

Construction of An Immune-Related CeRNA Network as a Prognostic Signature in Bladder Cancer

Ran-ran Zhou

The Third Affiliated Hospital of Southern Medical University

Hu Tian

The Third Affiliated Hospital of Southern Medical University

Cheng Yang

The Third Affiliated Hospital of Southern Medical University

Fan Peng

The Third Affiliated Hospital of Southern Medical University

Hao-yu Yuan

The Third Affiliated Hospital of Southern Medical University

Jun-hao Zhou

The Third Affiliated Hospital of Southern Medical University

Xi Zhi

The Third Affiliated Hospital of Southern Medical University

cundong liu (✉ cundongliu@163.com)

Southern Medical University <https://orcid.org/0000-0002-2098-1139>

Primary research

Keywords: bladder cancer, ceRNA, prognosis, immune infiltration

Posted Date: September 2nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-70360/v1>

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

[Read Full License](#)

Abstract

Background: Immunotherapy has been proved to be effective for bladder cancer (BLCA). However, the molecular network involved in BLCA tumor immune response remains unclear. This study aims to construct an immune-related ceRNA network and to identify the prognostic value.

Methods: Based on The Cancer Genome Atlas (TCGA), we used single-sample gene set enrichment analysis (ssGSEA), weighted gene co-expression network analysis (WGCNA) to determine immune-related mRNA, lncRNA and miRNA. Then least absolute shrinkage, and selection operator (LASSO) and Cox regression were performed to identify the mRNAs with high prognostic value, and accordingly, the risk score was calculated. Internal and external validation were performed both in TCGA and GSE13507 with Kaplan-Meier (KM) survival and Receiver Operating Characteristic (ROC) curve analysis. Using the immune-related mRNA, lncRNA and miRNA, a ceRNA network was established via MiRcode, starBase, miRDB, miRTarBase and TargetScan. Besides, we also explore the relationship between the risk score and immune cell infiltration via CIBERSORT algorithm.

Results: 5 mRNAs (PCGF3, FASN, DPYSL2, TGFBI and NTF3) were ultimately identified, and KM survival analysis displayed the 5-mRNA risk signature could predict the prognosis of BLCA with high efficacy both in TCGA ($p = 1.006e-13$) and GSE13507 ($p = 7.759e-04$). Using miRNA targeting molecular prediction database, an immune-related ceRNA network, including 5 mRNAs, 24 miRNAs and 86 lncRNAs, was constructed. Memory B cells, activated dendritic cells, and regulatory T cells infiltration into tumors were negatively correlated with risk score, while the infiltration levels of macrophages M0, M1 and M2 were positively correlated with risk score.

Conclusion: This study helped to better understand the molecular mechanisms of tumor immune response from the view of ceRNA hypothesis, and provided a novel prognostic signature for bladder cancer.

Introduction

Bladder cancer (BLCA) is among the most common malignancies worldwide with high morbidity and mortality^{1,2}. Most BLCA patients still need postoperative adjuvant therapy even after radical cystectomy which remains the main treatment for BLCA, especially for muscle invasive bladder cancer³. Adjuvant chemotherapy has been widely accepted as a feasible measure to improve the prognosis of BLCA⁴. Currently, immunotherapy mainly including traditional Bacillus Calmette-Guérin (BCG) and emerging immune checkpoint inhibitors (ICIs), which could strengthen anti-tumor immunological ability in theory, has been demonstrated to be effective for BLCA, providing new choice for patients with advanced bladder cancer or resistance to conventional medical practice^{5,6}. Thence, it is of high significance to explore the immune-related prognostic indicators and relevant biological mechanisms based on immunobiology research.

The competing endogenous RNA (ceRNA) hypothesis, in which lncRNA, mRNA and circRNA could regulate the expression level of each other by competitive binding miRNAs response elements⁷, has been proved to be involved in carcinogenesis of various tumors by emerging evidence^{8,9}. Establishment of a ceRNA network could provide novel insights into tumor regulatory networks. The tight relationship has been found between tumor-infiltrating immune cells and ceRNA networks in various tumors like head and neck squamous cell cancer¹⁰, soft tissue sarcoma¹¹ and mesothelioma¹², while this kind of research on BLCA has not been received enough attention. Previous study has successfully constructed a regulatory ceRNA network, which could be a prognostic predictor for BLCA patients and displayed potential biological mechanisms from the angle of ceRNA hypothesis¹³. However, this study was only based on The Cancer Genome Atlas (TCGA) database and external validation was lacking. Besides, the analysis of the roles ceRNA played in tumor immune response of BLCA, especially based on large datasets, has not been illustrated completely, which is urgently demanded.

In this study, we aimed to screen hub immune-related genes including lncRNA, mRNA and miRNA via single-sample gene set enrichment analysis (ssGSEA), weighted gene co-expression network analysis (WGCNA) and other bioinformatical tools, and to construct a lncRNAs-mRNAs-miRNAs ceRNA network based on TCGA. Moreover, an immune-relevant prognostic signature was established through Kaplan-Meier (KM) survival analysis, LASSO regression and Cox regression, and the robustness of this model has been verified in TCGA-BLCA dataset and GSE13507 dataset, respectively. We also explored the correlation of the risk signature with tumor-infiltrating cells (TICs) proportion and tumor immune microenvironment by means of CIBERSORT and ESTIMATE algorithm.

Materials And Methods

1. Data collection

The RNA-seq expression matrix and corresponding clinicopathological parameters of patients with BLCA were extracted from TCGA on UCSC website (<https://xena.ucsc.edu/>)¹⁴. The gene expression data were downloaded in the forms of “HTSeq-Counts” and “HTSeq-FPKM” for genomic expression divergence detection and prognostic model building. A total of 430 samples, including 411 BLCA samples and 19 normal samples, were included in this study. With the annotation file downloaded from GENECODE (version 22, GRCh38), Ensembl IDs were transformed into gene symbol and the expression matrix of lncRNAs and mRNAs were singled out. The external validation dataset was acquired from GSE13507¹⁵ in Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>). The series matrix file, along with the detailed survival information, of 165 enrolled BLCA patients in GSE13507 were downloaded directly from GEO website. All these data were analyzed with R software (version 3.5.3).

2. Identification of immune-related lncRNAs, mRNAs and miRNAs via ssGSEA

Gene markers of 29 immune-related signatures, including immune cells, immune-relevant pathways and immune-relevant functions, were obtained from previous studies^{16,17}. Gene Set Variation Analysis

(GSVA) package ¹⁸ of R was utilized to evaluate the activation and infiltration level of these immune-related biological events according to the expression profile of BLCA from TCGA. Based on the ssGSEA score matrix, 411 BLCA patients were grouped into high- and low-immune-infiltration with “hclust” R package. To evaluate the potency of ssGSEA grouping, Stromal score, Immune score, ESTIMATE score and tumor purity, which were obtained from ESTIMATE algorithm ¹⁹ on the basis of BLCA gene expression profile, were analyzed. The expression divergence of human leukocyte antigen (HLA) were verified between two immune-related groups by Wilcoxon sum rank test. KM survival analysis was used to compare the survival rate of these two groups. To ascertain the immune-related lncRNAs, mRNAs and miRNAs, “edgeR” package ²⁰ was implemented to detect the differentially expressed mRNAs (DEmRNAs), lncRNAs (DElncRNAs) and miRNAs (DEmiRNAs) between high- and low-infiltrating groups based on the criteria of $|\log_2FC| > 0.5$ and $p < 0.05$. The same filtering criteria was applied for genomic divergence analysis between normal and cancer groups in order to select the genes both correlated immune response and tumorigenesis, and venn diagrams were used to visualize the interaction analysis.

3. Weighted gene co-expression analysis (WGCNA)

To further identify the immune-related lncRNAs, WGCNA R package ²¹ was conducted to construct a lncRNA co-expression network to determine the gene modules with high correlation with immune score, which could reflect the ratio of immune component of tumor microenvironment (TME) and was acquired from ESTIMATE algorithm mentioned above. First, Pearson correlation coefficients of paired genes were calculated, and a weighted adjacency matrix was constructed in light of the formula $am_n = |cm_n|^\beta$ (am_n : adjacency between two paired genes; cm_n : pearson correlation coefficient of two genes; β : soft-thresholding parameter, which could strengthen the robust correlation and penalize the weak correlation of genes). Second, the optimal β value was selected to establish a scale-free co-expression network. After transforming the adjacency matrix into a topological overlap matrix (TOM), hierarchical clustering was performed to assign genes with analogous expression pattern into the same module ²² with a minimal gene size of 50. Next, the correlation between the modules and Immune score was explored, and the modules with high and statistically significant module significance (MS) score, which is defined as the average \log_{10} -transformed p-value of the correlation between the genes in the module and Immune score, were chosen. The lncRNAs in these modules were regarded as the immune-related lncRNAs.

4. Construction of a ceRNA network

MiRcode (<http://www.mircode.org/>) ²³, a database used for targeted miRNA prediction, was implemented to forecast the miRNAs with potential to bind with the lncRNAs obtained from WGCNA. To avoid false-positive predictive results as possible, then miRDB (<http://www.mirdb.org/>) ²⁴, miRTarBase (<http://www.mirtarbase.cuhk.edu.cn/>) ²⁵ and TargetScan (http://www.targetscan.org/vert_72/) ²⁶ were

all used for mRNAs prediction based on the miRNAs obtained from the previous step analysis. The interaction analysis was performed between these mRNAs and immune-relevant DE mRNAs. After detecting the hub prognostic mRNAs, the targeted miRNAs and lncRNAs were determined via miRcode and starbase (<http://www.starbase.sysu.edu.cn/>)²⁷, respectively. The interaction of the ceRNA regulatory network was visualized by means of Cytoscape (version 3.8.0).

5. Development and validation of an immune-related risk signature

399 BLCA patients with complete overall survival (OS) information from TCGA were randomly grouped into a training dataset (n = 280) and an internal validation dataset (n = 119). Considering the probable unrelated wherefore of death, the patients with the survival time of less than one month were removed from the study. The overlapped mRNAs, which have been codetermined by genomic difference detection and lncRNAs prediction by the mentioned-above methods, were chosen for further survival analysis. First, according to the median expression of genes, patients with BLCA were divided into two groups, and KM survival analysis was then performed with the criteria of $p < 0.05$ via “survival” R package. Following that, univariate Cox regression was utilized to explore the correlation between hub mRNAs expression level and OS using the same R package. The genes with $p < 0.05$ were singled out for least absolute shrinkage, and selection operator (LASSO) regression, which was conducted by “glmnet” R package. At last, multivariate Cox analysis with stepwise regression was carried out to set up a prognostic model, and the risk score of each BLCA patient was calculated. With “survivalROC” R package, Receiver Operating Characteristic (ROC) curves were drawn to validate the effectiveness and robustness of the risk signature in the training dataset, the internal validation dataset, the entire TCGA-BLCA dataset and the external validation dataset. Meanwhile, after ascertaining the optimal cut-off value of risk score by X-tile software²⁸, all BLCA patients including 399 cases from TCGA and 165 cases from GSE13507 were divided into high- and low-risk group. Subsequently, KM survival analysis was conducted to reconfirm the prognostic value of the risk score. Besides, we also used the external validation dataset (GSE13507) to explore the respective prognostic value of the mRNAs with $p < 0.05$ in the prognostic model. KM survival analysis was performed and X-tile helped to determine the optimal cut-off value. In addition, univariate Cox regression was performed to validate the prognostic value of the predicted miRNAs and lncRNAs in the ceRNA network via “survival” R package.

6. Construction of a risk-score-based nomogram

Clinical phenotype and the immune-related risk score from TCGA-BLCA were transformed into binary variables, respectively. Then univariate Cox regression and multivariate Cox regression were conducted not only to identify whether the risk score was an independent predictor for BLCA prognosis, but also to construct a nomogram integrating the risk score and other important clinical parameters by means of “survival”, “rms” and “foreign” R packages to obtain a better predictive performance. Similarly, the

variables with $p < 0.05$ in univariate analysis were picked into multivariate Cox model. ROC curve, KM survival curve and calibration curves were drawn to confirm the efficacy of the nomogram.

7. Correlation between risk score and BLCA immune response

In this part, we mainly evaluated the correlation of the gene-based risk score with the expression level of crucial immune-checkpoint-relevant genes, Immune score which could account for the proportion of immune component in TME, and the infiltrating level of TICs. To sufficiently explore the relationship between risk score and tumor immune response, all BLCA patients were labeled with high- or low-risk according to the median risk score. The expression profile of five critical immune checkpoints, including TIGIT, LAG3, HAVCR2, CTLA4 and PDCD1, were extracted from TCGA-BLCA gene expression matrix with $\log_2(\text{FPKM}+1)$ transformation. Pearson coefficients were calculated and only $p < 0.001$ was considered as statistically significance. Meanwhile, the expression level of these five genes in high-risk group was compared with that in low-risk group by Wilcoxon test. The linkage between the hub mRNAs in the gene risk signature and pivotal immune-checkpoint-relevant genes was also explored. The correlation with $R > 0.3$ and $p < 0.05$ was deemed to be significant

ESTIMATE, an algorithm developed to evaluate the relative abundance of stromal and immune components in tumor, was utilized to calculate the Immune score. The correlation was assessed both using difference test and correlation test.

To check the tight relationship between risk score and immune TME, CIBERSORT algorithm²⁹ was implemented to reckon the infiltration ratio of 22 different kinds of TICs with the filtering threshold of $p < 0.05$. Then the proportion difference was explored between high- and low-risk group. Similarly, to obtain reliable results as possible, we calculated the Pearson correlation coefficients between risk score and infiltration proportion of each TIC.

8. Functional annotation analysis

The Hallmark gene sets, which were available in the MSigDB database (<https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>)³⁰, were downloaded as reference datasets for Gene Set Enrichment Analysis (GSEA) and Gene Set Variation Analysis (GSVA). These analysis was performed through GSEA software³¹ (version 4.0.3) and "GSVA" R package, respectively. According to the median risk score, the enrolled BLCA patients from TCGA all divided into high- and low-risk group. For GSVA, $|\log_2\text{FC}| > 0.1$ and $p < 0.05$ were selected as cut-off criteria. For GSEA, $p < 0.001$ and false discovery rate (FDR) < 0.001 were set as the significance threshold. The common significantly changed pathways obtained by GSEA and GSVA were reconfirmed by measuring the Pearson correlation between risk score and ssGSEA score of enriched pathways.

Results

1. Identification of immune-relevant DEmRNAs, DElncRNAs and DEmiRNAs via ssGSEA

The workflow of the whole study is displayed on **Fig. 1**. Via unsupervised clustering algorithm, 411 bladder cancer patients were divided into two groups, including low-immune-infiltrating group (n = 314) and high-immune-infiltrating group (n = 97) (**Fig. 2a**). The expression level of HLA family in high-immune-infiltrating group was significantly higher than that in low-immune-infiltrating group ($p < 0.001$), verifying the utility of the immune grouping strategy (**Fig. 2b**). Besides, using KM survival analysis, BLCA patients with high immune infiltration have poorer survival than those with low immune infiltration, which is logical and consistent with previous reports (**Fig. 2c**). After confirming the feasibility of our grouping method, 3257 DElncRNAs (**Fig. 2e**), 212 DEmiRNAs (**Fig. 2k**) and 6012 DEmRNAs (**Fig. 2h**) were screened between high- and low-infiltration groups through “edgeR” package. Besides, 2940 DElncRNAs (**Fig. 2d**), 446 DEmiRNAs (**Fig. 2j**) and 8049 DEmRNAs (**Fig. 2g**) were detected in BLCA samples compared with the paracancerous tissue. Venn analysis indicated 1030 DElncRNAs (**Fig. 2f**), 2836 DEmRNAs (**Fig. 2i**) and 134 DEmiRNAs (**Fig. 2l**) were both correlated with tumorigenesis and tumor immune response, and these genes will be used for following investigation.

2. Immune-related lncRNAs obtained by WGCNA

From the analysis above, 1030 DElncRNAs were regarded to be associated with BLCA immune response. In order to further clarify the immune-relevant lncRNAs, WGCNA was performed based on Immune score. As shown in **Fig. S1a&b**, soft-thresholding value 2 was chosen to construct a scale-free network ($R^2 = 0.90$). We have identified a total of four divergent modules, including blue module, turquoise module, brown module and yellow module (**Additional file 1: Fig. S1c**). Among these modules, brown module ($R = 0.45$, $p = 3e-22$) and yellow module ($R = 0.54$, $p = 3e-32$) were picked for subsequent analysis (**Additional file 1: Fig. S1d**). A sum of 170 lncRNAs, including 102 lncRNAs in brown module and 68 lncRNAs in yellow module, were reconfirmed to be associated with BLCA immune response.

3. Overlapped mRNAs of predicted mRNAs and immune-related DEmRNAs

170 lncRNAs, which were obtained from the establishment of a scale-free co-expression network, were then used to predict the possible targeted miRNAs using miRcode website, and as a result, 75 miRNAs were obtained. With miRDB, TargetScan and miRTarBase, 75 predicted miRNAs from the previous step analysis were uploaded and 1933 common mRNAs were acquired. Venn analysis was conducted between 1933 predicted mRNAs and 2863 immune-relevant DEmRNAs, and 182 overlapped mRNAs were found and further explored in the following survival analysis.

4. Construction and validation of a five-gene prognostic signature

First, according to the median expression level of each gene among 182 overlapped mRNAs, 399 BLCA patients with detailed survival data were divided into two groups, and accordingly, KM survival analysis

could be put into practice. 48 genes were preliminarily believed to be correlated with prognosis of BLCA (**Additional file 4**). As displayed in **Table.1**, no significant divergence of clinical features between the training dataset and validation dataset was found with the filtering threshold of $p < 0.05$ by means of Chi-square test. After the stochastic grouping of 399 patients, univariate Cox regression with filtering criteria of $p < 0.05$ helped to identify 24 genes in the training dataset (**Additional file 5: Table. S1**). By performing LASSO (**Fig. 3a&b**) and multivariate Cox regression, a five-gene prognostic model was built as follows: $-0.4231 \times \exp(\text{PCGF3}) + 0.4164 \times \exp(\text{FASN}) + 0.2498 \times \exp(\text{DPYSL2}) + 0.1477 \times \exp(\text{TGFBI}) - 0.6484 \times \exp(\text{NTF3})$, where exp means the mRNA expression level of the particular gene. In addition, we found PCGF3, FASN, DPYSL2 and TGFBI were independent predictors in this model ($p < 0.05$; **Fig. 3c**), implying the tight relationship between these four genes and the process of tumor immunobiology. KM survival analysis displayed these four genes also had predictive ability for BLCA prognosis in GSE13507 cohort (**Fig. 6d-g**). The patients with higher expression of DPYSL2, TGFBI and FASN and lower expression of PCGF3 would suffer poorer prognosis, which was consistent with the analysis results from TCGA.

To validate the robustness of the five-gene model, internal and external validation was performed in TCGA-BLCA cohort and GSE13507 cohort, respectively. Time-dependent ROC analysis showed the areas under curves (AUCs) at 1-, 3- and 5-year was 0.672, 0.701 and 0.714 in the training dataset (**Fig. 3d**). The same analysis was conducted in internal validation dataset (**Fig. 3e**), the entire TCGA-BLCA dataset (**Fig. 3f**) and external validation dataset (**Fig. 3j**). KM log-rank test survival curves were also drawn to validate the predictive ability for BLCA prognosis. "2.05" was selected as the optimal cut-off value of risk score by X-tile, and then all patients, including ones in the training dataset, internal validation dataset and external validation dataset, were grouped into high- and low-risk. The cases in high-risk group significantly suffered less survival time than that in low-risk group no matter in the training dataset ($p = 1.554e-09$; **Fig. 3g**), the internal validation dataset ($p = 1.959e-05$; **Fig. 3h**), the entire TCGA-BLCA dataset ($p = 1.006e-13$; **Fig. 3i**) and GSE13507 cohort ($p = 7.759e-04$; **Fig. 3k**). In all, our five-gene model was a strong tool to predict the prognosis of BLCA patients across the datasets and platforms.

5. Establishment a risk-score-based nomogram

To clarify whether the risk-score, which was highly correlated with tumor immune response, was an independent predictive factor for BLCA prognosis compared with other clinicopathological traits, univariate and multivariate Cox analysis was performed. In the univariate Cox model, the risk score, along with age, pathological N stages, pathological T stages and tumor stages, were responsible for the prognosis of BLCA ($p < 0.05$; **Fig. 4a**). Age, pathological N stages and risk score were reconfirmed as independent prognostic factors through multivariate Cox regression analysis ($p < 0.05$; **Fig. 4b**), and based on the results of multivariate Cox analysis, a nomogram consisting of age, N stages, T stages and tumor stages was constructed to predict the 3- and 5-year survival through the quantitative scoring strategy (**Fig. 4c**). Using the nomogram, each patient with BLCA could get a total point. ROC analysis (**Fig. 4d**) and KM survival analysis (**Fig. 4e**), which is based on the median total points, demonstrated the higher total score represented the greater mortality. Calibration curves (**Fig. 4f&g**) indicated the high similarity between our model and ideal model.

6. Correlation between risk score and BLCA immune response

To prove the close relevance between the risk score and the immune infiltrating status of BLCA, Immune score, immune-checkpoint-related gene and infiltrating proportion of TICs were picked for further analysis. The Immune score (**Additional file 2: Fig. S2a**), Stromal score (**Additional file 2: Fig. S2b**) and ESTIMATE score (**Additional file 2: Fig. S2c**) were positively correlated with the risk score, and these scores in the high-risk group were significantly higher than those in the low-risk group (**Additional file 2: Fig. S2d-f**). Besides, to further explore the relationship between the risk score and immune TME, the infiltration ratio of 22 different TICs was calculated via CIBERSORT (**Additional file 3: Fig. S3**). The outcome from difference test and correlation test indicated that a sum of 6 types of TICs were significantly correlated with the risk score (**Fig. 5**). Among them, three kinds of TICs have positive correlation with risk score, including macrophages M0, M1 and M2, while three kinds of TICs have negative correlation, including memory B cells, activated dendritic cells, and regulatory T cells. We also detected the correlation among risk score and crucial immune-checkpoint-relevant genes (TIGIT, PDCD1, LAG3, HAVCR2 and CTLA4), and found the expression of critical immune checkpoints in the high-risk group was significantly higher than that in the low-risk group (**Fig. 6h&i; Additional file 6: Table. S2**), implying the poor prognosis of patients with high risk might be partly due to the increasing expression of immune checkpoints. All this evidence above strongly demonstrated our risk score was significantly associated with the process of BLCA immune response. To verify the immunological relationship between the immune-checkpoint-relevant genes and the four hub genes, which were independent predictors in our risk model, we calculated the Pearson correlation coefficients and the results of this analysis were shown in **Fig 6j** and **Additional file 7: Table. S3** ($R > 0.3; p < 0.05$). Remarkably, FASN and PCGF3 had negative correlation with the crucial immune-checkpoints, while DPYSL2 and TGFBI had positive correlation.

7. Establishment of an immune-related ceRNA network

Through miRcode online tool, 199 miRNAs were obtained based the five hub mRNAs which included DPYSL2, TGFBI, NTF3, PCGF3 and FASN. After integrating the immune-related DE miRNAs, which were acquired by our first-step analysis of the whole process, 24 overlapped miRNAs were identified (**Fig. 6c**). Similar process of analysis was conducted on the predicted lncRNAs. However, starBase instead of miRcode was used for lncRNA prediction. Ultimately, a mRNA-miRNA-lncRNA ceRNA network including 5 mRNAs, 24 miRNAs and 86 lncRNAs was constructed (**Fig. 6a**). The correlation analysis among the five hub mRNAs was displayed in **Fig. 6b**. **Additional file 8: Table. S4** showed the prognostic value of 24 miRNAs and 86 lncRNAs via univariate Cox regression.

8. Identification of risk-score-related pathways

GSEA displayed the ssGSEA scores of 9 pathways in high-risk group were significantly higher than those in low-risk group by "limma" R package (**Fig. 7a**). Top 10 terms ranked by Normalized Enrichment Score (NES) in high-risk group via GSEA software were chosen and visualized by R (**Fig. 7b**). The interaction analysis showed five pathways, including allograft rejection, complement, E2F targets, inflammatory response, and epithelial mesenchymal transition (EMT) pathways, were mostly associated with risk score

(Fig. 7c). Positive correlation between risk score and the ssGSEA score of the five significantly changed pathways was reconfirmed by Pearson significance test (Fig. 7d).

Discussion

Bladder cancer (BLCA), carrying a great social burden with about 170000 patients died annually³², has been widely demonstrated to be able to benefit from immunotherapy³³, so exploring the underlying mechanism involving in the immune response of BLCA is necessary and meaningful. The increasing researches of ceRNA network suggested the upset of the equilibrium of regulatory ceRNA network played nonnegligible roles in the progress of diseases, especially cancers. A recent study¹³ has constructed a ceRNA network, which was of high prognostic value, to achieve a deeper comprehension to the RNA regulatory network in BLCA. Besides, although several immune-relevant signature³⁴⁻³⁶ have been constructed to predict the survival of BLCA, few studies concentrated on immune-relevant genes revealing the prognosis based on ceRNA hypothesis. In this study, we integrated the TCGA-BLCA database and GEO database to determine an immune-related risk signature and constructed an immune-relevant ceRNA network including 5 mRNAs, 24 miRNAs and 86 lncRNAs, leading to a better understanding of the oncological immunology of BLCA from the angle of RNA interactions.

Utilizing ssGSEA strategy, which has been proved to be effective for tumor grouping based on immune infiltrating status in multiple studies^{17,35,37}, 411 BLCA patients were labeled with high-immune-infiltrating and low-immune-infiltrating, and then genomic difference detection was performed to achieve the latent immune-relevant lncRNA, mRNA and miRNA. To identify the genes correlated with tumorigenesis of BLCA, we compared the genomic expression difference between the paracancerous tissues and malignant bladder tissues by “edgeR” package of R. WGCNA was implemented to obtain the lncRNAs mostly relevant to the process of immune response. Multiple databases, including MiRDB, miRcode, miRTarBase, TargetScan and starBase, were used to predict the possible binding profile of the lncRNAs, mRNAs and miRNAs in order to ensure the accuracy of the prediction of target molecules as possible. KM survival analysis, univariate Cox regression, LASSO algorithm and multivariate Cox regression were implemented to develop a risk model based on 182 overlapped mRNAs. Through the internal and external validation, 5 mRNAs including PCGF3, FASN, DPYSL2, TGFBI and NTF3 were identified as important prognostic biomarkers and the risk score was successfully constructed. ESTIMATE algorithm, CIBERSORT algorithm and the exploring of the correlation between risk score and crucial immune checkpoints were conducted, verifying our risk score was significantly correlated with BLCA immune process. At last, we established an immune-relevant ceRNA network based on these 5 critical mRNAs, providing novel insights of the tumor immunity of BLCA.

Among these five mRNAs, PCGF3, FASN, DPYSL2 and TGFBI were independent predictors and displayed significant prognostic value in the external validation dataset. Polycomb Group Ring Finger 3 (PCGF3), involving in polycomb group multiprotein PRC1-like complex, could inhibit the transcription of many genes in the developmental process³⁸. It has been reported that PCGF3 could serve as a biomarker for

the efficacy of docetaxel, cisplatin and S-1 (DCS) on patients with advanced gastric cancer³⁹. Besides, some study reported the relatively high expression level of RING1 protein, such as PCGF3, could reveal a favorable prognosis in non-small cell lung cancer⁴⁰, implying PCGF3 might be an antioncogene during the tumorigenesis. In our study, we firstly found the expression of PCGF3 acted as a favorable predictor in BLCA and was negatively correlated with the expression of immune-checkpoint-relevant genes. Most of the current researches of PCGF3 are based on developmental biology, but the role PCGF3 played in cancer development, especially the process of antitumor immunity, is unclear.

Fatty Acid Synthase (FASN), Dihydropyrimidinase Like 2 (DPYSL2) and Transforming Growth Factor Beta Induced (TGFBI) were all associated with adverse clinical outcomes in BLCA patients. As an essential lipogenic enzyme synthesizing the fatty acids de novo, FASN was involved in energy metabolism of BLCA. Many studies have reported the metabolic reprogramming of tumor cells could inhibit the energy intake of immune cells⁴¹. In addition, it has been found the fatty acids are able to induce the transformation of M1 macrophages to M2 macrophages, promoting tumor growth and metastasis⁴². Recent study revealed blockade of FASN helped CD4 effector T cells effectively avoid restimulation-induced cell death (RICD)⁴³. All the proof above displayed the close relationship between FASN and tumor immune response, while few studies concentrated on this aspect in BLCA for the moment.

DPYSL2, a member of dihydropyrimidinase-like protein family, which was acted as a regulator of the nervous system development, was firstly found to be an indicator for poor prognosis of BLCA. Some researchers^{44,45} have demonstrated the linkage between DPYSL2 and mTOR signaling pathway, and the activation of mTOR signaling pathway is crucial for T cell differentiation⁴⁶, suggesting the potential function of DPYSL2 in immunobiology of bladder cancer. TGFBI is an exocrine protein and could promote the malignant phenotypes of BLCA^{47,48}. Recent study⁴⁹ has identified TGFBI as an immune-relevant biomarker for prognosis of clear cell renal cell carcinoma based on multiple datasets, verifying our interesting findings. In all, combining the literature and the outcomes of our study, we believe PCGF3, FASN, DPYSL2 and TGFBI could serve as important prognostic factors and involve in the process of immune response in BLCA.

Several ncRNAs of the 24 miRNAs and 86 lncRNAs in our ceRNA network have been reported to be correlated with immune response and tumorigenesis. For example, miR-216a, which served as a predictor for poor prognosis of BLCA (**Figure. 6A**), could promote the proliferation of renal cell carcinoma cell⁵⁰. We also found HCP5, a lncRNA obtained from the target prediction of miR-216a, was correlated with immune checkpoints, which was reported by a recent literature⁵¹. Additionally, the previous studies offered us reliable experimental buttress to the prediction in our ceRNA network. The FASN-3'-UTR-luciferase assay was successfully performed to validate the ability of miR-150-5p to target FASN in the mice experiment⁵². MiR-21 was able to induce the tumor growth by targeting TGFBI in non-small cell lung cancer cells⁵³ and pancreatic cancer cells⁵⁴, demonstrating the relative stability of our predictions though experimental investigation in BLCA is demanded.

Importantly, we constructed a gene-based risk scoring model, which showed high accuracy of prediction for overall survival both in the TCGA-BLCA cohort and GSE13507 cohort. By means of immune grouping and correlation test, we ensure the risk score is closely correlated with tumor immunity. Compared with other immune-related risk models³⁴⁻³⁶, we firstly made use of ceRNA hypothesis to establish a prognostic signature in order to make sure the risk score was immune-related. The analysis results of functional annotation indicated that the immune-related pathways, such as complement and inflammatory response, were significantly enriched in the high-risk group, which once again verified the immunological correlation of risk score. We also explored the immune cell infiltration divergence between low- and high-risk group, and found the higher infiltration proportion of macrophages M0, M1 and M2 and lower infiltration ratio of memory B cells, activated dendritic cells and regulatory T cells in the high-risk cohort than that in the low-risk group. Multiple researches revealed that tumor-associated macrophages (TAMs) could positively regulate the tumor growth and affect the effectiveness of BCG instillation in BLCA^{55, 56}. Unexpectedly, M1 macrophages, which were regarded as anti-tumor cells, were found to be correlated with poor prognosis of BLCA in our study. It was reported that M1 macrophages could induce the expression of PD-L1 in hepatocellular carcinoma (HCC) cells, leading to the resistance to anti-tumor immunity⁵⁷. Our research also disclosed that the higher infiltrating level of M1 macrophages and higher expression level of PD-L1, which is also known as PDCD1, could be discovered in the high-risk group. Hence, there may be similar biological mechanisms in the tumorigenesis of BLCA, which needed to be further experimental validated. Furthermore, we revealed that the risk score is positively correlated with the immune score, stromal score and ESTIMATE score. Therefore, the cases in the high-risk group might have a significantly higher infiltration ratio of immune cells and stromal cells, along with the lower ratio of tumor cells, supporting the conclusion that the risk score could indicate the immune status of TME in patients with BLCA.

Meanwhile, the limitations of the present study are nonnegligible. First, PCGF3 and DPYSL2 were firstly reported to be prognostic biomarker and correlated with tumor immune response in BLCA patients, and further studies needed to explore the underlying mechanisms and stability of their predictive value through multicenter and large-scale researches. Second, we predict the possible targeted miRNAs and probable lncRNA competitors of DPYSL2, TGFBI, NTF3, PCGF3 and FASN, while few of the predictions have been verified in BLCA. Last, an immune-related scoring model has been constructed and validated primitively, and clinical trials ought to be performed to reconfirm the robustness.

To sum up, our study has successfully set up an immune-related risk signature, including DPYSL2, TGFBI, NTF3, PCGF3 and FASN, which showed high prognostic value in BLCA. Importantly, an immune-relevant ceRNA network, containing 5 mRNAs, 24 miRNAs and 86 lncRNAs, was established, leading to a deeper understanding of RNA interactions in the process of tumor immune response in BLCA.

Declarations

Funding

This study is supported by National Natural Science Foundation of China (NO. 81772257), Youth Cultivation Program of Southern Medical University (NO. PY2018N076) and Medical Scientific Research Foundation of Guangdong Province (NO. A2019557).

Availability of data and materials

The data supporting our findings could be downloaded from UCSC browser and GEO database.

Ethics approval and consent to participate

No applicable

Consent for publication

No applicable

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Cun-dong Liu designed the study. Ran-ran Zhou developed the algorithm and drafted the manuscript. Hu Tian downloaded and analyzed the data. Cheng Yang checked the results. Fan Peng, Xi Zhi, Hao-yu Yuan and Jun-hao Zhou revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank everyone who did help in this study.

References

1. Kulkarni GS, Black PC, Sridhar SS, et al. Canadian Urological Association guideline: Muscle-invasive bladder cancer. *Can Urol Assoc J.* 2019;230–238.
2. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer.* 2015;136(5):E359–86.
3. Massari F, Santoni M, di Nunno V, et al. Adjuvant and neoadjuvant approaches for urothelial cancer: Updated indications and controversies. *Cancer Treat Rev.* 2018;68:80–5.
4. Galsky MD, Stensland KD, Moshier E, et al. Effectiveness of Adjuvant Chemotherapy for Locally Advanced Bladder Cancer. *J Clin Oncol.* 2016;34(8):825–32.
5. Afonso J, Santos LL, Longatto-Filho A, et al. Competitive glucose metabolism as a target to boost bladder cancer immunotherapy. *Nat Rev Urol.* 2020;17(2):77–106.

6. Thoma C. Combining epigenetic and immune checkpoint inhibitors in bladder cancer. *Nat Rev Urol*. 2019;16(9):507.
7. Salmena L, Poliseno L, Tay Y, et al. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*. 2011;146(3):353–8.
8. Hong X, Liu N, Liang Y, et al. Circular RNA CRIM1 functions as a ceRNA to promote nasopharyngeal carcinoma metastasis and docetaxel chemoresistance through upregulating FOXQ1. *Mol Cancer*. 2020;19(1):33.
9. Wang W, Hu W, Wang Y, et al. Long non-coding RNA UCA1 promotes malignant phenotypes of renal cancer cells by modulating the miR-182-5p/DLL4 axis as a ceRNA. *Mol Cancer*. 2020;19(1):18.
10. Wang Z, Liu T, Li G, et al. The exploration of new therapeutic targets for HPV-negative head and neck squamous cell cancer through the construction of a ceRNA network and immune microenvironment analysis. *J Cell Biochem*. 2020;121(5–6):3426–37.
11. Huang R, Meng T, Chen R, et al. The construction and analysis of tumor-infiltrating immune cell and ceRNA networks in recurrent soft tissue sarcoma. *Aging*. 2019;11(22):10116–43.
12. Huang R, Wu J, Zheng Z, et al. The Construction and Analysis of ceRNA Network and Patterns of Immune Infiltration in Mesothelioma With Bone Metastasis. *Front Bioeng Biotechnol*. 2019;7:257.
13. Jiang J, Bi Y, Liu XP, et al. To construct a ceRNA regulatory network as prognostic biomarkers for bladder cancer. *J Cell Mol Med*. 2020;24(9):5375–86.
14. Haeussler M, Zweig AS, Tyner C, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res*. 2019;47(D1):D853–8.
15. Kim WJ, Kim EJ, Kim SK, et al. Predictive value of progression-related gene classifier in primary non-muscle invasive bladder cancer. *Mol Cancer*. 2010;9:3.
16. Bindea G, Mlecnik B, Tosolini M, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity*. 2013;39(4):782–95.
17. Shen Y, Peng X, Shen C. Identification and validation of immune-related lncRNA prognostic signature for breast cancer. *Genomics*. 2020;112(3):2640–6.
18. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013;14:7.
19. Yoshihara K, Shahmoradgoli M, Martinez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612.
20. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40.
21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
22. Yip AM, Horvath S. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics*. 2007;8:22.

23. Jeggari A, Marks DS, Larsson E. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics*. 2012;28(15):2062–3.
24. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res*. 2020;48(D1):D127–31.
25. Huang HY, Lin YC, Li J, et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res*. 2020;48(D1):D148–54.
26. Agarwal V, Bell GW, Nam JW, et al. Predicting effective microRNA target sites in mammalian mRNAs. *Elife*. 2015;4.
27. Li JH, Liu S, Zhou H, et al. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res*. 2014;42(Database issue):D92–7.
28. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res*. 2004;10(21):7252–9.
29. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–7.
30. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50.
31. Subramanian A, Kuehn H, Gould J, et al. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics*. 2007;23(23):3251–3.
32. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394–424.
33. Patel VG, Oh WK, Galsky MD. Treatment of muscle-invasive and advanced bladder cancer in 2020. *CA Cancer J Clin*. 2020.
34. Cao R, Yuan L, Ma B, et al. Immune-related long non-coding RNA signature identified prognosis and immunotherapeutic efficiency in bladder cancer (BLCA). *Cancer Cell Int*. 2020;20:276.
35. Jiang W, Zhu D, Wang C, et al. An immune relevant signature for predicting prognoses and immunotherapeutic responses in patients with muscle-invasive bladder cancer (MIBC). *Cancer Med*. 2020;9(8):2774–90.
36. Xu N, Ke ZB, Lin XD, et al. Development and validation of a molecular prognostic index of bladder cancer based on immunogenomic landscape analysis. *Cancer Cell Int*. 2020;20:302.
37. Li J, Liu C, Chen Y, et al. Tumor Characterization in Breast Cancer Identifies Immune-Relevant Gene Signatures Associated With Prognosis. *Front Genet*. 2019;10:1119.
38. Almeida M, Pintacuda G, Masui O, et al. PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science*. 2017;356(6342):1081–4.

39. Kitamura S, Tanahashi T, Aoyagi E, et al. Response Predictors of S-1, Cisplatin, and Docetaxel Combination Chemotherapy for Metastatic Gastric Cancer: Microarray Analysis of Whole Human Genes. *Oncology*. 2017;93(2):127–35.
40. Zhou Y, Wan C, Liu Y, et al. Polycomb group oncogene RING1 is over-expressed in non-small cell lung cancer. *Pathol Oncol Res*. 2014;20(3):549–56.
41. Ngwa VM, Edwards DN, Philip M, et al. Microenvironmental Metabolism Regulates Antitumor Immunity. *Cancer Res*. 2019;79(16):4003–8.
42. Nomura DK, Long JZ, Niessen S, et al. Monoacylglycerol lipase regulates a fatty acid network that promotes cancer pathogenesis. *Cell*. 2010;140(1):49–61.
43. Voss K, Luthers CR, Pohida K, et al. Fatty Acid Synthase Contributes to Restimulation-Induced Cell Death of Human CD4 T Cells. *Front Mol Biosci*. 2019;6:106.
44. Pham X, Song G, Lao S, et al. The DPYSL2 gene connects mTOR and schizophrenia. *Transl Psychiatry*. 2016;6(11):e933.
45. Liu Y, Pham X, Zhang L, et al. Functional variants in DPYSL2 sequence increase risk of schizophrenia and suggest a link to mTOR signaling. *G3. (Bethesda)*. 2014;5(1):61–72.
46. Yang K, Blanco DB, Chen X, et al. Metabolic signaling directs the reciprocal lineage decisions of alphabeta and gammadelta T cells. *Sci Immunol*. 2018;3:25.
47. Lang K, Kahveci S, Bonberg N, et al. TGFBI Protein Is Increased in the Urine of Patients with High-Grade Urothelial Carcinomas, and Promotes Cell Proliferation and Migration. *Int J Mol Sci*. 2019;20(18).
48. Zou J, Huang R, Li H, et al. Secreted TGF-beta-induced protein promotes aggressive progression in bladder cancer cells. *Cancer Manag Res*. 2019;11:6995–7006.
49. Du GW, Yan X, Chen Z, et al. Identification of transforming growth factor beta induced (TGFBI) as an immune-related prognostic factor in clear cell renal cell carcinoma (ccRCC). *Aging*. 2020;12(9):8484–505.
50. Chen P, Quan J, Jin L, et al. miR-216a-5p acts as an oncogene in renal cell carcinoma. *Exp Ther Med*. 2018;15(4):4039–46.
51. Xu S, Wang Q, Kang Y, et al. Long Noncoding RNAs Control the Modulation of Immune Checkpoint Molecules in Cancer. *Cancer Immunol Res*. 2020;8(7):937–51.
52. Heinz RE, Rudolph MC, Ramanathan P, et al. Constitutive expression of microRNA-150 in mammary epithelium suppresses secretory activation and impairs de novo lipogenesis. *Development*. 2016;143(22):4236–48.
53. Yan L, Ma J, Wang Y, et al. miR-21-5p induces cell proliferation by targeting TGFBI in non-small cell lung cancer cells. *Exp Ther Med*. 2018;16(6):4655–63.
54. Zhang Y, Zhang R, Ding X, et al. EFNB2 acts as the target of miR-557 to facilitate cell proliferation, migration and invasion in pancreatic ductal adenocarcinoma by bioinformatics analysis and verification. *Am J Transl Res*. 2018;10(11):3514–28.

55. Suriano F, Santini D, Perrone G, et al. Tumor associated macrophages polarization dictates the efficacy of BCG instillation in non-muscle invasive urothelial bladder cancer. *J Exp Clin Cancer Res.* 2013;32:87.
56. Qiu S, Deng L, Liao X, et al. Tumor-associated macrophages promote bladder tumor growth through PI3K/AKT signal induced by collagen. *Cancer Sci.* 2019;110(7):2110–8.
57. Zong Z, Zou J, Mao R, et al. M1 Macrophages Induce PD-L1 Expression in Hepatocellular Carcinoma Cells Through IL-1beta Signaling. *Front Immunol.* 2019;10:1643.

Tables

Table 1
Clinicopathologic characteristics of 398 BLCA patients from TCGA

Parameters	Type	Training dataset	Testing dataset	Entire dataset	p value
Age	<= 64	105(37.5%)	44(37.29%)	149(37.44%)	0.968
	> 64	175(62.5%)	74(62.71%)	249(62.56%)	
Gender	female	70(25%)	36(30.51%)	106(26.63%)	0.256
	male	210(75%)	82(69.49%)	292(73.37%)	
M stage	M0	130(46.43%)	58(49.15%)	188(47.24%)	0.636
	M1	6(2.14%)	4(3.39%)	10(2.51%)	
	unknown	144(51.43%)	56(47.46%)	200(50.25%)	
pN stage	N0	171(61.07%)	59(50%)	230(57.79%)	0.149
	N1	31(11.07%)	15(12.71%)	46(11.56%)	
	N2	45(16.07%)	30(25.42%)	75(18.84%)	
	N3	6(2.14%)	1(0.85%)	7(1.76%)	
	unknown	27(9.64%)	13(11.02%)	40(10.05%)	
Stage	Stage I	1(0.36%)	0(0%)	1(0.25%)	0.123
	Stage II	85(30.36%)	40(33.9%)	125(31.41%)	
	Stage III	107(38.21%)	31(26.27%)	138(34.67%)	
	Stage IV	85(30.36%)	47(39.83%)	132(33.17%)	
	unknown	2(0.71%)	0(0%)	2(0.5%)	
pT stage	T0	0(0%)	1(0.85%)	1(0.25%)	0.641
	T1	2(0.71%)	0(0%)	2(0.5%)	
	T2	79(28.21%)	35(29.66%)	114(28.64%)	
	T3	136(48.57%)	55(46.61%)	191(47.99%)	
	T4	41(14.64%)	17(14.41%)	58(14.57%)	
	unknown	22(7.86%)	10(8.47%)	32(8.04%)	
Histological type	Carcinoma	1(0.36%)	0(0%)	1(0.25%)	0.497
	Papillary adenocarcinoma	1(0.36%)	0(0%)	1(0.25%)	

Parameters	Type	Training dataset	Testing dataset	Entire dataset	p value
	Papillary transitional cell carcinoma	49(17.5%)	14(11.86%)	63(15.83%)	
	Squamous cell carcinoma	1(0.36%)	0(0%)	1(0.25%)	
	Transitional cell carcinoma	228(81.43%)	104(88.14%)	332(83.42%)	

Figures

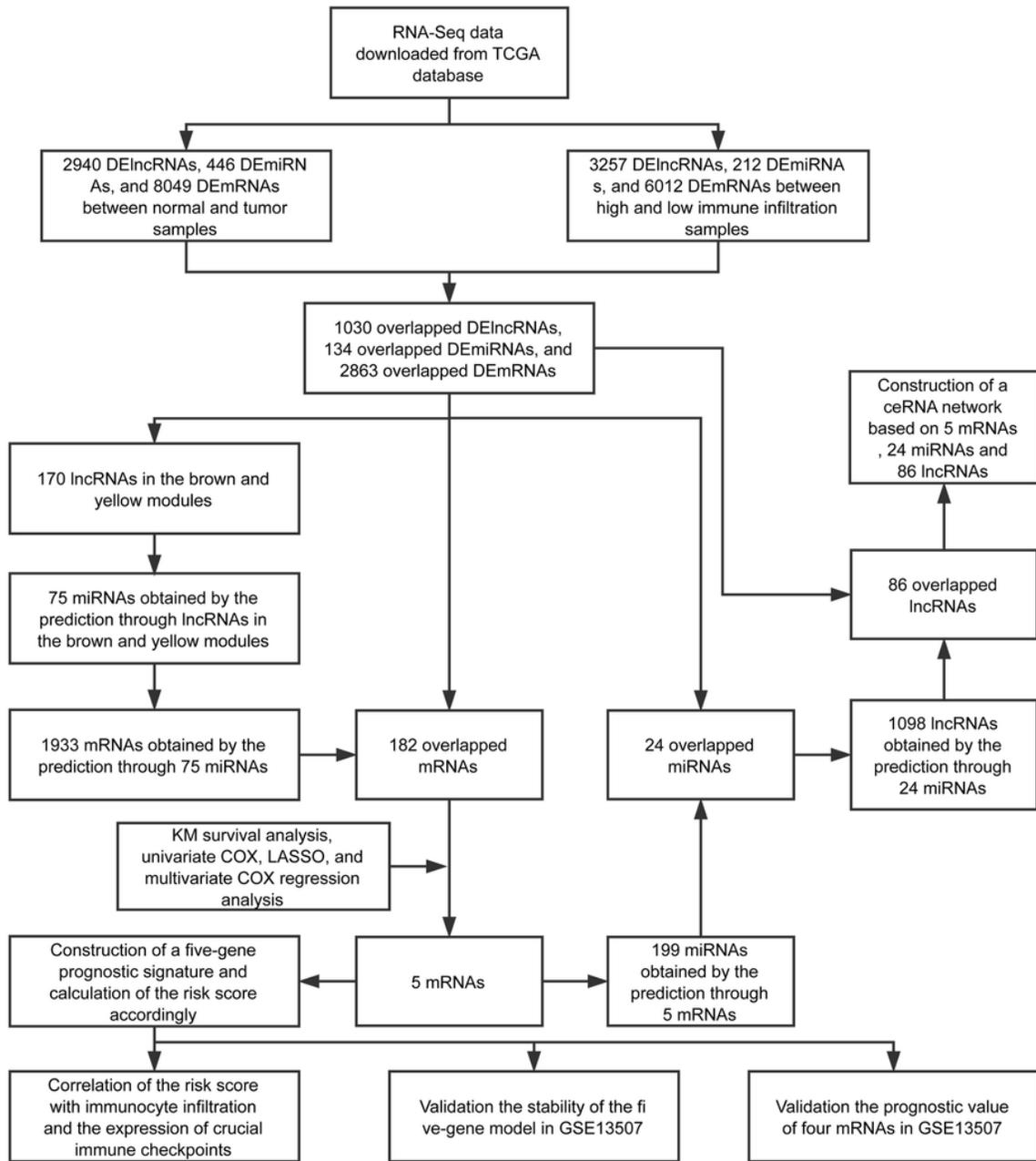


Figure 1

The workflow diagram of the study process.

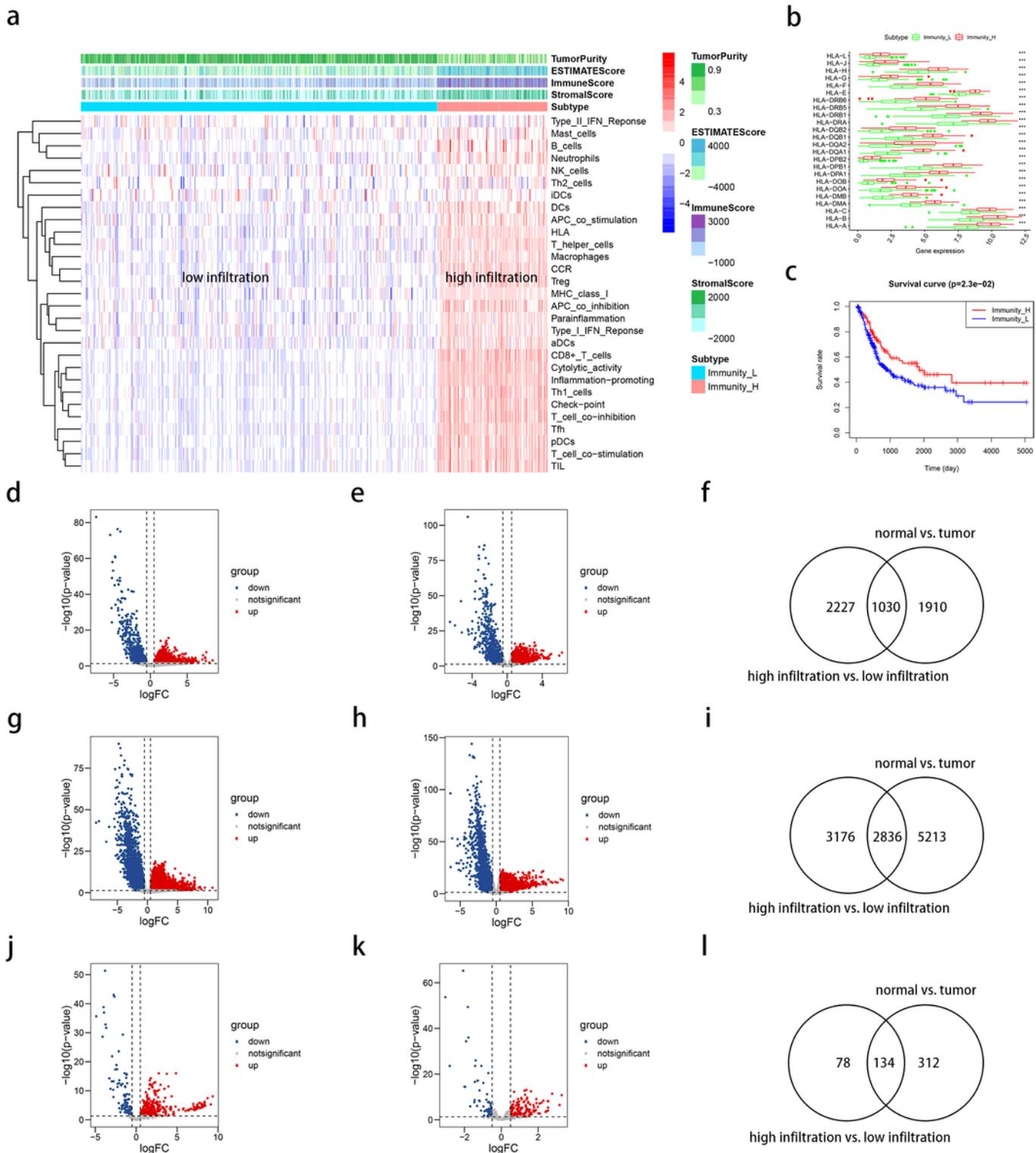


Figure 2

The immunogenomic landscape of the TME in BLCA patients. a. Unsupervised clustering of 29 immune signatures for 411 BLCA patients from the TCGA-BLCA cohort: high infiltration (red, n = 97) and low infiltration (blue, n = 314). Parameters including tumor purity, ESTIMATEScore, ImmuneScore and StromalScore are shown above the heatmap. b. Expression levels of HLA genes in the TCGA-BLCA cohort patients with a high-infiltrating TME (red) vs that in the patients with a low-infiltrating TME (blue). The

differences between the two groups were compared by the Wilcoxon test (***, $P < 0.001$). c. Kaplan-Meier curves of the overall survival of the TCGA-BLCA cohort patients with different infiltrating TME (high-infiltrating TME vs low-infiltrating TME, log-rank test, $P = 0.023$). d. 2940 DElncRNAs between normal and tumor samples. e. 3257 DElncRNAs between high-infiltrating and low-infiltrating samples. f. Venn plot showing 1030 overlapped DElncRNAs. g. 8049 DEmRNAs between normal and tumor samples. h. 6012 DEmRNAs between high-infiltrating and low-infiltrating samples. i. Venn plot showing 2836 overlapped DEmRNAs. j. 446 DEmiRNAs between normal and tumor samples. k. 212 DEmiRNAs between high-infiltrating and low-infiltrating samples. l. Venn plot showing 134 overlapped DEmiRNAs.

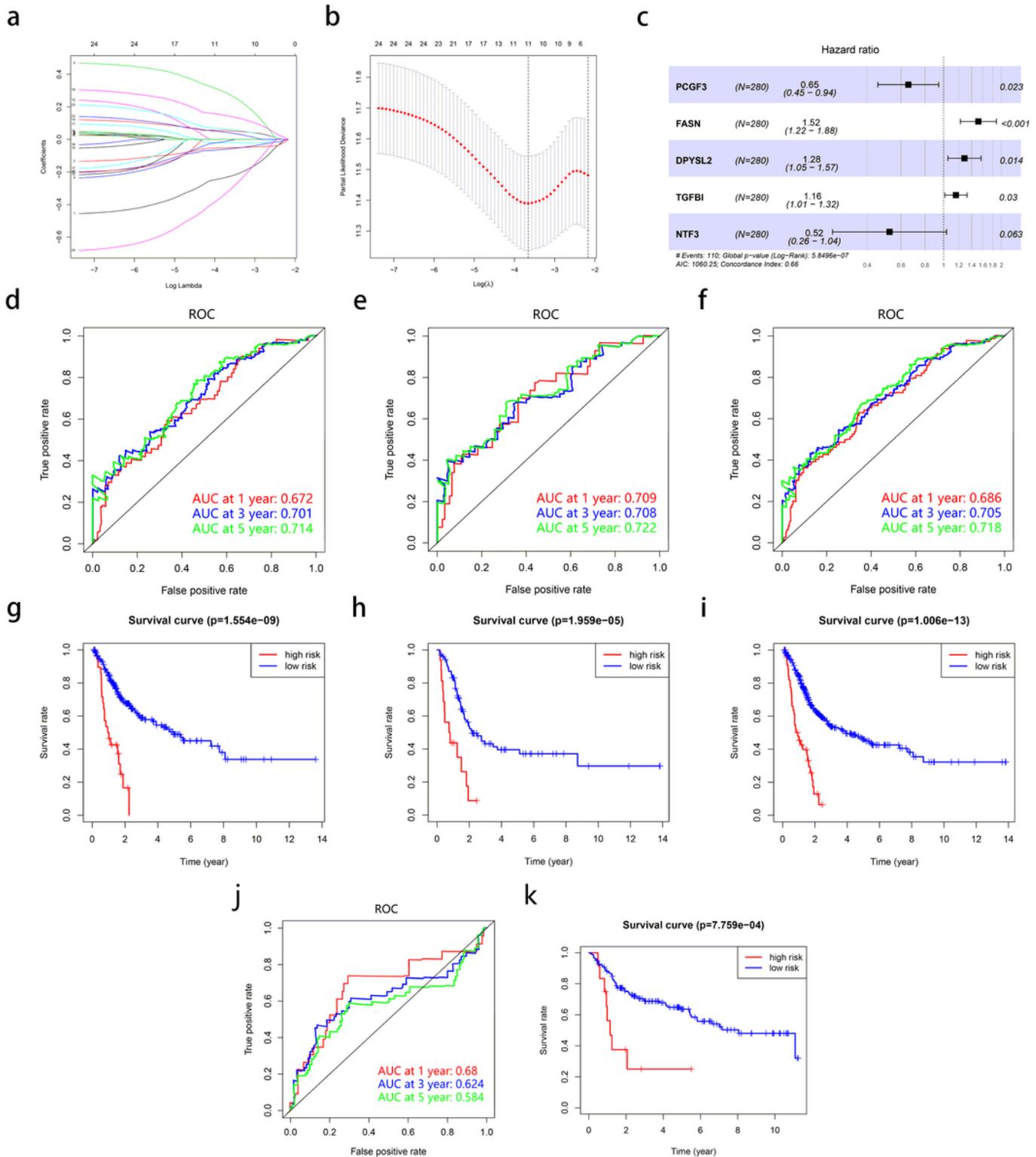


Figure 3

The risk score based on the five-gene signature is a promising predictor for prognosis of BLCA patients. a-b. LASSO Cox regression identified 11 genes most correlated with overall survival in the training dataset. c. Forest plot showing the hazard ratios (HRs) and 95% confidence intervals (CI) in multivariate Cox regression analysis. d-i. Validation the predictive stability of the five-gene signature in TCGA dataset. Time-dependent ROC curves for 1-, 3-, and 5-year OS predictions for the five-gene risk signature in TCGA-

BLCA training (d), validation (e) and entire (f) cohort. Kaplan-Meier survival curves for patients assigned to high- and low-risk group in TCGA-BLCA training (g), validation (h) and entire (i) cohort. 2.05 is regarded as the optimal cut-off value via X-tile software. j-k. Time-dependent ROC curve (j) and Kaplan-Meier survival curve (k) shows the 5-gene risk signature can predict the prognosis of patients with BLCA from GSE13507.

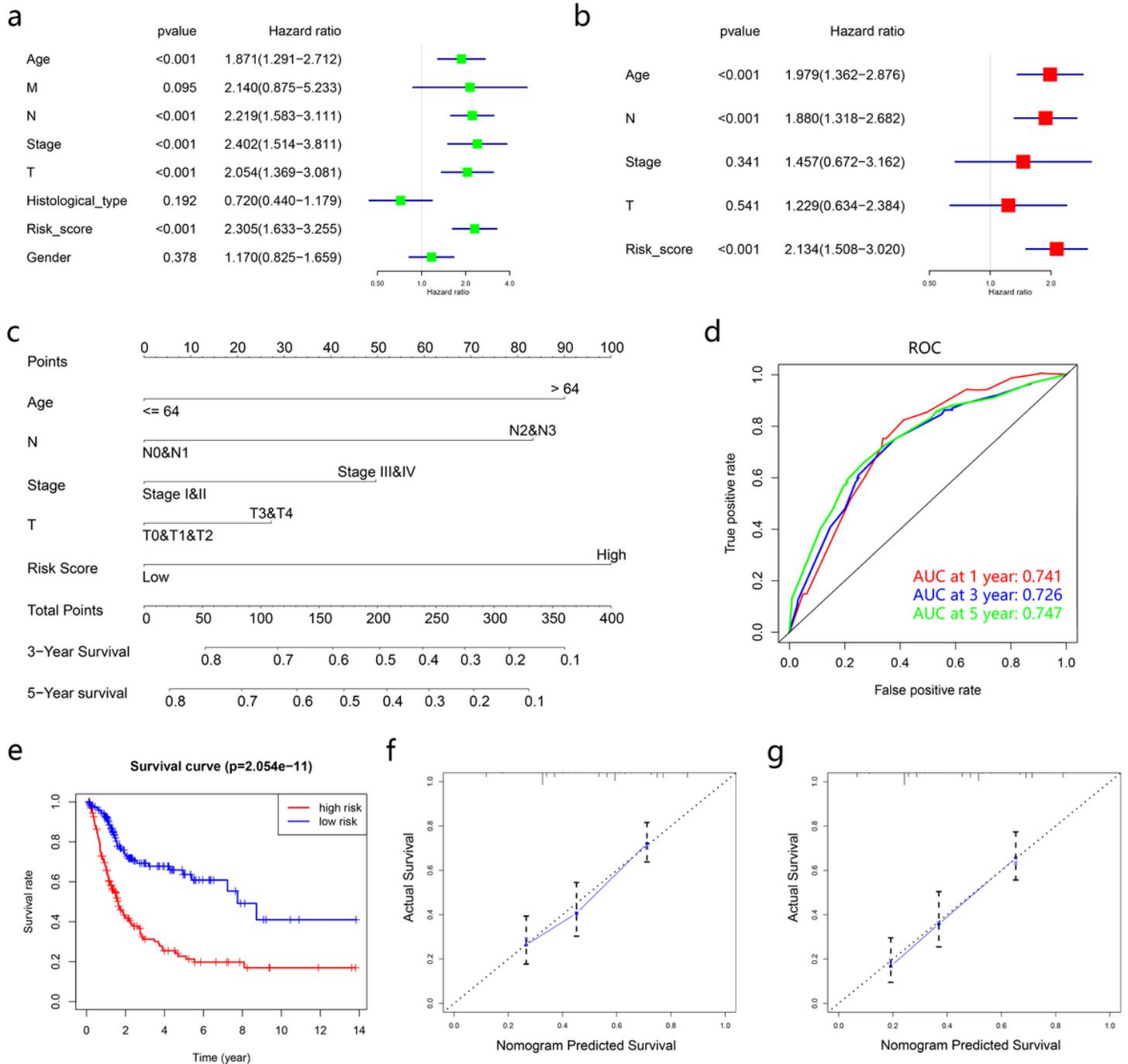


Figure 4

Development of a nomogram based on the five-gene signature to predict overall survival (OS) of patients with BLCA. a. Univariable Cox regression analysis for the entire TCGA-BLCA cohort. b. Multivariable Cox regression analysis for the entire TCGA-BLCA cohort. c. A nomogram including risk score and other

clinical features for predicting 3- and 5-years OS for BLCA. d. Time-dependent ROC curve for 1-, 3-, and 5-year OS predictions for the nomogram. e. Kaplan-Meier survival curve for the patients with BLCA. The cohorts were stratified at median cut-off of the points calculated by nomogram to form high-risk and low-risk groups. Calibration plot of nomogram for predicting probabilities of 3- (f) and 5-year (g) OS of TCGA-BLCA patients.

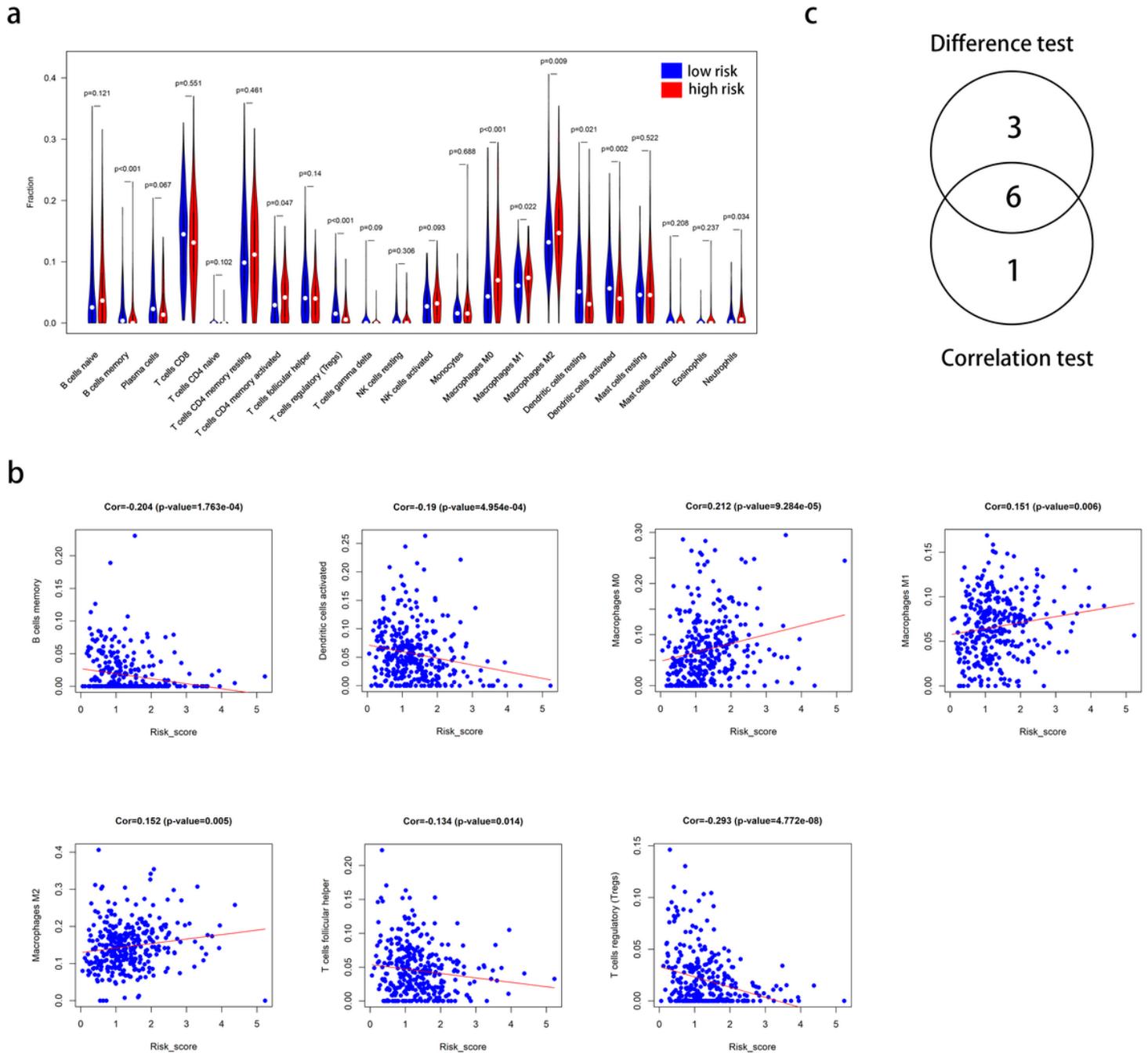


Figure 5

Correlation of tumor-infiltrating immune cells (TICs) proportion with risk score. a. Violin plot displaying the ratio divergence of 22 types of immune cells between BLCA samples with low or high risk score relative to the median of risk score, and Wilcoxon rank sum was used for the significance test. b. Scatter

plot showing the correlation of 7 kinds of TICs proportion with risk score ($p < 0.05$). Pearson coefficient was used for the correlation test. c. Venn plot revealing six kinds of TICs correlated with risk score codetermined by difference test and correlation test displayed in violin and scatter plots, respectively.

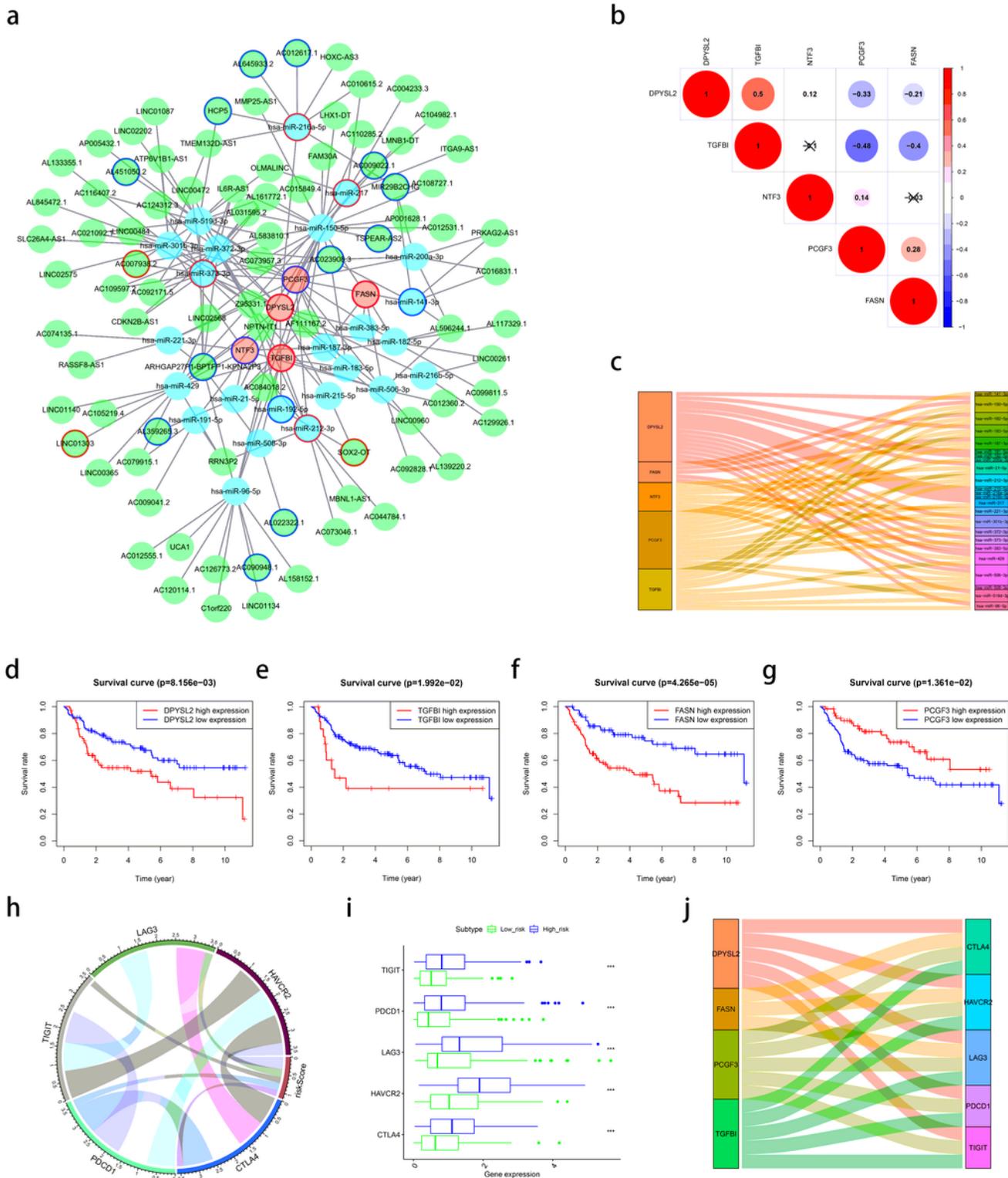


Figure 6

Construction of a ceRNA network. a. A ceRNA network including 5 mRNAs, 24 miRNAs and 86 lncRNAs. Red nodes represent mRNAs, blue nodes represent miRNAs, and green nodes represent lncRNAs. The

nodes with red border mean the genes with HR >1, and the nodes with blue border mean the genes with HR <1 (p < 0.05). b. The Pearson correlation of five mRNAs expression levels. c. Sankey diagram showing the linkage among five mRNAs and 24 miRNAs. d-g. Kaplan-Meier survival curves showing the prognostic value of DPYSL2 (d), TGFBI (e), FASN (f) and PCGF3 (g) in BLCA cohorts from GSE13507. h. Chord diagram displaying the correlation between the risk score based on 5 mRNAs and immune-checkpoint-related genes including PDCD1, CTLA4, HAVCR2, LAG3 and TIGIT. i. Boxplots indicating the expression levels of five crucial immune-checkpoint-relevant genes in high-risk group are significantly higher than those in low-risk group. The differences between the two groups were compared by the Wilcoxon test (***, P = 0). j. The correlation between four mRNAs and crucial immune checkpoints. The correlation with R > 0.3 and p < 0.05 was considered statistically significant.

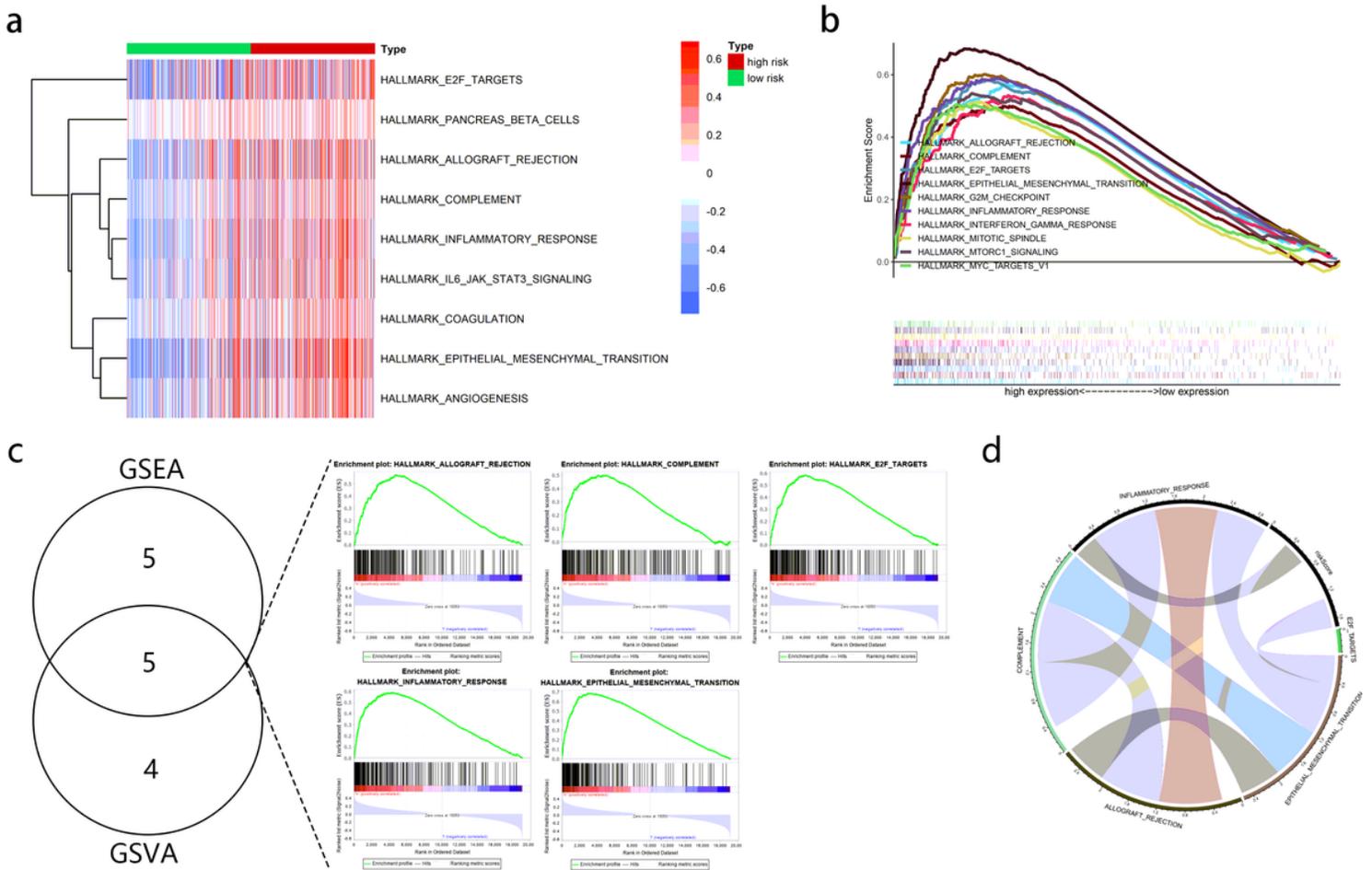


Figure 7

GSVA and GSEA for TCGA-BLCA samples with high and low risk score. a. The GSVA analysis shows 9 pathways have the relatively close correlation with risk score. b. The most related Top 10 pathways from GSEA analysis. A permutation of 1000 times were conducted. c. The 5 intersection pathways between GSEA and GSVA, including HALLMARK_COMPLEMENT, HALLMARK_E2F_TARGETS, HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION, HALLMARK_INFLAMMATORY_RESPONSE and HALLMARK_ALLOGRAFT_REJECTION. D. Chord diagram showing the correlation between risk score and

activation levels of the five interaction pathways ($p < 0.001$). Pearson coefficient was used for significance test.

Supplementary Files

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

- [Additionalfile8Table.S4.docx](#)
- [Additionalfile7Table.S3.docx](#)
- [Additionalfile6Table.S2.docx](#)
- [Additionalfile5TableS1.docx](#)
- [Additionalfile4.xls](#)
- [Additionalfile3Fig.S3.tif](#)
- [Additionalfile2Fig.S2.tif](#)
- [Additionalfile1Fig.S1.tif](#)