American Journal of Cancer Science

http://ivyunion.org/index.php/ajcs/

Research Article

# An Integrated Clinical-mRNA-lncRNA-miRNA Signature for Muscle-Invasive Bladder Cancer Prognosis

Xing Liu[1], Min Zhou[2], Xi Luo[1], Yiren Wang[1], Fang Deng[1], Xiaoning Peng[3,4*], and Xiaomin Zeng[1,5*]

[1] Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Changsha, China

[2] Department of Vascular Surgery, Zhongshan Hospital, Fudan University, Shanghai, China

[3] Department of Pathology and Pathophysiology, Hunan Normal University School of Medicine, Changsha, China

[4] Department of Pathophysiology, Jishou University School of Medicine, Jishou, China

[5] Hunan Provincial Key Laboratory of Clinical Epidemiology (Central South University), Changsha, China

Abstract: An increasing number of evidence suggests that clinical variables alone are not enough to predict the survival of patients with muscle invasive bladder cancer (MIBC), and the expression of mRNAs, long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) also plays an important role in the onset of MIBC. This study aims to establish a more accurate model for predicting the overall survival of MIBC based on clinical information and genetic characteristics. In this study, the RNAs profiles and clinical variable data of patients with MIBC were downloaded from the Cancer Genome Atlas (TCGA) database. Univariate Cox regression analysis, differential expression analysis and elastic net-regulated Cox regression analysis were used to identify the clinical variables and RNAs (mRNAs, lncRNAs and miRNAs) related to the prognosis of MIBC. Prognostic models of MIBC were established by multivariate Cox regression and ridge regression analysis using the identified prognostic clinical variables and RNAs. Three clinical variables, 25 mRNAs, 3 lncRNAs and 2 miRNAs related to the prognosis of MIBC were identified, and an integrated signature, a clinical variable signature, and an mRNA-lncRNA-miRNA signature were established based on the identified clinical variables and/or RNAs. Among the three models, the integrated signature had the highest predictive accuracy (5-year the area under the curve (AUC)=0.835, 95%CI: 0.776-0.894) among the three models ($P < 0.05$). The patients in the TCGA MIBC cohort were classified into high- or low-risk groups by the integrated signature, and it was found that the patients in the low-risk group had a significantly longer overall survival time compared with the patients in the high-risk group ($P < 0.001$). Applying published gene signatures and TCGA data, a new and more accurate integrated clinical-mRNA-lncRNA-miRNA signature for MIBC prognostic was established.

**Keywords**: Bladder cancer; RNA; Prognosis; Signature; The Cancer Genome Atlas

**Received**: January 10, 2020; **Accepted**: February 12, 2020; **Published**: February 26, 2020

**Competing Interests**: The authors have declared that no competing interests exist.



***Correspondence to:** Xiaomin Zeng (zxiaomin@csu.edu.cn); Xiaoning Peng (pxiaoning@hunnu.edu.cn)

# 1. Introduction

Bladder cancer (BC), a kind of malignant tumor that develops on the bladder mucosa, ranks first among urogenital system tumors in China and next only to prostate cancer in western countries [1, 2]. Previous studies show that 15%-30% of BC are muscle invasive bladder cancer (MIBC) with high degree of malignancy, and are usually treated with radical cystectomy plus pelvic lymph node dissection and systemic chemotherapy thereafter [3]. However, the postoperative 5-year survival rate was only 66% [4]. Due to high recurrence and metastasis rate of MIBC, it is important to carry out studies onthe improvement of prognosis. Currently, the prognosis of patients with MIBC is predicted mainly based on tumor size, multiple lesions, age and other clinical variables [5-7]. Recently, several studies showed that biological alterations in specific genes or molecules could affect the prognosis of MIBC patients [8-10]. Thus, it is necessary to investigate the prognosis-related clinical variables and biomarkers of MIBC, to obtain a prognostic model with high prediction accuracy and assist in the development of rational individualized treatment regimen.

With the development of high throughput sequencing technique, gene expression of mRNAs and non-coding RNAs (ncRNA) were widely used in the prognosis of patients with MIBC [8, 9, 11]. Liu et al. [12] established a prognostic model integrated with TNM stage and six selected PCG-lncRNA-miRNA signature of patients with BC, which was superior to the prognostic model of TNM stage signature and the prognostic model of PCG-lncRNA-miRNA signature in predictive ability, but only one clinical variable (TNM stage) was presented in this model. Xiong et al. [7] also found that a prognostic model integrated with clinical-mRNA-miRNA signature was superior to the clinical-alone signature in predicting the overall survival (OS) of BC, but lncRNAs signature was missing in this model. According to our knowledge, there was no integrated prognostic model of clinical variables and mRNAs, lncRNAs and miRNAs signature for MIBC. Thus, we obtained the clinical variable data and mRNAs, miRNAs and lncRNAs expression profiles of patients with MIBC from The Cancer Genome Atlas (TCGA) database. We aimed to construct a more accurate prognosis model and provide new genomics clues for the formation and metastasis study of BC.

# 2. Materials and methods

## 2.1 Patient information and RNA profiles

Demographic information (genders, races, smoking history), clinical data (diagnostic age, clinical stage of tumor, neoplasm histologic grade, TNM staging, diagnostic subtypes and OS and RNA profiles of cancer tissue and paracancerous tissue of MIBC patients (mRNAs, lncRNAs andmiRNAs profiles at Level 3) were downloaded from the TCGA database (https://portal.gdc.cancer.gov/). The exclusion criteria were survival time less than 30 days after initial pathologic diagnosis. The qualified MIBC patients were included in the TCGA MIBC cohort for this study. Survival time of patients was defined as a time span between the initial diagnosis and all-cause death.

## 2.2 Data preprocessing

The data obtained from the TCGA database were pre-processed as follows: (1) for each of the clinical variables, when the data missing rate was less than 10%, the missing data was filled by the Classification and Regression Trees method; (2) the RNA sequence data was generated by Illumina

HiSeq RNA-Seq platform, and the RNA sequence was annotated (annotation file: GENCODE.V29) to obtain mRNA and lncRNA profiles, then the mRNA and lncRNA profiles was standardized by the Transcripts Per Million Reads method; and the miRNA sequence data was generated by Illumina HiSeq miRNA-Seq platform; (3) the expression of each RNA gene in patients was divided into low expression and high expression by using the median of expression of each RNA gene as the cut-off point for univariate Cox regression analysis.

## 2.3 Identification of prognostic variables

The clinical variables were filtered out by univariate Cox regression at first, then the clinical variables with *P* value ≤ 0.05 were enrolled in a multivariate Cox regression. Finally, the clinical variables with *P* value ≤ 0.05 in multivariate Cox regression were identified as the significant factors of the prognosis of MIBC.

mRNA, lncRNA and miRNA related to the prognosis of MIBC were identified by the following three steps: (1) the RNAs with |log2FC| > 2 (Fold Change, FC) and false discovery rate (FDR) < 0.01 were preliminarily screened as the differentially expressed genes. (2) for the differentially expressed genes, the RNAs with a univariate Cox regression *P* value < 0.01 were taken as the significant genes of the MIBC prognosis. (3) we fitted an elastic net-regulated Cox regression with 10,000 iterations and 10 cross-validations using the significant genes of the MIBC prognosis with a univariate Cox regression *P* value < 0.01. RNAs with elastic net-regulated Cox regression coefficient≠0 were taken as the identified candidate genes related to the prognosis of MIBC.

## 2.4 Prognosis index and model development

For each patient of MIBC, a prognosis index (PI) was calculated as an integrated indicator of the identified candidate clinical variables and (or) RNAs. Weighted prognostic index (WPI) was defined as the standard form of the PI. Specifically,

$$PI = \sum_i (\beta_i \times V_i)$$

$$WPI = \frac{PI - Mean(PI)}{SD(PI)}$$

where, PI is the prognostic index of the each observation. $\beta_i$ is the regression coefficient of the *i*th variable. Mean (PI) and SD (PI) are the mean and standard deviation of the PI vector, respectively. For the clinical variable signature, $V_i$ is the observed value of the *i*th clinical variable, and $\beta_i$ is the multivariate Cox regression coefficient of the *i*th clinical variable. For the RNA signature, $V_i$ is the expression value of each RNA, and $\beta_i$ is the ridge regression coefficient of the *i*th RNA. For the integrated clinical-RNA signature, $V_i$ is the expression value of each RNA or observed value of clinical variable, and $\beta_i$ is the ridge regression coefficient of the *i*th RNA or clinical variable.

## 2.5 Model evaluation

The area under the curve (AUC) of the time-dependent receiver operating characteristic (ROC) of the three prognostic models were separately calculated. The efficiency of the three models to predict the prognosis of MIBC was evaluated by comparing the AUC values. The model with the maximum AUC value was defined as the best prognosis model, and its PI and WPI were used as $PI_{\text{best}}$ and $WPI_{\text{best}}$. The mean of the $WPI_{\text{best}}$ was taken as the cut-off point ($WPI_{\text{cut}}$) if the distribution characteristics of

the $WPI_{best}$ values was symmetrical, or the median of the $WPI_{best}$ was taken as the $WPI_{cut}$ if the distribution characteristics of the $WPI_{best}$ values was skewed. According to the $WPI_{cut}$, the MIBC patients were divided into two subgroups: a low-risk group ($WPI_{best} \leq WPI_{cut}$) and a high-risk group ($WPI_{best} > WPI_{cut}$). We compared the OS time of the 2 groups using the Kaplan-Meier survival curves and log-rank test.

All the above analyses were performed using RStudio 1.1.463.

### 2.6 Gene function enrichment analysis

Gene ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment were carried out based on validated genes using WebGestalt 2019 [13] (http://www.WebGestalt.org). The validated genes for GO and KEGG enrichment included as follows: (1) the identified candidate mRNAs in this study; (2) the target mRNAs of the identified candidate miRNAs in this study obtained from miRDB database; (3) the mRNAs obtained from TCGA database in this study were correlated with the identified candidate lncRNAs.

## 3. Results

### 3.1 Patient information and RNA profiles

A total of 409 patients with MIBC were reviewed, and 19 of them were excluded due to the following reasons: missing survival time (n=2), incomplete mRNA, lncRNA and miRNA profiles (n=7), and death within 30 days of the initial pathologic diagnosis (n=10). Finally, 390 patients were included in the TCGA MIBC cohort for further study. Of this cohort, 19 patients had RNAs profiles of paracancerous tissues.

The demographic and clinical characteristics of these patients were shown in Table 1. The RNA profiles of this cohort included 19,364 mRNA, 5,217 lncRNA and 1,882 miRNA profiles.

**Table 1** Survival analysis results of demographic variables and clinical variables.

| Variables | Patients | Univariate Cox | | Multivariate Cox | |
|---|---|---|---|---|---|
| | n(%) | HR(95%CI) | *P* value | HR(95%CI) | *P* value |
| Diagnostic Age | 390(67.83)* | 1.03(1.02-1.05) | < 0.001 | 1.03(1.02-1.05) | < 0.001 |
| Clinical Stage | | 1.71(1.40-2.08) | < 0.001 | 1.47(1.20-1.79) | < 0.001 |
| Ⅰ | 0（0.00） | | | | |
| Ⅱ | 123(31.54) | | | | |
| Ⅲ | 136(34.87) | | | | |
| Ⅳ | 129(33.08) | | | | |
| Missed | 2(0.51) | | | | |
| TNM-N stage | | 1.24(1.12-1.37) | < 0.001 | 1.21(1.07-1.36) | 0.003 |
| N0 | 226(57.95) | | | | |
| N1 | 43(11.03) | | | | |
| N2 | 75(19.23) | | | | |
| N3 | 7(1.79) | | | | |

| | | | |
|---|---|---|---|
| NX | 35(8.97) | | |
| Missed | 4(1.03) | | |
| TNM-T stage | | 1.22(1.01-1.47) | 0.042 |
| T1 | 8(2.05) | | |
| T2 | 185(47.44) | | |
| T3 | 150(38.46) | | |
| T4 | 41(10.51) | | |
| TX | 6(1.54) | | |
| TNM-M stage | | 1.20(1.03-1.40) | 0.018 |
| M0 | 186(47.69) | | |
| M1 | 10(2.57) | | |
| MX | 192(49.23) | | |
| Missed | 2(0.51) | | |
| Diagnostic Subtype | | 1.55(1.08-2.21) | 0.017 |
| Papillary | 124(31.80) | | |
| Non-papillary | 261(66.92) | | |
| Missed | 5(1.28) | | |
| Gender | | 1.20(0.86-1.66) | 0.277 |
| Female | 103(26.41) | | |
| Male | 287(73.59) | | |
| Human Race[a] | | | |
| White | 312(80.00) | 1.05(0.67-1.65) | 0.818 |
| Asian | 39(10.00) | 0.65(0.33-1.28) | 0.215 |
| Black[#] | 22(5.64) | 1.37(0.78-2.42) | 0.271 |
| Missed | 17(4.36) | | |
| Neoplasm Histologic Grade | | 2.76(0.68-11.16) | 0.154 |
| Low | 18(4.62) | | |
| High | 369(94.62) | | |
| Missed | 3(0.76) | | |
| Smoking History | | 1.35(0.94-1.92) | 0.101 |
| Non-smoking | 105(26.93) | | |
| Smoking | 272(69.74) | | |
| Missed | 13(3.33) | | |

[a]: dumb variable. [*]: n (mean, year). [#]: black or African American.

## 3.2 Prognostic clinical variables for MIBC

The following clinical variables were significantly associated with the prognosis of MIBC: diagnostic age (HR=1.03, 95%CI: 1.02-1.05), clinical stage (HR=1.47, 95%CI: 1.20-1.79) and TNM-N stage

(HR=1.21, 95%CI: 1.07-1.36). With the increase of age, clinical stage and TNM-N stage in MIBC patients, there was increased risk of all-cause mortality (Table 1).

### 3.3 Prognostic RNAs for MIBC

A total of 3,908 differentially expressed genes between cancer and the paracancer screened including 3,781 (1,994, up-regulated; 1,787, down-regulated) differentially expressed mRNAs, 53 differentially expressed lncRNAs (21, up-regulated; 32, down-regulated) and 74 differentially expressed miRNAs (16, up-regulated; 58, down-regulated) were obtained. The results of the univariate Cox regression showed that there were 107 RNAs (78 mRNAs, 14 lncRNAs and 15 miRNAs) were identified as potential predictors of OS in patients with MIBC ($P < 0.01$). For further, the 107 RNAs were screened by fitting elastic net-regulated Cox regression. The results revealed that the regression coefficient of 30 selected candidate RNAs (25 mRNAs, 3 lncRNAs and 2 miRNAs) was over 0, suggesting that these 30 candidate RNAs were related to the prognosis of MIBC (Table 2).

**Table 2** Thirty candidate RNAs for the integrated signature.

| Number | Gene symbol | HR[!] | 95%CI | *P* value* | Regulation[#] | Coefficient[§] |
|---|---|---|---|---|---|---|
| mRNA | | | | | | |
| 1 | *LSM7* | 0.45 | 0.32-0.64 | < 0.001 | Down | -0.0817 |
| 2 | *TMEM259* | 0.47 | 0.33-0.65 | < 0.001 | Down | -0.1928 |
| 3 | *PRRG2* | 0.47 | 0.33-0.67 | < 0.001 | Down | -0.1609 |
| 4 | *CCDC61* | 0.47 | 0.33-0.67 | < 0.001 | Down | -0.1080 |
| 5 | *OCIAD2* | 0.47 | 0.33-0.68 | < 0.001 | Down | -0.0132 |
| 6 | *GEMIN7* | 0.49 | 0.34-0.70 | < 0.001 | Down | -0.0336 |
| 7 | *SPINT2* | 0.50 | 0.35-0.74 | < 0.001 | Down | -0.0613 |
| 8 | *ZNF600* | 0.51 | 0.35-0.73 | < 0.001 | Down | -0.0553 |
| 9 | *C19orf25* | 0.52 | 0.37-0.72 | < 0.001 | Down | -0.2209 |
| 10 | *KLHDC4* | 0.54 | 0.38-0.76 | < 0.001 | Down | -0.0579 |
| 11 | *TMC6* | 0.54 | 0.37-0.78 | 0.001 | Down | -0.0125 |
| 12 | *RBPMS* | 0.57 | 0.40-0.80 | 0.001 | Up | -0.0649 |
| 13 | *PID1* | 1.77 | 1.30-2.40 | < 0.001 | Up | 0.0556 |
| 14 | *DIXDC1* | 1.80 | 1.30-2.49 | < 0.001 | Up | 0.0015 |
| 15 | *OGN* | 1.81 | 1.34-2.45 | < 0.001 | Up | 0.0057 |
| 16 | *NHSL2* | 1.81 | 1.32-2.49 | < 0.001 | Up | 0.0843 |
| 17 | *SERPINB12* | 1.81 | 1.31-2.49 | < 0.001 | Up | 0.2234 |
| 18 | *SETBP1* | 1.83 | 1.35-2.49 | < 0.001 | Up | 0.0908 |
| 19 | *TMOD1* | 1.83 | 1.35-2.49 | < 0.001 | Up | 0.1079 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 20 | *CASQ2* | 1.88 | 1.39-2.54 | < 0.001 | Up | 0.0391 |
| 21 | *PTGIS* | 1.91 | 1.35-2.70 | < 0.001 | Up | 0.0568 |
| 22 | *ITGA7* | 1.91 | 1.29-2.82 | 0.001 | Up | 0.1859 |
| 23 | *SCN4B* | 1.95 | 1.44-2.64 | < 0.001 | Up | 0.3103 |
| 24 | *MAP1B* | 2.08 | 1.42-3.04 | < 0.001 | Up | 0.1689 |
| 25 | *EPHB1* | 2.15 | 1.55-2.98 | < 0.001 | Up | 0.1686 |
| lncRNA | | | | | | |
| 26 | *AC097534.2* | 0.71 | 0.53-0.96 | 0.002 | Up | -0.3506 |
| 27 | *AC021016.2* | 0.72 | 0.53-0.99 | 0.004 | Up | -0.1956 |
| 28 | *LINC01184* | 1.78 | 1.10-2.87 | 0.001 | Up | 0.5666 |
| miRNA | | | | | | |
| 29 | *hsa-mir-651* | 0.62 | 0.46-0.85 | 0.003 | Down | -0.0220 |
| 30 | *hsa-mir-1976* | 0.50 | 0.34-0.72 | < 0.001 | Down | -0.0139 |

Abbreviations: HR, hazard ratio; CI, confidence interval.

*Univariate Cox regression *P* value < 0.01 was considered statistically significant.

# type of regulation (upregulated or downregulated) in Bladder cancer patients vs normal controls.

§Elastic net-regulated Cox regression coefficient.

‖Protective RNAs had a hazard ratio < 1 and risky RNAs had a hazard ratio > 1 in bladder cancer patients as determined by univariate Cox regression.

**Table 3** AUC values of the three kind of prognostic models.

| Prognostic Model | 1-year[a] | | 3-year[b] | | 5-year[c] | |
|---|---|---|---|---|---|---|
| | AUC | 95%CI | AUC | 95%CI | AUC | 95%CI |
| 1. clinical variable signature | 0.713 | 0.651-0.776 | 0.699 | 0.631-0.769 | 0.707 | 0.622-0.793 |
| 2. mRNA-lncRNA-miRNA signature | 0.706 | 0.639-0.774 | 0.793 | 0.734-0.852 | 0.793 | 0.725-0.861 |
| 3. integrated signature | 0.763 | 0.702-0.823 | 0.827 | 0.772-0.882 | 0.835 | 0.776-0.894 |

[a]: The AUC values of Model 1 and Model 2 were compared with that of Model 3, and P values were 0.145 and < 0.001, respectively.

[b]: The AUC values of Model 1 and Model 2 were compared with that of Model 3, and P values were < 0.001 and 0.021, respectively.

[c]: The AUC values of Model 1 and Model 2 were compared with that of Model 3, and P values were < 0.001 and 0.024, respectively.

## 3.4 Prognostic model for MIBC

Clinical variable signature (Model 1): the value of the PI for Model 1 was calculated by the following formula: $PI=0.033*X_1+0.382*X_2+0.187*X_3$, and $X_1$, $X_2$ and $X_3$ in the formula were the identified clinical variables, which were diagnostic age, clinical stage and TNM-N stage, respectively (Table 1). As to Model 1, the AUC values of ROC at 1-year, 3-year and 5-year were 0.713 (95%CI: 0.651-0.776), 0.699 (95%CI: 0.631-0.769) and 0.707 (95%CI: 0.622-0.793), respectively. This result indicated that Model 1 can accurately predict the prognosis of MIBC (Table 3, Fig. 1).

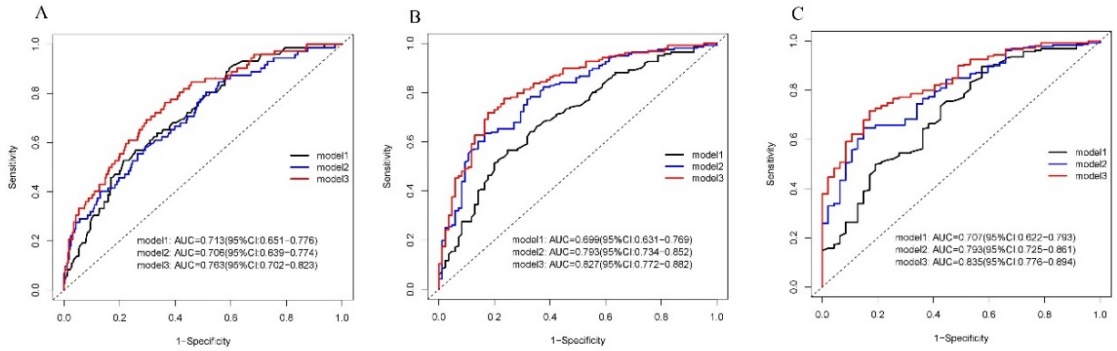


**Fig. 1** ROC curves for the clinical variable signature (Model 1), mRNA-lncRNA-miRNA signature (Model 2) and integrated signature (Model 3). A: 1-year ROC curve; B: 3-year ROC curve; C: 5-year ROC curve.

**Table 4** mRNA-lncRNA-miRNA signature ridge regression model.

| No. | Gene symbol | Coefficient | No. | Gene symbol | Coefficient |
|---|---|---|---|---|---|
| 1 | *LSM7* | -0.11236 | 16 | *NHSL2* | 0.10920 |
| 2 | *TMEM259* | -0.14759 | 17 | *SERPINB12* | 0.18034 |
| 3 | *PRRG2* | -0.14256 | 18 | *SETBP1* | 0.12176 |
| 4 | *CCDC61* | -0.08467 | 19 | *TMOD1* | 0.10711 |
| 5 | *OCIAD2* | -0.06095 | 20 | *CASQ2* | 0.05853 |
| 6 | *GEMIN7* | -0.08146 | 21 | *PTGIS* | 0.10672 |
| 7 | *SPINT2* | -0.11538 | 22 | *ITGA7* | 0.20852 |
| 8 | *ZNF600* | -0.07962 | 23 | *SCN4B* | 0.25273 |
| 9 | *C19orf25* | -0.19050 | 24 | *MAP1B* | 0.16588 |
| 10 | *KLHDC4* | -0.07556 | 25 | *EPHB1* | 0.15126 |
| 11 | *TMC6* | -0.11435 | 26 | *AC097534.2** | -0.34230 |
| 12 | *RBPMS* | -0.13521 | 27 | *AC021016.2** | -0.24680 |
| 13 | *PID1* | 0.09729 | 28 | *LINC01184** | 0.53358 |
| 14 | *DIXDC1* | 0.06041 | 29 | *has-mir-651*$^{\#}$ | -0.06155 |
| 15 | *OGN* | 0.04545 | 30 | *hsa-mir-1976*$^{\#}$ | -0.07795 |

*: lncRNA. $^{\#}$: miRNA

mRNA-lncRNA-miRNA signature (Model 2): Model 2 was established based on the identified 30 RNAs in this study (Table 4). Regarding Model 2, the AUC values of ROC at 1-year, 3-year and 5-year were 0.706 (95%CI: 0.639-0.774), 0.793 (95%CI: 0.734-0.852) and 0.793 (95%CI: 0.725-0.861), respectively. This result suggested that this RNAs signature (Model 2) can accurately predict the prognosis of MIBC (Table 3, Fig. 1).

Integrated signature (Model 3): Model 3 was established based on the identified 3 clinical variables and 30 RNAs in this study (Table 5). For Model 3, the AUC values of ROC at 1-year, 3-year and 5-year were greater than those of the other two models, and the minimum AUC value of ROC was 0.763 (95%CI: 0.702-0.823), which was the AUC value at 1-year.

**Table 5** Integrated signature ridge regression model.

| No. | Symbol | Coefficient | No. | Symbol | Coefficient |
|---|---|---|---|---|---|
| 1 | *LSM7* | -0.09845 | 18 | *SETBP1* | 0.16407 |
| 2 | *TMEM259* | -0.13468 | 19 | *TMOD1* | 0.07921 |
| 3 | *PRRG2* | -0.19806 | 20 | *CASQ2* | 0.04033 |
| 4 | *CCDC61* | -0.07815 | 21 | *PTGIS* | 0.08012 |
| 5 | *OCIAD2* | -0.09788 | 22 | *ITGA7* | 0.12968 |
| 6 | *GEMIN7* | -0.08696 | 23 | *SCN4B* | 0.24782 |
| 7 | *SPINT2* | -0.15221 | 24 | *MAP1B* | 0.15625 |
| 8 | *ZNF600* | -0.05400 | 25 | *EPHB1* | 0.13183 |
| 9 | *C19orf25* | -0.13733 | 26 | *AC097534.2** | -0.31203 |
| 10 | *KLHDC4* | 0.00340 | 27 | *AC021016.2** | -0.19513 |
| 11 | *TMC6* | -0.10502 | 28 | *LINC01184** | 0.50612 |
| 12 | *RBPMS* | -0.13213 | 29 | *has-mir-651*$^{\#}$ | 0.00075 |
| 13 | *PID1* | 0.11719 | 30 | *hsa-mir-1976*$^{\#}$ | -0.05994 |
| 14 | *DIXDC1* | 0.09981 | 31 | Clinical stage | 0.16541 |
| 15 | *OGN* | -0.00609 | 32 | Diagnostic age | 0.01550 |
| 16 | *NHSL2* | 0.13559 | 33 | TNM -N stage | 0.15239 |
| 17 | *SERPINB12* | 0.23608 | | | |

*: lncRNA, $^{\#}$: miRNA.

Among the three models, the maximum AUC value of ROC was 0.835 (95%CI: 0.776-0.894), which was the AUC value of Model 3 at 5-year. These results suggested that the Model 3 was superior to the Model 1 or the Model 2, then Model 3 was defined as the best prognosis model for MIBC (Table 3, Fig. 1).

## 3.5 Evaluation for the best prognostic model

The PI and WPI of the integrated signature (Model 3) were taken as the $PI_{\text{best}}$ and $WPI_{\text{best}}$ of the

best model. The distribution of the $WPI_{\text{best}}$ values of the TCGA MIBC cohort (n=390) was positive skewness, therefore, the median of the $WPI_{\text{best}}$ with a value of -0.046 was used as the $WPI_{\text{cut}}$ for the cohort classification. The patients of the cohort were divided into two subgroups, 195 patients with $WPI_{\text{best}}$ value less than or equal to $WPI_{\text{cut}}$ were classified into low-risk group, and the other 195 patients were classified into low-risk group. The results of the Kaplan-Meier survival curves and the log-rank test ($P < 0.001$) showed that the patients in the low-risk group had better prognosis than high-risk group patients (Fig. 2).

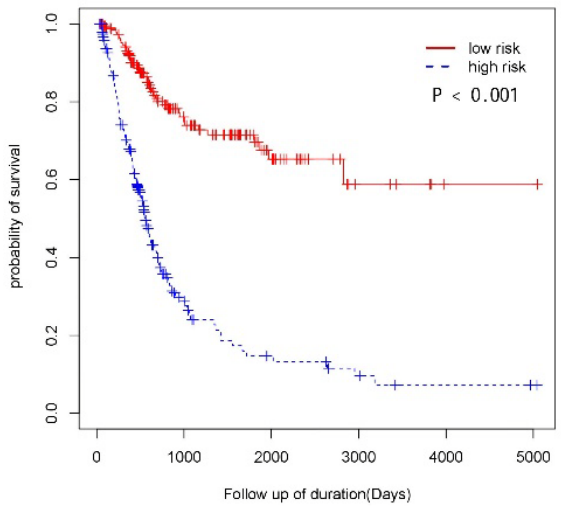


**Fig.2** Kaplan-Meier Survival curves for the low-risk group and the high-risk group segregated by the clinical-mRNA-lncRNA-miRNA signature (Model 3) in the TCGA MIBC cohort.

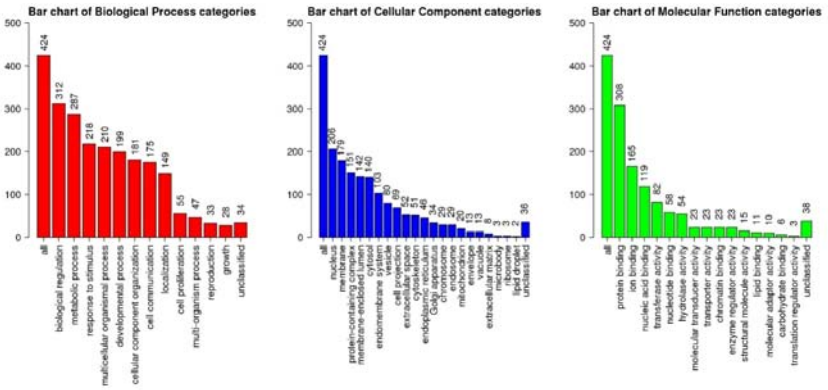


**Fig.3** GO analysis of the 424 validated genes

## 3.6 GO and KEGG enrichment analyses

GO and KEGG pathway enrichment analyses were conducted for the 424 validated genes which were associated with the prognosis of MIBC in this study. The functions of these 424 genes were mainly enriched in the process of biological regulation and metabolism. For cell composition analysis, the products of the 424 genes were located in the nucleus and cell membrane. In addition, most of the 424 genes were related to ion binding, nucleic acid binding and transferase activity (Fig. 3). The function of these 424 genes was enriched in tumor transcriptional disorders, MAPK signaling pathway and AMPK signaling pathway (Fig. 4).

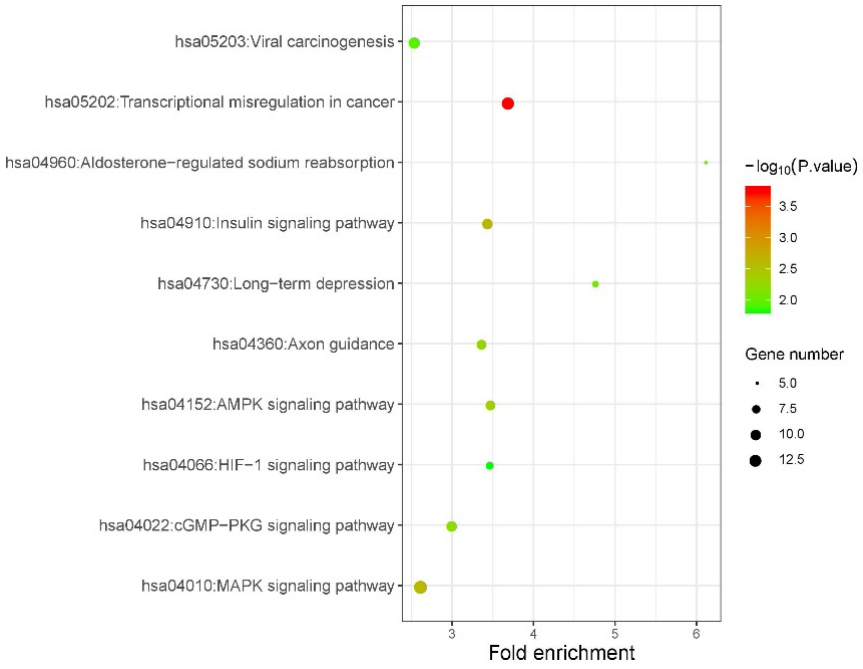


**Fig.4** KEGG analysis of the 424 validated genes

# 4. Discussion

Based on the data of TCGA database, we analyzed the relevance between the clinical variables and mRNA, lncRNA and miRNA expression and the prognosis of patients with MIBC. The identified clinical variables and mRNAs, lncRNAs and miRNAs related to the prognosis of MIBC were used to establish the prognostic model. As the first established prognostic model integrating clinical variables and mRNAs, lncRNAs and miRNAs signature, the integrated signature had the highest predictive accuracy (5-year AUC=0.835, 95%CI: 0.776-0.894) among the three models ($P < 0.05$). The integrated signature was used to classify the prognosis of MIBC patients, the survival rate of patients in the low-risk subgroup was higher than that in the high-risk subgroup ($P < 0.001$), suggesting the good predictive effect of this model.

The integrated signature established in this study included 3 clinical variables and 25 mRNAs, 3 lncRNAs and 2 miRNAs signature related to the prognosis of MIBC. In accordance with previous studies, we also found diagnostic age, clinical stage and TNM-N stage were independent prognostic variables in MIBC patients [5, 6]. Recently, several studies showed that the expression of lncRNAs affected the prognosis of BC [11, 14]. Therefore, our integrated signature added lncRNA signature. The introduction of lncRNA signature improved the predictive ability of the model. Compared with the literature [12], our integrated signature increased the effect of diagnostic age and clinical stage and improved the predictive accuracy of prognosis.

In this study, we identified 25 mRNAs related to the prognosis of MIBC, including 12 protective mRNAs (HR < 1) (*LSM7*, *TMEM259*, *PRRG2*, *CCDC61*, *OCIAD2*, *GEMIN7*, *SPINT2*, *ZNF600*, *C19orf25*, *KLHDC4*, *TMC6*, *RBPMS*) and 13 risk mRNAs (HR > 1) (*PID1*, *DIXDC1*, *OGN*, *NHSL2*, *SERPINB12*, *SETBP1*, *TMOD1*, *CASQ2*, *PTGIS*, *ITGA7*, *SCN4B*, *MAP1B*, *EPHB1*). Of the

prognostic mRNAs, *DIXDC1* was previously reported to a novel component of the Wnt pathway, which played key roles in development, cell growth, differentiation, polarity formation, neural development, and carcinogenesis [15]. Moreover, it has been reported that inhibition of *DIXDC1* by *microRNA-582-5p* and *-3p* suppresses the proliferation and invasion of BC [16]. Furthermore, *PTGIS* is a major regulator in hypoxic cancer progression by activating transcription of various oncogenes [17]. In the competitive endogenous RNA net analysis of BC, *MAP1B* interacted as a node with lncRNA and miRNA to regulate transcription [18]. The interaction between *MAP1B* and *p53* affected the apoptosis and proliferation of cells [19]. Alpha-7 integrin is a protein that is encoded by the *ITGA7*, and appears to be a tumor suppressor that operates by suppressing tumor growth and retarding migration [20]. *ITGA7* were forecast to play important roles in the occurrence and progression of BC [21]. LncRNAs had longer sequences and more complex spatial structures than other types of RNAs and participated in cell proliferation, transformation and other functional activities, and the change of lncRNA expression level had an important impact on the occurrence and development of many kinds of tumors [22-24]. Of the three lncRNAs identified in this study, 2 were protective lncRNAs (*AC097534.2* and *AC021016.2*) (HR < 1) and 1 was risk lncRNA (*LINC01184*) (HR > 1). With oncogenes or anti-oncogenes, miRNAs could affect the proliferation and differentiation of cancer cells [25-27]. The expression of 2 miRNAs (*hsa-mir-651* and *hsa-mir-1976*) identified in this study was down-regulated in patients with MIBC, and the high expression of these 2 miRNAs gene could reduce the prognostic death risk of patients with MIBC (HR < 1), both of which were protective miRNAs (anti-oncogenes). The role and mechanism of the 3 lncRNAs and 2 miRNAs in the occurrence and development of MIBC need to be verified by further experimental study. The results of GO and KEGG analysis of the 424 genes related to the prognosis of MIBC in this study showed that the functions of these genes mainly affected the nucleic acid binding and transferase activity, and was enriched in MAPK and AMPK signaling pathway of tumor. Considering that MAPK signaling pathway played a key role in tumor formation and metastasis [28], it was suggested that RNAs identified in this study might play an important role in the occurrence and development of MIBC.

This study has several potential limitations. First, the predictive accuracy of the prognostic model of integrated signature was needed for further evaluation using other BC data. Second, the function of the identified RNAs needs to be confirmed by experiments.

In summary, a prognostic model of integrated signature for MIBC established in our study has a better predictive effect of prognosis than that of clinical variable signature and mRNA-lncRNA-miRNA signature. The results of our study can provide new genomics clues for the formation and metastasis of MIBC.

# Acknowledgments

The Cancer Genome Atlas project contributed the data for analyses of our study. We appreciate the help of reviewers.

# Conflict of interest

The authors declare that they have no competing interests.

# Funding

Curriculum Construction Project of Innovation and Entrepreneurship Education in Central South University (2018CXKZ04).


# References


1. Flaig TW, Spiess PE, Agarwal N, et al. NCCN Guidelines Insights: Bladder Cancer, Version 5.2018. ***J Natl Compr Canc Netw***. 2018, 16:1041-1053
2. Siegel RL, Miller KD, Jemal A. Cancer statistics. ***CA. Cancer J Clin***. 2018, 68:7-30
3. Gregg JR, Dahm P, Chang SS. Guideline-based management of non-muscle invasive bladder cancer. ***Indian J Urol***. 2015, 31:320-326
4. Shariat SF, Karakiewicz PI, Palapattu GS, et al. Outcomes of radical cystectomy for transitional cell carcinoma of the bladder: a contemporary series from the bladder cancer research consortium. ***J Urol***. 2006, 176:2414-2422; discussion 2422.
5. Leliveld AM, Doornweerd BH, Bastiaannet E, et al. Treatment and outcome in muscle invasive bladder cancer: a population-based survey. ***World J Urol***. 2010, 28:439-444
6. Reheman Y, Zhang L, Wu EW, et al. Analysis of prognostic factors in 102 cases of muscle-invasive bladder cancer (in Chinese). ***Oncology Progress***. 2016, 14:146-149+154
7. Xiong J, Xiong K, Bing Z. Clinical and RNA expression integrated signature for urothelial bladder cancer prognosis. ***Cancer Biomark***. 2018, 21:535-546
8. Andrew AS, Marsit CJ, Schned AR, et al. Expression of tumor suppressive microRNA-34a is associated with a reduced risk of bladder cancer recurrence. ***Int J Cancer***. 2015, 137:1158-1166
9. Oh JJ, Ji SH, Choi DK, et al. A six-week course of bacillus Calmette-Guerin prophylaxis is insufficient to prevent tumor recurrence in nonmuscle invasive bladder cancer with strong-positive expression of p53. ***Oncology***. 2010, 79:440-446
10. Wang HF, Wu JH, Gai JW, et al. MAN1B1 is associated with poor prognosis and modulates proliferation and apoptosis in bladder cancer. ***Gene***. 2018, 679:314-319
11. de Jong JJ, Liu Y, Robertson AG, et al. Long non-coding RNAs identify a subset of luminal muscle-invasive bladder cancer patients with favorable prognosis. ***Genome Med***. 2019, 11:60
12. Liu CP, Zhang JH, Zheng SC, et al. A novel clinical multidimensional transcriptome signature predicts prognosis in bladder cancer. ***Oncol Rep***. 2018, 40:2826-2835
13. Liao Y, Wang J, Jaehnig EJ, et al. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. ***Nucleic Acids Res***. 2019, 47:W199-w205
14. Lian P, Wang Q, Zhao Y, et al. An eight-long non-coding RNA signature as a candidate prognostic biomarker for bladder cancer. ***Aging (Albany NY)***. 2019, 11:6930-6940
15. Wang L, Li H, Chen Q, et al. Wnt signaling stabilizes the DIXDC1 protein through decreased ubiquitin-dependent degradation. ***Cancer Sci***. 2010, 101:700-706
16. Uchino K, Takeshita F, Takahashi RU, et al. Therapeutic effects of microRNA-582-5p and -3p on the inhibition of bladder cancer progression. ***Mol Ther***. 2013, 21:610-619
17. Lu M, Ge Q, Wang G, et al. CIRBP is a novel oncogene in human bladder cancer inducing expression of HIF-1alpha. ***Cell Death Dis***. 2018, 9
18. Zhu N, Hou J, Wu Y, et al. Integrated analysis of a competing endogenous RNA network reveals key lncRNAs as potential prognostic biomarkers for human bladder cancer. ***Medicine (Baltimore)***. 2018, 97:e11887

19. Lee SY, Kim JW, Jeong MH, et al. Microtubule-associated protein 1B light chain (MAP1B-LC1) negatively regulates the activity of tumor suppressor p53 in neuroblastoma cells. ***FEBS Lett***. 2008, 582:2826-2832
20. Ren B, Yu YP, Tseng GC, et al. Analysis of integrin alpha7 mutations in prostate cancer, liver cancer, glioblastoma multiforme, and leiomyosarcoma. ***J Natl Cancer Inst***. 2007, 99:868-880
21. Jia Z, Ai X, Sun F, et al. Identification of new hub genes associated with bladder carcinoma via bioinformatics analysis. ***Tumori***. 2015, 101:117-122
22. Lv M, Zhong Z, Huang M, et al. lncRNA H19 regulates epithelial-mesenchymal transition and metastasis of bladder cancer by miR-29b-3p as competing endogenous RNA. ***Biochim Biophys Acta Mol Cell Res***. 2017, 1864:1887-1899
23. Yang L, Zhang J, Jiang A, et al. Expression profile of long non-coding RNAs is altered in endometrial cancer. ***Int J Clin Exp Med***. 2015, 8:5010-5021
24. Li HJ, Sun XM, Li ZK, et al. LncRNA UCA1 promotes mitochondrial function of bladder cancer via the MiR-195/ARL2 signaling pathway. ***Cell PhysiolBiochem***. 2017, 43:2548-2561
25. Li Q, Wang H, Peng H, et al. MicroRNAs: key players in bladder cancer. ***Mol Diagn Ther***. 2019, 23:579-601
26. Wei XC, Lv ZH. MicroRNA-132 inhibits migration, invasion and epithelial-mesenchymal transition via TGFbeta1/Smad2 signaling pathway in human bladder cancer. ***Onco Targets Ther***. 2019, 12:5937-5945
27. Enokida H, Yoshino H, Matsushita R, et al. The role of microRNAs in bladder cancer. ***Investig Clin Urol***. 2016, 57(Suppl 1):S60-76
28. Sun Y, Liu WZ, Liu T, et al. Signaling pathway of MAPK/ERK in cell proliferation, differentiation, migration, senescence and apoptosis. ***J Recept Signal Transduct Res***. 2015, 35:600-604