Comprehensive analysis of the functions and prognostic significance of RNA-binding proteins in bladder urothelial carcinoma

Changgang Guo1,2*, Ting Shao3*, Xiling Jiang4*, Dadong Wei1, Zhanhua Wang1, Mingyang Li1, Guochang Bao1,2

1Department of Urology, Affiliated Hospital of Chifeng University, Chifeng 024000, China; 2Urology Research Center, Chifeng University, Chifeng 024000, China; 3Department of Gynecology, Affiliated Hospital of Chifeng University, Chifeng 024000, China; 4Department of Stomatology, Affiliated Hospital of Chifeng University, Chifeng 024000, China. *Equal contributors.

Received August 5, 2020; Accepted October 16, 2020; Epub November 15, 2020; Published November 30, 2020

Abstract: Alterations in RNA-binding proteins (RBPs) are reported in various cancer types; however, the role of RBPs in bladder urothelial cancer (BLCA) remains unknown. This study aimed to systematically examine the function and prognostic significance of RBPs in BLCA using bioinformatics analyses. RNA sequencing and clinical data for BLCA were downloaded from The Cancer Genome Atlas (TCGA) database, and differentially expressed RBPs (DERBPs) between normal and cancer tissues were identified. A total of 388 DERBPs were identified, including 219 upregulated and 169 downregulated RBPs. All RBPs were screened for the prognostic model establishment and 9 RBPs (TRIM71, YTHDC1, DARS2, XPOT, ZNF106, FTO, IPO7, EFTUD2, and CTU1) were regarded as prognosis-related hub RBPs in BLCA. Further analysis revealed worse overall survival (OS) in the high-risk cohort compared to the model-based low-risk cohort. The area under the receiver operating characteristic (ROC) curve was 0.752 in the training group and 0.701 in the testing group, which supports the strength of its predictive ability. A nomogram was established according to nine prognosis-related RBPs, which showed strong predictive ability for BLCA. The C-indices of the nomogram were 0.7033 in the training group, and 0.6295 in the testing group. The prognosis-related hub RBPs may be involved in oncogenesis, development, and metastasis of BLCA. Our results will be of great significance in revealing the pathogenesis of BLCA, and developing new therapeutic targets and prognostic molecular markers for BLCA.

Keywords: RNA-binding proteins, bladder urothelial cancer, differentially expressed RBPs, overall survival, predictive ability

Introduction

Bladder urothelial cancer (BLCA) is the tenth most common malignant tumor in the world. More than five hundred thousand new cases of bladder cancer and two hundred thousand related deaths are estimated to have occurred in 2018; it is more common in men than in women [1]. Based on pathological diagnosis, bladder cancer can be categorized into non-muscle invasive bladder cancer (NMIBC) and muscle invasive bladder cancer (MIBC). Most BLCA cases originate from epithelial cells, of which approximately 90% are urothelial tumors, whereas squamous and glandular tumors are the less common histologic subtypes; bladder cancer very rarely originates from mesenchymal cells [2]. General treatment includes operation, intravesical treatment, radical treatment, immunotherapy and radiotherapy, and various other therapies chosen according to cancer-risk assessment [3]. High-risk patients with NMIBC have 60-70% chance of recurrence and 10-45% chance of progression to muscle invasive or metastatic disease within 5 years [4]. Unfortunately, the recurrence rate of BLCA is quite high. Treatment needs to be repeated frequently, which in turn inevitably leads to resistance [5].

Important players in RNA-mediated, post-transcriptional regulation are RNA-binding proteins
Prognostic relevance of RBPs in BLCA

(RBPs) [6]. These proteins, among other diverse biological functions, facilitate regulation by miRNAs and sRNAs [7, 8]. To date, more than 1500 RBP genes have been identified in the human genome through genome-wide analysis [9]. Over the past decade, many studies have revealed abnormal expression of RBPs in tumors, suggesting their involvement in carcinogenesis. IGF2BP1 causes an increase in proliferation and tumorigenesis and the leukemia cell line with low expression of IGF2BP1 has less ability to form colonies and initiate tumors [10]. MSI1 is reported to be a potential therapeutic target for glioblastoma, since luteolin has been shown to inhibit the RNA-binding characteristics of MSI1 and destroy the cancer phenotype in glioblastoma [11]. The hnRNP K has both oncogenic and tumor suppressor properties. However, it mostly behaves as a tumor suppressor in acute leukemia [12]. RBPs constitute a key factor of the post-transcriptional process and play an important role in the regulation of RNA in gastrointestinal [13] and colorectal cancers [14]. Despite the emergence of RBPs as key regulators of every cancer hallmark, very little is known about their potential mechanisms and downstream carcinogenic targets, particularly with regard to bladder cancer. Therefore, all relevant BLCA data were downloaded from TCGA and a comprehensive analysis was conducted to investigate the potential molecular function and clinical significance of RBPs in BLCA. In this study, we selected a number of DERBPs related to BLCA, which have provided new insights into the pathogenesis of the disease, some of which may be potential biomarkers for the diagnosis and prognosis of BLCA.

Material and methods

Data preprocessing and identification of differentially expressed RBPs

The RNA-sequencing dataset and corresponding clinical data were downloaded from TCGA (https://portal.gdc.cancer.gov/); it included 19 normal bladder tissue samples and 411 BLCA samples. The raw data of BLCA were preprocessed using the limma package [15] in R. We used the Wilcoxon test in R to select DERBPs between normal bladder and BLCA tissues, considering |log₂FC (fold change)| ≥ 0.5 and FDR (false discovery rate) < 0.05. Finally, we applied R and the pheatmap R package to draw a volcano map and heatmap of the DERBPs.

Prognosis-related RBP identification

Univariate Cox regression analysis or the Kaplan-Meier test was performed for DERBPs using survival R package. Values with P < 0.01 were considered to correspond to prognosis-related candidate hub RBPs in the univariate Cox regression test. The Kaplan-Meier test was used to evaluate the prognostic value of DERBPs, and a p-value < 0.05 was considered to indicate candidate hub RBPs related to prognosis. Thereafter, a multivariate Cox regression test was applied to prognosis-related candidate hub RBPs in order to further identify the prognosis-related hub RBPs.

Prognostic model establishment and evaluation

All patients with BLCA, from TCGA, were randomly divided into a training and testing group. A multivariate Cox proportional hazards regression model was established, based on prognosis-related RBPs in the training group, which calculated the risk score to evaluate patient prognosis using the survival, caret, glmnet [16], survminer, and survivalROC packages in R. Using the model, we calculated the risk score of each patient with BLCA based on the following formula: Risk score = β1 × Exp1 + β2 × Exp2 + βi × Expi, where β is regression coefficient and Exp is expression level.

On the basis of median risk score from the formula, the training group was divided into a low- and high-risk cohort; thereafter, the testing group was also divided into a low- and high-risk cohort, depending on the median score of training group and risk score from the formula. Patients in the testing group served as a validation cohort to verify the predictive ability of the model. Difference in overall survival rate between high- and low-risk cohorts was compared by log-rank test using survival and survminer R packages in the training and testing group, respectively. The ROC curve was constructed using the survival ROC R package to evaluate the predictive ability of the model in both training and testing groups, and the pheatmap R package was used to draw the risk plot and heatmap. Finally, based on the nine hub RBPs, a nomogram was constructed to predict
the possibility of OS using the rms R package. C-indices were used to estimate the predictive performance of the nomogram.

**Mutation analysis and prognostic value of clinical parameters**

Mutation analysis of nine hub RBPs was executed using the cBioPortal platform (http://www.cbioportal.org) [17]. We applied the survival R package for Cox regression analysis to assess the prognostic significance of different clinical parameters in the training and testing groups of patients with BLCA, respectively.

**Hub RBP expression levels and validation**

We also analyzed the hub RBP’s expression level of TCGA-BLCA using the GEPIA online tool (http://gepia.cancer-pku.cn/index.html) [18]. The online database Human Protein Atlas (http://www.proteinatlas.org/) was utilized to explore the expression of hub RBPs at a translational level [19].

**Results**

**Differentially expressed RBP identification**

The research design is shown in Figure 1A. In this study, we performed a comprehensive analysis of crucial functions and prognostic significance of RBPs in BLCA. Data regarding BLCA were acquired from TCGA, including 411 bladder cancer samples and 19 normal bladder samples. Relevant packages in R were utilized to process the data and select the DERBPs. A total of 388 (out of 1542) RBPs [9] fulfilled the screening criteria of the study, consisting of 219 upregulated and 169 downregulated RBPs. The heatmap and volcano map of DERBPs are displayed in Figure 1B, 1C.

**Prognosis-related RBP screening**

A total of 388 DERBPs were identified. In order to study the prognostic value of these RBPs, univariate Cox regression analysis was performed, and 19 candidate hub RBPs related to prognosis were obtained (Figure 2A). Multivariate Cox regression analysis was performed on the 19 RBPs, of which 9 hub RBPs were identified as independent predictors of BLCA (Figure 2B; Table 1). Among them, four RBPs (FTO, IPO7, YTHDC1, and ZNF106) were downregulated, and five RBPs (CTU1, DARS2, EFTUD2, TRIM71, and XPOT) were upregulated.

**Prognosis-related model construction and analysis**

A total of 404 patients with BLCA were randomly divided into a training (202 patients) and testing group (202 patients). The 9 prognosis-related hub RBPs were utilized to establish a predictive model based on training-group data. We calculated the risk score of every patient based on the following formula: Risk score = (0.2707 × ExpTRIM71) + (-0.1148 × ExpYTHDC1) + (0.0417 × DARS2) + (0.0272 × ExpXPOT) + (0.1341 × ExpZNF106) + (0.2806 × ExpFTO) + (-0.023 × ExpIPO7) + (0.0521 × ExpEFTUD2) + (-0.0812 × ExpCTU1).

Next, we aimed to evaluate the predictive ability. Results in the training group indicated that patients in the high-risk cohort had a worse OS than those in the low-risk cohort (Figure 3B). ROC analysis demonstrated the prognostic value of the nine hub RBPs. Area under the ROC curve (AUC) of the model was 0.752 in the training group (Figure 3C), suggesting it had better diagnostic capability. For the training group, Figure 3A shows the expression heatmap, patient survival status, and risk scores for the low- and high-risk cohorts based on nine RBPs. In order to evaluate whether the risk score model had the same prognostic significance in the testing group, the same formula was used in the latter; high-risk cohort patients were found to have worse OS than those in the low-risk cohort, and area under the ROC curve was 0.701 (Figure 4A-C). It thus suggested better sensitivity and specificity of the model for predicting prognosis.

**A nomogram based on nine RBPs**

In order to develop a quantitative approach for predicting prognosis in bladder cancer, nine RBPs were integrated to construct a nomogram (Figure 5A). The C-indices of the nomogram were 0.7033 in the training group, and 0.6295 in the testing group (validation cohort) (Figure 5B). Based on multivariate Cox regression analysis, the point scale in the nomogram was used to assign values to individual variables. By drawing a vertical line between the prognosis axis and total-point axis, we could calculate the
Prognostic relevance of RBPs in BLCA

A

BLCA clinical information in TCGA

BLCA RNA sequencing data in TCGA

Differently expressed RNA binding proteins (RBPs) between BLCA and normal bladder tissue

Univariate cox regression analysis to select significant candidate hub RBPs

Validation of hub RBPs expression

Multivariate cox regression analysis to identify prognosis related hub RBPs

Multivariate cox proportional hazards regression model

Mutation analysis

Validation cohort TCGA BLCA (Testing group)

ROC

Nomogram

Log-rank test

KM-plotter

Diagnostic value analysis

Clinical decision-making analysis

Prognostic risk assessment

Validation the prognostic value

B

C

Volcano
estimated overall survival rate of 1, 3, and 5 years, which could eventually help doctors make clinical decisions for patients with BLCA.

*Mutation analysis and prognostic value of clinical parameters*

Mutation analysis of the hub genes TRIM71, YTHDC1, DARS2, XPOT, ZNF106, FTO, IPO7, EFTUD2, and CTU1 was performed using the cBioPortal platform. Results indicated that in 226 samples from 404 patients with BLCA, the 9 hub RBPs had changed (56%) (Figure 6B). The high mRNA levels of DARS2 was the maximum alteration among the 9 hub RBPs. Cox regression analysis was used to evaluate the effect of different clinical characteristics on the prognosis of patients with BLCA. Univariate Cox regression analysis results suggested age, stage, and risk score to be related to the OS of patients with BLCA, in both training and testing groups (Figure 6C, 6E). Multivariate Cox regression analysis results indicated age, stage, and risk score to be independent prognostic factors associated with OS in the training and testing groups (Figure 6D, 6F).

*Hub RBP expression levels and validation*

We analyzed the hub RBP expression levels in TCGA-BLCA using the GEPIA online tool, and the result indicated that CTU1, DARS2, EFTUD2, TRIM71, and XPOT expression levels in BLCA tissue were significantly higher than those in
normal bladder tissue. Whereas, FTO, IPO7, YTHDC1, and ZNF106 expression levels in BLCA tissue were significantly lower than those in normal bladder tissue (Figure 7). We used the immunohistochemical results of Human Protein Atlas database to explore the expression of hub RBPs in BLCA, and found that CTU1, DARS2, EFTUD2, TRIM71, and XPOT levels in bladder cancer tissues were significantly higher than in normal bladder tissues. However, the antibody staining levels of FTO, IPO7, YTHDC1, and ZNF106 in bladder cancer tissues were relatively reduced (Figure 8).

Discussion

Although early diagnosis and multimodal treatment of bladder cancer have recently achieved promising results, metastatic diseases are usually incurable, and the 5-year survival rate remains only 15% [20]. Metastasis and recurrence are the main causes of death in patients with bladder cancer, especially MIBC [21]. Therefore, it would be highly significant to understand the molecular mechanism of bladder cancer further, and develop effective early-screening and diagnostic approaches to
Prognostic relevance of RBPs in BLCA

enhance treatment effectiveness and quality of life in patients. RNA-binding proteins play an important role in the regulation of various RNA processes, including splicing, transport, translation, and degradation of coding and non-coding RNAs [22]. RBPs and RNAs assemble into a dynamic complex, called ribonucleoprotein (RNP), which regulates almost every stage of RNA lifecycle [23]. An important regulatory mechanism of IncRNAs is RNA-binding protein-mediated post-transcriptional regulation [24]. This post-transcriptional regulation is a vital approach of coding and non-coding RNAs, and is mainly promoted by RNA-binding proteins, since they dynamically coordinate the maturation, transport, and stability of all RNA types [9]. Identification of pathogenic gene variation in cancer has always been the subject of in-depth study, and colorectal cancer [25], prostate cancer [26], glioblastoma [27], ovarian cancer [28], and melanoma [29] have been reported to be related to RNA-binding proteins.

Figure 4. Validation of the prognostic signature in the testing group. A. The distribution of risk scores; the distribution of survival time and survival status in the low- and high-risk cohorts; heatmap of the expression of nine prognosis-related RBPs between low- and high-risk cohorts. B. The patients in the high-risk cohort had significantly shorter OS than those in the low-risk cohort. C. The ROC curve of model for forecasting OS based on risk score.
Prognostic relevance of RBPs in BLCA

Figure 5. Nomogram construction and assessment. A. Nomogram model for predicting the probability of 1-, 3-, and 5-year OS in BLCA patients. B. The C-index of the nomogram in two cohorts.

However, understanding of the mechanisms of RBPs in BLCA is currently very limited.

In our study, RNA sequencing data of BLCA were integrated to identify the DERBPs between bladder cancer tissues and normal bladder tissues. Univariate Cox regression analysis was used to screen candidate hub RBPs related to prognosis, and multivariate Cox regression analysis was used to identify hub RBPs related to prognosis; finally, we identified the following nine hub RBPs: TRIM71, YTHDC1, DARS2, XPOT, ZNF106, FTO, IPO7, EFTUD2, and CTU1. Using multivariate Cox regression analysis, according to the data from the training group, the risk score model was constructed with the 9 RBPs to predict the prognosis of patients with BLCA. In the training group, the ROC curve of the nine-RBP risk score model had a moderate ability to predict OS (AUC = 0.752), and high-risk patients with BLCA showed remarkably worse overall survival time. The C-indices of the nomogram were 0.7033 in the training group, and 0.6295 in the testing group. The nomogram was established to enable professionals to predict 1-, 3-, and 5-year OS for patients with BLCA. Based on the predicted results by the risk score model, high-risk score patients had worse prognosis, suggesting that the treatment plan and individualized treatment would possibly require adjustment. We further demonstrated that CTU1, DARS2, EFTUD2, TRIM71, and XPOT expression levels in BLCA tissue were significantly higher than those in normal bladder tissue. Whereas, FTO, IPO7, YTHDC1, and ZNF106 expression levels in BLCA tissue were significantly lower than those in normal bladder tissue. Moreover, CTU1, DARS2, EFTUD2, TRIM71, and XPOT expression...
Prognostic relevance of RBPs in BLCA

A

B

C

D

E

F

Prognostic relevance of RBPs in BLCA

Figure 6. Mutation analysis of nine RBPs, and the prognostic effect of different clinical parameters. (A) Mutation frequency of hub RBP genes. (B) Mutation frequency of each RBP gene. Age, tumor stage, and risk score were correlated with OS of BLCA patients by univariate analysis in the training (C) and testing (E) group. Age, tumor stage, and risk score were the independent prognostic indicators by multivariate analysis in the training (D) and testing (F) group.

Figure 7. Expression level of nine RBPs in TCGA-BLCA using the GEPIA online tool. CTU1, DARS2, EFTUD2, TRIM71, and XPOT expression levels in BLCA tissue were significantly higher than those in normal bladder tissue. In contrast, expression levels of FTO, IPO7, YTHDC1, and ZNF106 were the opposite.
was significantly higher in bladder cancer tissues than in normal bladder tissues using Human Protein Atlas database. However, the staining levels of FTO, IPO7, YTHDC1, and ZNF106 in bladder cancer tissues were relatively lower.

The hub RBPs have been reported in many studies. ELP3 and CTU1/2, partner enzymes in uridine 34 (U34) mcm3^s^ -tRNA modification, are upregulated and promote metastasis in human breast cancers [30]. CTU1 copy number amplifications were identified in 25% of myxopapillary ependymomas [31]. Qin et al. demonstrated DARS2 as a hepatocarcinoma gene that could promote the progression of the hepatocarcinoma cell cycle and inhibit the apoptosis of hepatocarcinoma cells [32]. EFTUD2 gene expression was upregulated in hepatocarcinoma, and had prognostic significance in patients with hepatocarcinoma [33]. Liu et al. found the expression of FTO in esophageal squamous cell carcinoma (ESCC) to be higher than in adjacent normal tissues, and the corresponding survival curve showed the high expression of FTO to tend toward poor progno-
sis. In terms of function, FTO silencing inhibited the growth and migration of ESCC cells in CCK8 and Transwell assays, whereas FTO overexpression showed an opposite result [34]. Inhibition of IPO7 by siRNA is known to lead to reduced proliferation of prostate cancer cells [35]. Torres-Fernández et al. had indicated TRIM71 to be correlated with advanced stages and poor prognosis in hepatocellular carcinoma. TRIM71 could inhibit the mRNA expression of cell cycle inhibitor and tumor suppressor CDKN1A/p21, and promote the proliferation of tumor cells [36]. XPOT belongs to the Ran-GTPase exportin family that mediates export of tRNA from the nucleus to the cytoplasm, and high expression of XPOT in hepatocellular carcinoma is associated with worse prognosis [37]. Celona et al. had reported ZFP106 knockout mice to have severe degeneration of motoneurons while transgenic recovery of ZFP106 specifically inhibited the degeneration [38].

In summary, the study proposed new insights regarding the functions of RBPs in BLCA oncogenesis and development. In addition, the model indicated better predictive ability in terms of survival, which may be helpful in the exploitation of novel BLCA prognostic biomarkers. However, this research had some limitations. Firstly, our findings are only based on RNA sequencing without other omics data. Secondly, the risk score model was established based on the TCGA BLCA data, and prospective studies should be conducted to prove it. Thirdly, the TCGA data lacked some clinical characteristics that may have reduced the statistical validity and reliability of multivariate Cox regression analysis. Finally, since we had adopted a bioinformatics approach, further biological experiments are required to verify the claims.

**Conclusion**

In conclusion, we comprehensively investigated the function and prognostic significance of DERBP s in BLCA through extensive bioinformatics analysis. The hub RBPs may be involved in oncogenesis, development, and metastasis of BLCA. A risk score model, or RBP-related prognostic model, was established, and might be used as an independent prognostic factor for BLCA. Our results will be of great significance in revealing the pathogenesis of BLCA, and developing new therapeutic targets and prognostic molecular markers for BLCA.

**Acknowledgements**

We thank the TCGA database and authors for the availability of the data. This study is supported by the National Natural Science Foundation of China (81960208), and the Natural Science Foundation of Inner Mongolia Autonomous Region (2019MS08046).

**Disclosure of conflict of interest**

None.

**Address correspondence to:** Guochang Bao, Department of Urology, Affiliated Hospital of Chifeng University, and Urology Research Center, Chifeng University, No. 42, Wangfu Street, Songshan District, Chifeng 024000, China. Tel: +86-1394769-0228; Fax: +86-476-5973310; E-mail: guochangbao58@163.com

**References**


Prognostic relevance of RBPs in BLCA


Prognostic relevance of RBPs in BLCA

[34] Liu S, Huang M, Chen Z, Chen J, Chao Q, Yin X and Quan M. FTO promotes cell proliferation and migration in esophageal squamous cell carcinoma through up-regulation of MMP13. Exp Cell Res 2020; 389: 111894.