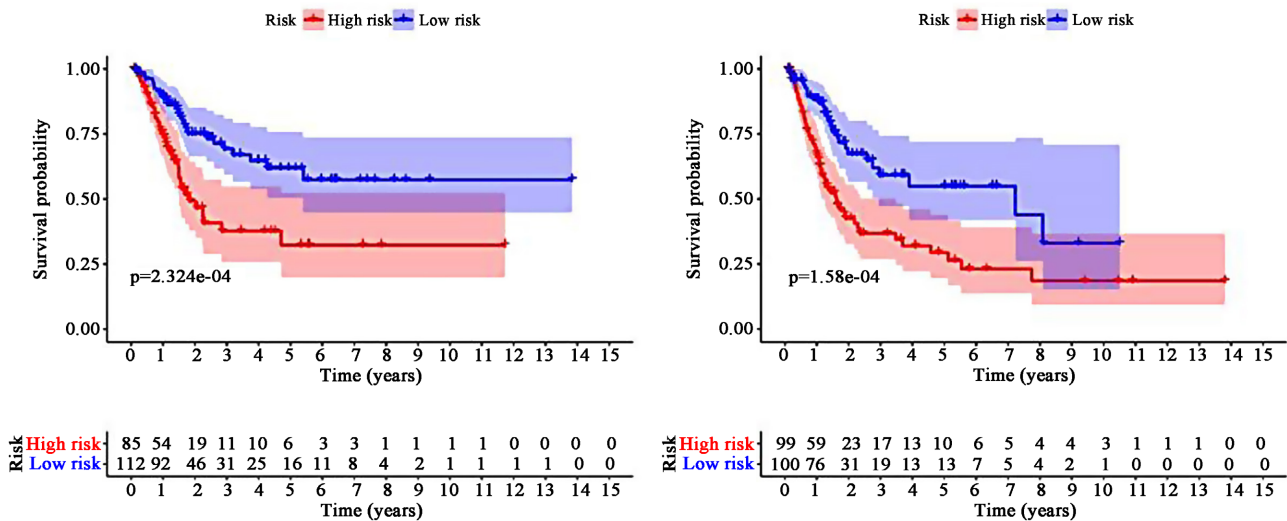


Figure 3. Survival curves of low-risk and high-risk subgroups.

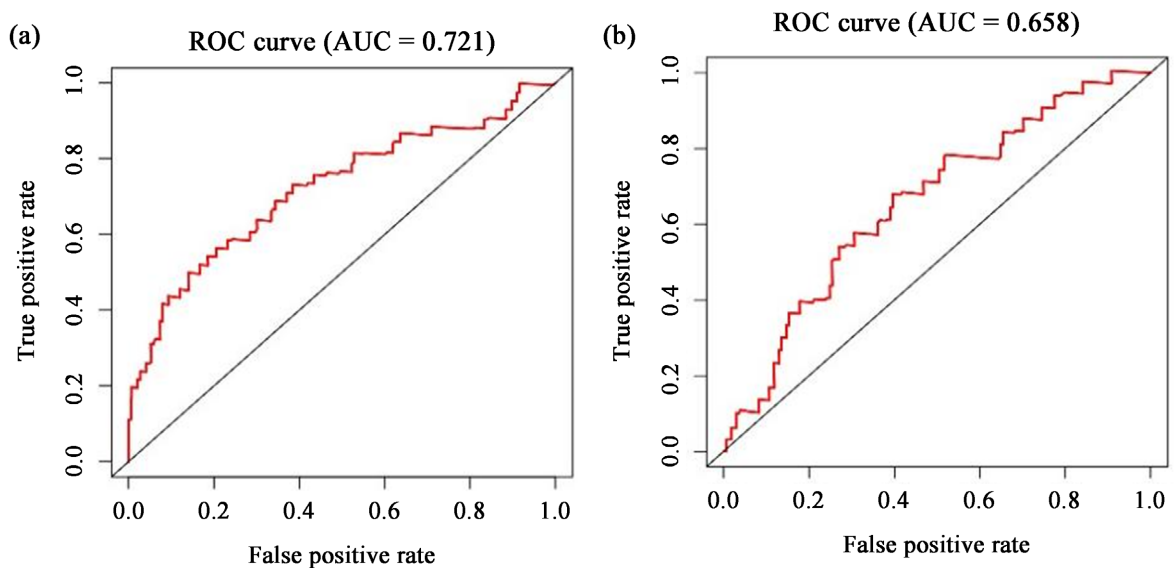


Figure 4. ROC curves for predicting OS based on risk score.

that in the training group and the test group, age, stage, and risk score were all related to the OS of BLCA patients (Figure 5). The results of multivariate Cox regression analysis showed that age, stage, and risk score were independent prognostic factors related to OS in the training and testing groups (Figure 5).

4. Discussions

Despite recent promising results in early diagnosis and multimodal treatment of bladder cancer, metastatic disease is often incurable and its 5-year survival rate is still only 15% [10]. Metastasis and recurrence are the main causes of death in bladder cancer patients, especially MIBC [11]. Therefore, it is of great significance to further understand the molecular mechanism of bladder cancer and develop effective early screening and diagnosis methods to improve the treatment

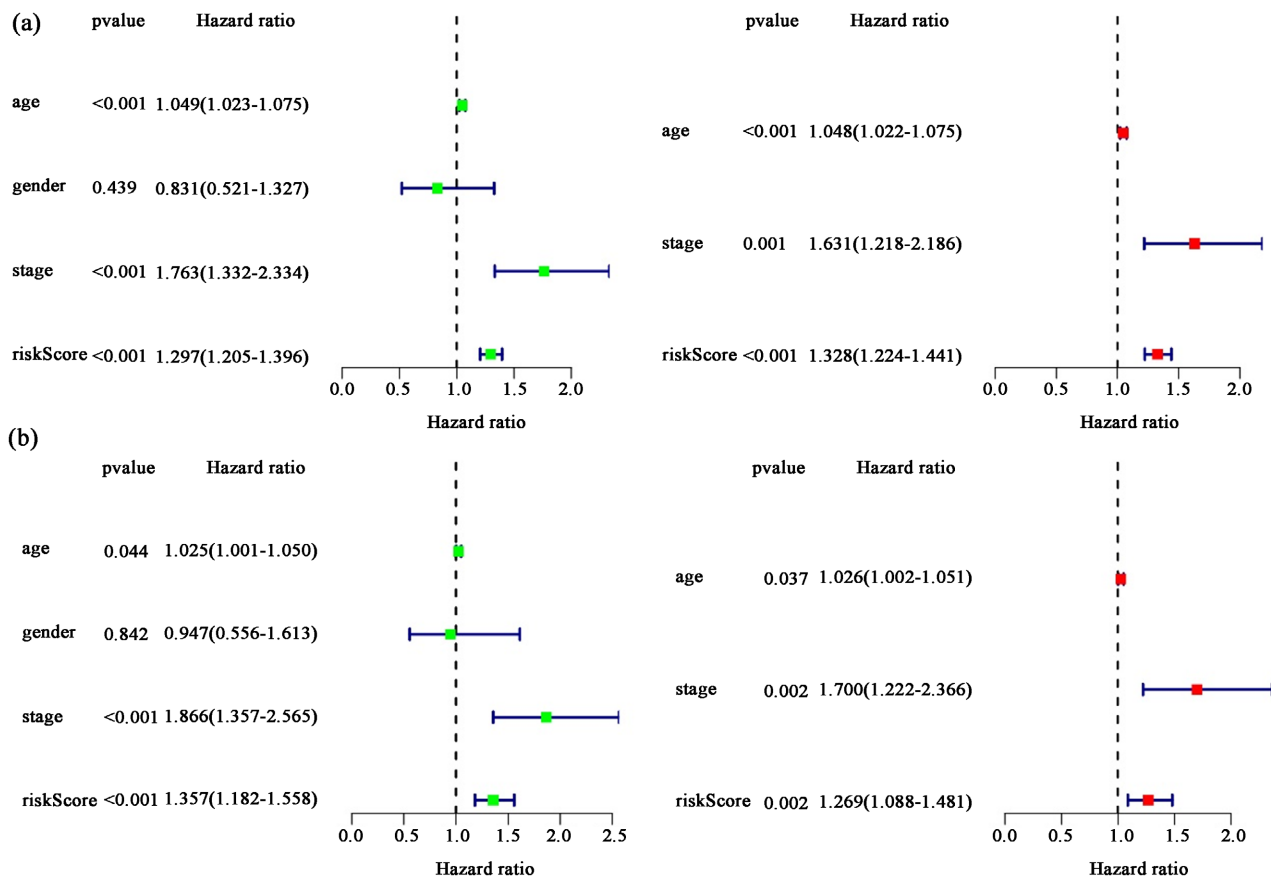


Figure 5. Results of univariate independent analysis (left) and multivariate independent analysis (right) of train group (a) and test (b).

effect and quality of life of patients. In this study, we first screened 385 RBPs that were differentially expressed between BLCA and normal tissues from the TCGA database. Then, we systematically analyzed the biological pathways of these differential RBPs and constructed a PPI network. Subsequently, we performed univariate and multivariate Cox regression analyses to further identify 6 independent RBPs associated with prognosis. To further understand their biological functions and clinical significance, we also performed survival and ROC analysis on the 6 hub RBPs. Finally, we constructed a risk model based on these 6 prognostic center RBPs to predict the prognosis of BLCA patients. Our findings may provide new biomarkers for prognostic assessment of BLCA patients.

In our study, BLCA RNA sequencing data were integrated to identify DERBP between bladder cancer tissue and normal bladder tissue. Univariate Cox regression analysis was used to screen candidate pivotal RBPs related to prognosis, and multivariate Cox regression analysis was used to identify pivotal RBPs related to prognosis. Finally, we identified the following 6 central RBPs: RPL9, OAS1, YTHDC1, DARS2, RBMS3, SMAD4. Using multivariate Cox regression analysis, based on the training group data, 6 RBPs were used to construct a risk scoring model to predict the prognosis of BLCA patients. In the training group, the ROC curve of the RBP risk scoring model has moderate OS predictive power (AUC = 0.721),

and the overall survival time of high-risk BLCA patients is significantly shortened. In the test group, as a validation cohort, the ROC curve of the RBP risk scoring model also had moderate OS predictive power (AUC = 0.685), and the overall survival time of high-risk BLCA patients was significantly shortened. The C index of the nomogram in the training group was 0.7033, and in the test group, it was 0.6295. The purpose of establishing the nomogram is to enable professionals to predict the 1-year, 3-year, and 5-year OS of BLCA patients. According to the prediction results of the risk score model, the prognosis of patients with high-risk scores is poor, indicating that the treatment plan and individualized treatment may need to be adjusted. We further proved that using GEPIA, the expression levels of DARS2, RBMS3, and SMAD4 in BLCA tissues were significantly higher than those in normal bladder tissues. The expression levels of RPL9, OAS1, and YTHDC1 were significantly lower than those of normal bladder tissue. In addition, using the Human Protein Atlas database, the expression of DARS2, RBMS3, and SMAD4 in bladder cancer tissue was significantly higher than that in normal bladder tissue. However, the staining levels of RPL9, OAS1, and YTHDC1 in bladder cancer tissues are relatively low.

A previous study reported that OAS1 was identified as an interferon-induced antiviral enzyme, which was recently associated with 5-azacytidine (AZA) sensitivity, and its lack led to NCI-60 anti-AZA cancer cell lines [12]. YTHDC1 is an N⁶-methyladenosine binding protein located in the YT body near the nuclear spot, which regulates mRNA splicing by recruiting splicing factors to target mRNA [13]. DARS2 promotes cell cycle progression and inhibits apoptosis of liver cancer cells through the miR-30e-5p/MAPK/NFAT5 pathway [14]. The loss of RBMS3 in epithelial ovarian cancer not only induces chemoresistance to platinum but also promotes recurrence through miR-126-5p/ β -catenin/CBP signaling. In addition, the loss of RBMS3 is related to the low overall survival rate and recurrence-free survival of patients with epithelial ovarian cancer [15]. Another study found that RBMS3 inhibits the proliferation, migration, and invasion of breast cancer cells through the Wnt/ β -catenin signaling pathway [16]. Studies have shown that siRNA inhibition of RPL9 can inhibit the growth of colorectal cancer (CRC) cells and the formation of long-term colonies by increasing the number of sub-G1 cells and strongly inducing apoptotic cell death [17]. Transforming growth factor- β (TGF- β) regulates cell function and plays a key role in the development of pancreatic cancer. Smad4 belongs to the family of signal transduction proteins. It is phosphorylated and activated by the transmembrane serine-threonine receptor kinase through transforming growth factor β (TGF- β) signaling across multiple pathways. This gene acts as a tumor suppressor gene [18]. SMAD4, as one of the Smads signal transducer families from TGF- β , mediates pancreatic cell proliferation and apoptosis [19]. However, the functions and molecular mechanisms of these central RBPs in BLCA are still poorly understood.

All in all, this study provides new insights into the function of RBP in the oc-

currence and development of BLCA tumors. In addition, the model has better predictive power in terms of survival, which may help to develop new BLCA prognostic biomarkers. However, this study has some limitations. First, our findings are based only on RNA sequencing, and no other omics data. Second, a risk-scoring model should be established based on TCGA BLCA data, and prospective studies should be conducted to prove this. In addition, TCGA data lacks certain clinical characteristics, which may reduce the statistical validity and reliability of multiple Cox regression analyses. Finally, because we have adopted bioinformatics methods, further biological experiments are needed to verify their claims.

5. Ethics Approval and Consent to Participate

The data used in our study were obtained from public databases TCGA. Therefore, ethical approval was not required.

Acknowledgements

We would like to thank the teachers of the Research Center of the First Affiliated Hospital of Yangtze University and Yangtze University School of Medicine for their help and the patients and researchers who participated in TCGA for providing the data.

Funding

This study was supported by the Youth Talents of the Hubei Provincial Health Council (Grant No. WJ2021Q014, WJ2017Q041).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] John, B.A. and Said, N. (2017) Insights from Animal Models of Bladder Cancer: Recent Advances, Challenges, and Opportunities. *Oncotarget*, **8**, 57766-57781. <https://doi.org/10.18632/oncotarget.17714>
- [2] Banerjee, S. and Southgate, J. (2019) Bladder Organoids: A Step towards Personalised Cancer Therapy? *Translational Andrology and Urology*, **8**, S300-s302. <https://doi.org/10.21037/tau.2019.06.10>
- [3] Tse, J., *et al.* (2019) Current Advances in BCG-Unresponsive Non-Muscle Invasive Bladder Cancer. *Expert Opinion on Investigational Drugs*, **28**, 757-770. <https://doi.org/10.1080/13543784.2019.1655730>
- [4] Crijnen, J. and De Reijke, T.M. (2018) Emerging Intravesical Drugs for the Treatment of Non Muscle-Invasive Bladder Cancer. *Expert Opinion on Emerging Drugs*, **23**, 135-147. <https://doi.org/10.1080/14728214.2018.1474201>
- [5] Gerstberger, S., Hafner, M. and Tuschl, T. (2014) A Census of Human RNA-Binding Proteins. *Nature Reviews Genetics*, **15**, 829-845. <https://doi.org/10.1038/nrg3813>

- [6] Pereira, B., Billaud, M. and Almeida, R. (2017) RNA-Binding Proteins in Cancer: Old Players and New Actors. *Trends in Cancer*, **3**, 506-528. <https://doi.org/10.1016/j.trecan.2017.05.003>
- [7] Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., *et al.* (2019) STRING v11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Research*, **47**, D607-D603. <https://doi.org/10.1093/nar/gky1131>
- [8] Bader, G.D. and Hogue, C.W. (2003) An Automated Method for Finding Molecular Complexes in Large Protein Interaction Networks. *BMC Bioinformatics*, **4**, Article No. 2. <https://doi.org/10.1186/1471-2105-4-2>
- [9] Gerstberger, S., Hafner, M. and Tuschl, T. (2014) A Census of Human RNA-Binding Proteins. *Nature Reviews Genetics*, **15**, 829-845. <https://doi.org/10.1038/nrg3813>
- [10] Nadal, R. and Bellmunt, J. (2019) Management of Metastatic Bladder Cancer. *Cancer Treatment Reviews*, **76**, 10-21. <https://doi.org/10.1016/j.ctrv.2019.04.002>
- [11] Richters, A., Aben, K.K.H. and Kiemenev, L.A.L.M. (2020) The Global Burden of Urinary Bladder Cancer: An Update. *World Journal of Urology*, **38**, 1895-1904. <https://doi.org/10.1007/s00345-019-02984-4>
- [12] Banerjee, S., Gusho, E., Gaughan, C., Dong, B., Gu, X., Holvey-Bates, E., *et al.* (2019) OAS-RNase L Innate Immune Pathway Mediates the Cytotoxicity of a DNA-Demethylating Drug. *Proceedings of the National Academy of Sciences of the United States of America*, **116**, 5071-5076. <https://doi.org/10.1073/pnas.1815071116>
- [13] Xiao, W., Adhikari, S., Dahal, U., Chen, Y.S., Hao, Y.J., Sun, B.F., *et al.* (2016) Nuclear m⁶A Reader YTHDC1 Regulates mRNA Splicing. *Molecular Cell*, **61**, 507-519. <https://doi.org/10.1016/j.molcel.2016.01.012>
- [14] Qin, X., Li, C., Guo, T., Chen, J., Wang, H.T., *et al.* (2017) Upregulation of DARS2 by HBV Promotes Hepatocarcinogenesis through the miR-30e-5p/MAPK/NFAT5 Pathway. *Journal of Experimental & Clinical Cancer Research*, **36**, Article No. 148. <https://doi.org/10.1186/s13046-017-0618-x>
- [15] Wu, G., Cao, L., Zhu, J., Tan, Z., *et al.* (2019) Loss of RBMS3 Confers Platinum Resistance in Epithelial Ovarian Cancer via Activation of miR-126-5p/beta-Catenin/CBP Signaling. *Clinical Cancer Research*, **25**, 1022-1035. <https://doi.org/10.1158/1078-0432.CCR-18-2554>
- [16] Yang, Y., Quan, L. and Ling, Y. (2018) RBMS3 Inhibits the Proliferation and Metastasis of Breast Cancer Cells. *Oncology Research*, **26**, 9-15. <https://doi.org/10.3727/096504017X14871200709504>
- [17] Baik, I.H., Jo, G.-H., Seo, D., Ko, M.J., *et al.* (2016) Knockdown of RPL9 Expression Inhibits Colorectal Carcinoma Growth via the Inactivation of Id-1/NF- κ B Signaling Axis. *International Journal of Oncology*, **49**, 1953-1962. <https://doi.org/10.3892/ijo.2016.3688>
- [18] McCarthy, A.J. and Chetty, R. (2018) Smad4/DPC4. *Clinical Pathology*, **71**, 661-664. <https://doi.org/10.1136/jclinpath-2018-205095>
- [19] Xia, X., Wu, W. and Huang, C. (2015) SMAD4 and Its Role in Pancreatic Cancer. *Tumor Biology*, **36**, 111-119. <https://doi.org/10.1007/s13277-014-2883-z>