The Characteristics of Natural Killer Cells and Prediction of Prognosis in Bladder Urothelial Carcinoma

Yuxin Zhang  
Fujian Medical University

Yankai Zeng  
First Affiliated Hospital of Xiamen University

Shuaishuai Song  
Xiamen University

Zhiyong Zhang  
Xiamen University

Yuedong Chen (chenyuedong8@126.com)  
Fujian Medical University

Article

Keywords: BLCA, NRGs, TCGA, GEO

Posted Date: April 11th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2765540/v1

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

**Background** Bladder urothelial carcinoma (BLCA) is a malignant tumor occurring in the bladder and derived from urothelial cells. It is one of the ten most common cancers in the world. There has been no significant progress in the treatment of patients with recurrence of BLCA in the past decades. With the progress of systemic treatment, the overall survival period (OS) of patients has improved, but the prognosis is still poor, which needs improvement clinically. The development of immunotherapy has changed the status initially. Natural killer (NK) cells, as an important type of innate immunomodulatory cells, have proved their potential in the treatment of several malignant tumors, so which received widespread attention in recent years.

**Objective:** The clinical data in TCGA and GEO was used to provide ideas for the diagnosis, treatment and evaluation of prognosis indicators of BLCA. And the relationship between BLCA and the expression of NRGs was investigated.

**Methods:** In this study, the normal and tumor samples of TCGA-BLCA and GSE3167 were analyzed to find out differential expression genes based on the "limma" R package. The genes and the natural killer cell-related genes (NRGs) were intersected to obtain NRGs with differential expression. We used univariate COX regression analysis, LASSO and multifactor COX analysis to obtain the risk score and build the prognosis gene model. GSE31684 queue has been verified externally. In order to further confirm the reliability of genes, we analyzed their mechanism of action, divided them into groups by the risk score, and performed immune analysis on subgroups to evaluate the immune treatment response.

**Results:** There are differences in the expression of LRP1 and INHBB, which are independent risk factors for the prognosis of BLCA. The nomogram was established based on the characteristics of 2-NRGs and clinicopathological characteristics. The low risk group has more abundant immune infiltration, suggesting a better immune response and better prognosis, which verifies the reliability of grouping. The sensitivity of the high-risk and low-risk group to the two target sites, PDL1 and FGFR3, is significantly different. And the risk score is closely related to the relevant chemotherapy drugs. These suggested that the risk score can screen the patients well at the drug level, and also providing a meaningful reference for the patients in drug selection.

**Conclusion:** NK cell core genes, LRP1 and INHBB, may play a crucial role in the occurrence and progression of BLCA. We constructed nomogram according to the NK core genes and verified its feasibility, which can provide a reference for clinical prognosis.

1 Introduction

Bladder urothelial carcinoma (BLCA) is one of the ten most common cancers in the world, which is the most common tumor in the urinary system \textsuperscript{[1]}. According to statistics, 40% of patients receiving surgical treatment have poor prognosis \textsuperscript{[2]}. The majority of these patients' reactions to chemotherapy are brief and transient. Only 12–15 months is the median overall survival for metastatic bladder carcinoma \textsuperscript{[3,4]}. With
the progress of systemic treatment, the overall survival period (OS) of patients has improved, but the prognosis is still poor [5]. The development of immunotherapy has changed the status initially. Although immune drugs have made remarkable progress on several tumors, it does not show multidimensional value in BLCA. The most important reason is that patients can not be well screened, classified and treated accurately. Immunotherapy is a method to enhance the function of immune cells to kill tumor cells by intervening in the patient's immune system or the ongoing cancer defense mechanism. Natural killer (NK) cells, as an important type of innate immunomodulatory cells, have proved their potential in the treatment of several malignant tumors, so which received widespread attention in recent years [6, 7]. NK cells have the ability to kill cells and regulate immune response through cytokines, which plays a key role in anti-cancer immunity [8, 9]. NK cells are considered as potential new targets for effective combined immunotherapy [10]. Extensive experiments have proved that chimeric antigen receptor NK cell therapy (CAR-NK) has significant safety and effectiveness, which cannot be achieved by CAR-T in the term of safety. In addition, NK cells induced by pluripotent stem cells can show different anti-tumor effects. With the continuous development of bioinformatics, people have various definitions of biomarkers. The role of NK cell related genes (NRGs) in BLCA is not clear at present. We aim to demonstrate the value of NRGs for assessing the prognosis of BLCA patients through a comprehensive analysis of genomic data, as well as to develop new tools for improving treatment options.

2 Methods

2.1 Data acquisition

The gene expression profile and clinical data of BLCA cohort including 421 tumor patients and 19 normal controls was downloaded from TCGA database (https://portal.gdc.cancer.gov). The dataset, GSE3167, downloaded from the GEO database (https://www.ncbi.nlm.nih.gov) included the gene expression of 41 BLCA patients and 19 normal controls. The above is used to search for differential genes. The genes related to NK cells were screened from the InnateDB database [11] http://www.innatedb.com/ for gene screening. 398 BLCA samples with complete clinical data were screened from TCGA database for further analysis. We downloaded the gene profile and clinical data of 93 BLCA patients in GSE31684 from GEO database. GSE31684 is considered as an external validation data set. The half maximum inhibitory concentration (IC50) of various drugs obtained from the CellMiner database [12–14] (https://discover.nci.nih.gov/cellminer/home.do/). This study was conducted in accordance with the Helsinki Declaration (revised in 2013).

2.2 Acquisition of differential genes

Due to the different sequencing methods used in different databases, the TCGA dataset and GSE3167 were analyzed independently. Based on the threshold | log2FC | > 0.5 and adjusted p value < 0.05, the differential genes in TCGA database and GEO database were obtained. The NRGs set, TCGA differential gene set (DEGs) and GEO DEGs were intersected to obtain the differential genes relevant to NK cells for subsequent analysis.
2.3 Model formula

We riddled the DEGs related to survival through univariate COX regression analysis. By performing univariate Cox regression analysis, we identified genes associated with survival, followed by LASSO regression analysis, and tenfold cross-validation was used to determine the penalty regularization parameter. Based on the optimal lambda value and corresponding coefficient, we build a risk characteristic model based on 2-NRGs. The NRGs risk score of each patient is calculated as follows: risk score = RNA1 expression value × Coef RNA1 + RNA2 expression value × Coef RNA2+... RNAn expression value × Coef RNAn.

2.4 Model validation

According to the equation above, all patients with BLCA were scored for risk. According the calculated median, all patients with BLCA were divided into low risk group and high risk group, and the survival curves of the two subgroups were drawn. According to different disease prognosis, bladder cancer can be divided into Muscle Invasive Bladder Cancer (MIBC) and NO Muscle Invasive Bladder Cancer (NMIBC). Because the prognosis of patients with NMIBC has been well predicted, we focused on the ability of the score to evaluate patients with MIBC or lymphatic metastases. Finally, we used the risk score and clinical characteristics to construct a graph to predict the 1-year, 3-year and 5-year survival curve of patients in the TCGA-BLCA cohort. We used ROC curve analysis to evaluate the prediction ability of the model.

2.5 Gene set enrichment analysis

The samples of patients with BLCA were grouped according to the score and the DEGs were screened. We calculated the correlation coefficient between scores and differential genes to understand the relationship between them better. The gene with p < 0.05 was selected for further functional analysis. And then we used gene ontology (GO) enrichment and Kyoto encyclopedia of genes and genes (KEGG) pathway analysis to study them[15–17]. Finally, we visualized GO and KEGG results.

2.6 Immune analysis

The immune infiltration score of BLCA patients was screened and visualized (p < 0.05). The risk score of BLCA patients was divided into two groups according to risk characteristics, and the relationship between immune cells and risk score was tested by Spearman correlation analysis. We downloaded the tumor purity, stromal cell level and immune cell infiltration level of BLCA patients from the ESTIMATE database, and observed the differences between high-risk and low-risk groups.

2.7 Drug sensitivity analysis

We evaluated the relationship between common immunotherapeutic drug targets and subgroups. The half maximal inhibitory concentration (IC50) from CellMiner database of various drugs and risk scores were analyzed using R's "limma" package.

2.8 Statistical analyses
R software (V 4.1.2) was utilized for data analysis. The significance between the two groups was identified using Wilcoxon test. The survival time differences between the two risk groups were estimated using Kaplan-Meier curves and log-rank test. Independent factors of OS were determined using univariate and multivariate COX regression analyses. P < 0.05 was the cut-off of statistical significance.

3 Results

3.1 Experimental workflow and acquisition of differentially expressed NRGs

The process is shown schematically in Fig. 1. Patient information is shown in Table 1. We used R's "limma" package to analyze the gene expression difference in BLCA patients. 318 NRGs were obtained with | log₂FC | > 0.5 after the intersection. The heat map (Fig. 2A,2B) and volcano map (Fig. 2C) were plotted by the “pheatmap” of R package. We used univariate COX analysis to 318 NRGs with differential expressed to identify NRGs associated with prognosis (P < 0.05), and twenty genes were screened (Table 2). LASSO regression analysis was performed for NRGs associated with prognosis. LASSO regression curves (Fig. 2D) and cross-validation plots (Fig. 2E) were obtained. Finally, two genes, LRP1 and INHBB, were obtained. Multiple COX regression factor analysis was performed on the two indicators to verify again, which proved both were risk factors for BLCA patients(Fig. 2F).
Table 1
The characteristics of patients

<table>
<thead>
<tr>
<th>The characteristics of patients</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
</tr>
<tr>
<td>&lt; 65</td>
<td>149</td>
</tr>
<tr>
<td>&gt;=65</td>
<td>248</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>293</td>
</tr>
<tr>
<td>Female</td>
<td>104</td>
</tr>
<tr>
<td><strong>Stage</strong></td>
<td></td>
</tr>
<tr>
<td>Stage I</td>
<td>3</td>
</tr>
<tr>
<td>Stage II</td>
<td>126</td>
</tr>
<tr>
<td>Stage III</td>
<td>138</td>
</tr>
<tr>
<td>Stage IV</td>
<td>130</td>
</tr>
<tr>
<td><strong>T</strong></td>
<td></td>
</tr>
<tr>
<td>T0</td>
<td>1</td>
</tr>
<tr>
<td>T1</td>
<td>4</td>
</tr>
<tr>
<td>T2</td>
<td>114</td>
</tr>
<tr>
<td>T3</td>
<td>190</td>
</tr>
<tr>
<td>T4</td>
<td>58</td>
</tr>
<tr>
<td>TX</td>
<td>30</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>231</td>
</tr>
<tr>
<td>N1</td>
<td>43</td>
</tr>
<tr>
<td>N2</td>
<td>75</td>
</tr>
<tr>
<td>N3</td>
<td>7</td>
</tr>
<tr>
<td>NX</td>
<td>41</td>
</tr>
<tr>
<td><strong>M</strong></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>193</td>
</tr>
</tbody>
</table>
### The characteristics of patients

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>11</td>
</tr>
<tr>
<td>MX</td>
<td>193</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GENE ID</th>
<th>p.value</th>
<th>HR (95% CI for HR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>INHBB</td>
<td>0.0013</td>
<td>1.5 (1.2–1.9)</td>
</tr>
<tr>
<td>ZYX</td>
<td>0.0058</td>
<td>0.69 (0.53–0.9)</td>
</tr>
<tr>
<td>PRDX1</td>
<td>0.0095</td>
<td>0.7 (0.53–0.92)</td>
</tr>
<tr>
<td>PSMC4</td>
<td>0.013</td>
<td>0.33 (0.14–0.79)</td>
</tr>
<tr>
<td>PMP2</td>
<td>0.013</td>
<td>0.53 (0.32–0.88)</td>
</tr>
<tr>
<td>LIF</td>
<td>0.013</td>
<td>0.68 (0.5–0.92)</td>
</tr>
<tr>
<td>RORB</td>
<td>0.017</td>
<td>150 (2.5–8800)</td>
</tr>
<tr>
<td>IL6ST</td>
<td>0.02</td>
<td>13000 (4.3·10^7)</td>
</tr>
<tr>
<td>CDH1</td>
<td>0.023</td>
<td>0.62 (0.4–0.94)</td>
</tr>
<tr>
<td>CXCR1</td>
<td>0.027</td>
<td>3 (1.1–8.2)</td>
</tr>
<tr>
<td>LRP1</td>
<td>0.032</td>
<td>2.1 (1.1–4.1)</td>
</tr>
<tr>
<td>PSMD11</td>
<td>0.034</td>
<td>0.75 (0.58–0.98)</td>
</tr>
<tr>
<td>AXL</td>
<td>0.036</td>
<td>0.014 (0.00026–0.76)</td>
</tr>
<tr>
<td>IL22</td>
<td>0.037</td>
<td>0.67 (0.46–0.98)</td>
</tr>
<tr>
<td>SEMA4F</td>
<td>0.043</td>
<td>0.26 (0.071–0.96)</td>
</tr>
<tr>
<td>CSF1</td>
<td>0.043</td>
<td>1.4 (1.1–1.8)</td>
</tr>
<tr>
<td>PSMD2</td>
<td>0.045</td>
<td>1.1·10^{-6} (1.8·10^{-12}–0.72)</td>
</tr>
<tr>
<td>BTC</td>
<td>0.048</td>
<td>21 (1·10^{-4})</td>
</tr>
<tr>
<td>ANGPT1</td>
<td>0.048</td>
<td>0.77 (0.6–1)</td>
</tr>
</tbody>
</table>

### 3.2 Establishment and verification of model

There were two cross-validation points for LASSO regression analysis at least. We used the multivariate COX regression model to identify and calculate the coefficient of the central gene (Fig. 2F), and constructed the prognostic index (PI)=\((0.768 \times \text{LRP1 exp.})+(0.4 \times \text{INHBB exp.})\). According to the median
risk score, patients with BLCA were divided into two groups: high-risk group and low-risk group. The prognosis of the high risk group in the TCGA cohort was significantly better than that of the low group (Fig. 3A). And then we verified with external data GSE31684. With the increase of risk, the survival period of patients with BLCA showed an increase in mortality (Fig. 3B, C, D), and the prognosis of the low-risk group was better (P < 0.05) (Fig. 3E). Based on the above results, we can see that the model is quite superior, and the expression of LRP1 and INHBB is positively correlated with risk score. To verify the accuracy of the risk score better, patients were divided into different subgroups for analysis. The series of patients with no lymphatic metastasis and no muscular infiltration had better prognosis, and fewer patients with this data type. Therefore, only five subgroups including male, female, T2-4, clinical grade 2–4, N1-N3 were explore in this study. The results all showed that the high-risk group had a poor prognosis, which was in line with the expected results (Fig. 4).

3.3 Establishment of nomograms in combination with clinical characteristics

Considering that the constructed risk model is significantly related to poor prognosis, we carried out univariate and multivariate Cox analysis to determine whether the prognostic characteristics of NGRs are independent predictors of survival. Multivariate cox analysis showed that the risk score and stage were significantly correlated with prognosis (p < 0.05) (Fig. 5A). Result risk score (p = 0.008) is an independent and reliable predictor of risk. In order to the clinical application of risk model, we built a nomogram (Fig. 5B) based on grading and risk score to predict the survival probability of BLCA patients in 1, 3 and 5 years. The results show that the prediction accuracy of the model is 1 year = 0.702, 2 years = 0.737, 5 years = 0.720, showing a good prediction effect (Fig. 5C).

3.4 Enrichment analysis

The differential genes between the two subgroups were evaluated by KEGG enrichment analysis and GO function analysis to clarify the correlation between biological activity, signal pathway and risk score. The significantly enriched items were screened with FDR < 0.05 and p < 0.05. Biological process (BP) mainly includes the composition of extracellular matrix and structure. Cell components (CC) mainly include extracellular matrix, actin and focal adhesion kinase. Molecular function (MF) mainly includes the regulatory activity of GTPase and the regulatory activity of nucleotide triphosphatase (Fig. 6A). The pathways enriched by KEGG mainly include PI-3K-AKT pathway and actin regulation (Fig. 6B). We found that there was a strong correlation between risk score with cell chemotaxis and cell energy supply. Since this gene is a set of NK related genes, we then systematically analyzed the immune landscape of two subgroups of BLCA patients.

3.5 Risk score and immune cell infiltration and immune environment
Tumor immune cell infiltration is widely recognized as one of the important immune characteristics of tumor immune microenvironment (TME). In order to understand the distribution and features of the relative amounts of 22 tumor-infiltrating immune cells in the TCGA-BLCA cohort, we calculated the level of immune cell infiltration in each sample using the CIBERSORT (Fig. 7A). In the high-risk group, NK cells decreased and macrophages increased (Fig. 7B). We also discussed the difference in the expression of immune checkpoints between the two groups, because immune checkpoints are important for the effectiveness of tumor immunotherapy. The gene expression of two immune checkpoints in high-risk population was significantly up-regulated, including PD1 and FGFR3. NTRK3 in the low-risk group was up-regulated (Fig. 7C). Upregulation of immune checkpoints is a key feature of TME, which may indicate that high-risk patients are in the inflammatory microenvironment. Targeted treatment of immunocheckpoints with elevated expression may be beneficial to patients with this subtype of tumor. Subsequently, the interstitial score, immune score and estimated score of the low-risk group were higher (p < 0.001), indicating that the overall immune level and immunogenicity of TME in the low-risk group were higher (Fig. 7D, E, F), and the high immune level, indicating a good prognosis and supporting the grouping with good risk score.

3.6 Drug sensitivity analysis

The drug sensitivity analysis displayed the risk score was negatively correlated with the IC50 of AT-13387, CUDC, Palbocicli, By-Product of CUDC-305, but positively correlated with the IC50 of Motesanib, Deforolimius, Apitolisib, AZD-8055 and so on (Fig. 8).

4 Discussion

NK cells play an important role in tumor microenvironment and immune monitoring, and their related genes are receiving more and more attention [18]. NK cell, which does not be limited by major histocompatibility complex, can directly act on tumor cells without antigen stimulation and antibody [19]. At present, there has been NK cell therapy as follows: recruitment and activation of NK cells [20], CAR-NK [21], natural killer cell membrane-cloaked virus-mimicking nanogenerator [22]. However, NRGs in BLCA have not been comprehensively studied. The purpose of this experiment is to explore the application of NRGs in BLCA.

Therefore, we used mRNA expression data from TCGA-BLCA dataset to identify important prognostic genes, and designed a two-biomarker prognostic model based on NRGs. In this study, we integrated the differential expression NRG profile in the TCGA-BLCA and GEO-GSE3167 data set. And we screened the two genes, LRP1 and INHBB using LASSO regression analysis and COX risk regression analysis. An external data set, GSE31684, was subsequently used to confirm the model's excellent prediction of patients. We constructed nomogram to apply it in clinical practice better. Later, we continued to verify the reliability of the model, and analyzed the prognosis, immunological characteristics of subgroups and drug sensitivity analysis.
LRP1, a member of the low density lipoprotein receptor family of proteins encoded by this gene, is involved in a variety of cellular processes, including intracellular signal transduction, lipid homeostasis and the elimination of apoptotic cells \[^{23}\]. It was included in one of the prognostic evaluation genes of BLCA by many authors \[^{24, 25}\]. INHBB is a protein-coding gene, mainly involved in synthesis of transforming growth factor β (TGF-β) superfamily members \[^{26}\]. INHBB and LRP1 are involved in a variety of tumors and immune functions, and a better prognosis model can be obtained by integrating their risk scores.

Subsequently, the differential genes of the two subgroups were used for enrichment analysis. Interestingly, the results were mainly concentrated in the cell transfer and metabolic pathways. Cell metastasis suggests that these two genes may be related to cancer invasion, which is the same as the experimental results of Kita \[^{26}\]. INHBB can activate SMAD pathway to promote cancer metastasis \[^{27, 28}\]. In addition, INHBB has been reported to be associated with lymph metastasis, and its mechanism may be related to p53 \[^{29}\]. EHSP90a which has been reported to promote the activity of cancer cells plays a role in tumor metastasis and immune cell trend by promoting the AKT pathway involved by LRP1 \[^{30, 31}\]. In addition, it is also a drug-related gene \[^{32}\]. After consulting the literature, we learned that whether the function of NK cells can work is closely related to cell metabolism functions, such as cell hypoxia, energy supply, amino acid cycle, etc \[^{33, 34}\]. FasL-related GTP-related signal pathway is also involved \[^{35}\]. The results and enrichment results corroborated with each other. Therefore, it is speculated that the regulatory activity of GTPase and the regulatory activity of nucleotide triphosphatase are the important links that hinder NK cells from playing their roles.

Since immune cells are the cellular basis of immunotherapy, a thorough understanding of immune infiltration in TME is critical to uncover the underlying molecular mechanisms and provide new immunotherapy strategies to improve clinical outcomes. Both genes are related to immunity \[^{36, 37}\], so we carried out immunoinfiltration analysis on the collected patients data. It was shown that the tumor immune cells were mainly CD8 positive T cells and B cells. Subgroup analysis showed that it was also related to M1 polarization of macrophages. M1 polarization of macrophages suggests poor prognosis \[^{38}\]. The increase of multiple targeted markers in high-risk subgroups indicates good sensitivity to targeted drugs. Finally, we conducted a drug sensitivity analysis, and the most relevant risk score is Motisanil, which provides a reference for further exploring the role of genes. Studies have proved that this drug can affect the apoptosis of bladder cancer through PI3K/AKT pathway, and it is considered to be a possible drug to treat bladder cancer \[^{39}\].

5 Conclusions

2-NRGs signature and nomograms demonstrate excellent predictive performance and offer new perspectives for assessing pre-immune efficacy, which will facilitate future precision immuno-oncology research.
Declarations

Author Contributions:


Competing interests:

The author(s) declare no competing interests.

Data availability:

The datasets generated and/or analysed during the current study are available in the TCGA and GEO repository.

References


Figures
Figure 1

The procedure of this experiment
Figure 2

The difference of gene expression and Lasso analysis. A Volcano plot of differentially expressed TCGA DEGs. (B) Volcano plot of differentially expressed GEODEGs. (C) Veen diagram. (D) Ten-time cross-validation for tuning parameter selection in the LASSO mode. (E) LASSO coefficient profiles. (F) Results of the multivariate regression analysis.
Figure 3

Construction and validation of NRGs signature. (A) KM curve compares the overall BLCA patients between low- and high-risk groups in the TCGA cohort. (B) Distribution of risk scores between low- and high-risk groups in the GSE31684 group. (C) Survival status of BLCA patients in the low- and high-risk groups in the GSE31684 group. (D) The gene expression in the low- and high-risk groups in the GSE31684 group. (E) KM curve compares the overall BLCA patients between low- and high-risk groups in the GSE31684 cohort.
Figure 4

The risk score is a valuable marker for poor prognosis in various subgroups divided by clinicopathological characteristics. The risk score could distinguish high-risk patients in a variety of subgroups divided by clinicopathological characteristics including gender A B , MIBC C D , and lymph metastasis E.
Figure 5

Establishment of the nomogram. A Multivariate COX regression analysis of the signature and different clinical feature. (B) Nomogram for predicting 1, 3, and 5-year OS of patients with HNSCC. (C) Time-dependent ROC curves analysis.

Figure 6

Enrichment analysis. A GO classifications of the differentially expressed proteins. The number of proteins involved in each of the biological process, cellular, component. B KEGG analysis of differentially expressed genes.
Figure 7

Risk score predicts tumor microenvironment and immune cell infiltration. (A) Immune cell in BLCA. (B) Differences in immune cell. (C) Immune checkpoint differences between high- and low-risk groups. (D) Stromal score. (E) Immune score. (F) Estimate score.
Figure 8

Correlation between the expression of risk score and IC50 of multiple drugs