

A stemness-related seven-gene signature for predicting prognosis and treatment response in muscle invasive bladder cancer

Maolang Tian

Department of Head and Neck Oncology and Department of Radiation Oncology, Cancer Center and State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu 610041, China

Jinlan He

Department of Head and Neck Oncology and Department of Radiation Oncology, Cancer Center and State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu 610041, China

Jiaqi Han

Department of Head and Neck Oncology and Department of Radiation Oncology, Cancer Center and State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu 610041, China

Hong Zhu (✉ zhuhonghuaxi@163.com)

Sichuan University West China Hospital <https://orcid.org/0000-0002-9754-8544>

Research

Keywords: Muscle invasive bladder cancer, Overall survival, Gene expression signature, Cancer stem cell, Nomogram.

Posted Date: May 6th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-25781/v1>

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

[Read Full License](#)

Abstract

Background: Muscle invasive bladder cancer (MIBC) is an aggressive cancer characterized by therapeutic resistance and poor prognosis, which are possibly due to the existence of cancer stem cells (CSCs). In this study, we aimed to characterize the expression of cancer stemness-related genes and develop a multi-gene risk signature to predict clinical outcome and treatment response in MIBC.

Methods: The mRNA expression data and clinical data of MIBC patients were collected from The Cancer Genome Atlas (TCGA) database and Gene Expression Omnibus (GEO) database, which included the TCGA training cohort (n = 333) and three GEO validation cohorts, GSE13507 (n = 165), GSE32548 (n = 127), and GSE48075 (n = 72). A list of 166 stemness-related genes were obtained from the Cancer Single Cell State Atlas (CancerSEA) database and prognostic genes for overall survival (OS) were identified by univariate Cox analysis. Then, the least absolute shrinkage and selection operator (LASSO) regression and stepwise multivariate Cox regression were performed to generate a multi-gene risk signature. Kaplan-Meier curve, time-dependent receiver operating characteristic (ROC) curve, multivariate analysis, and stratification analysis were used to evaluate the performance of the gene signature. We also explored the relationship between risk score and response to chemotherapy and radiotherapy in MIBC patients. Moreover, independent prognostic factors for OS were combined together into a nomogram to improve predictive performance.

Results: Firstly, a total of 25 prognostic genes were identified. Then, a seven-gene risk signature (EGFR, FOXA2, HES1, MME, RBM6, SMOC2, and TFRC) was constructed and it could robustly classify MIBC patients into high -risk and low-risk groups with different clinical outcomes. ROC curves showed that the seven-gene signature had a robust predictive accuracy in four cohorts. Besides, high risk score was significantly associated with advanced clinical stage and treatment failure. As an independent risk factor for OS, the stemness-related seven-gene signature could achieve better prognostic accuracy when integrated with clinical factors.

Conclusions: We developed and validated a robust stemness-related gene signature which could robustly predicate clinical outcome and shed light on the cancer stemness in bladder cancer.

Background

Bladder cancer (BC) is one of the most common genitourinary malignancies, with an estimated 549,000 new cases and 200,000 deaths worldwide in 2018 [1]. Currently, BC is categorized into two major pathological types: non-muscle-invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC). Nearly 80% of newly diagnosed patients present with NMIBC and the remaining 20% of patients present with MIBC [2]. Patients with NMIBC generally have good prognosis with 5-year survival rate above 90% [3]. On contrast, patients with MIBC have a less favorable prognosis with 5-year survival less than 50%, which is mainly due to disease progression, metastasis, and treatment failure [3]. About half of the MIBC patients are with occult metastasis when diagnosed [4]. Moreover, MIBCs are usually chemo-

resistant, especially for recurrent diseases [5]. Accumulating evidences suggest that bladder cancer stem cells (BCSCs) are possibly responsible for tumor heterogeneity, therapeutic resistance, and metastasis in bladder cancer.

Cancer stem cells (CSCs) are a small population of cancer cells with stem-cell like properties, including an undifferentiated state, resistance to cell apoptosis, and self-renewal properties [6]. CSCs are intimately correlated with tumor heterogeneity as CSCs are able to generate heterogeneous lineages of cancer cells [6]. BCSCs were first identified using the markers of normal stem cells [7]. BCSCs are thought to exist mainly in MIBC as well as the late stages during tumor development [8]. Biological mechanisms of CSCs in NMIBC and MIBC are distinctly different. NMIBC might originate from BCSCs with HRAS or FGFR3 mutations, whereas MIBC might be initiated by BCSCs with p53/Rb/PTEN gene mutations [9]. Up to date, a variety of potential markers have been identified for BCSCs, which include CD44 [5], EMA[10], and ALDH1A1 [11]. As BCSCs are spared from systematic chemotherapy and radiotherapy, CSCs are responsible for treatment failure, recurrence, and metastasis and might serve as potential targets for anticancer therapies. However, the underlying mechanism of CSCs in bladder cancer remains largely inconclusive. Therefore, identification of potential biomarkers associated with BCSCs could help to shed light on the properties of BCSCs and provide therapeutic targets for clinical management.

Recently, the rapid development of RNA sequencing (RNA-seq) technologies including single cell RNA-seq and mRNA-seq have contributed to a deeper understanding of tumor heterogeneity at single cell levels or genetic levels. Yang et al. conducted single-cell RNAseq on 59 cancer cells from three BC specimens and confirmed clonal homogeneity of BCSCs and identified 21 key altered genes in BCSCs [12]. The Cancer Single Cell State Atlas (CancerSEA) database, which store and reanalyze the published single cell RNA-seq data of cancers, has been established so as to explore the functional heterogeneity of cancer cells at single-cell resolution [13]. Up to date, this database has summarized critical genes related with 14 functional states (e.g. metastasis, stemness, hypoxia) by reanalyzing the sequencing data of 41,900 cancer cells from 25 cancer types.

In the present study, we aimed to utilize cancer stemness-related genes from CancerSEA database to develop a prognostic multi-gene risk signature in MIBC. With bioinformatics methods and public transcriptome data, we established a novel seven-gene signature, which is superior to TNM staging method in prognosis prediction. As an independent risk factor for overall survival, the seven-gene signature could robustly predict prognosis in MIBC patients. By emphasizing the genetic biomarkers of cancer stemness, our findings might provide new insights into cancer stem cells in bladder cancer.

Methods

Collection of mRNA expression data and stemness-related genes

TCGA BLCA (bladder cancer) project which contained mRNA sequencing data and clinical data of 411 patients were obtained from The Cancer Genome Atlas (TCGA) data portal (<https://portal.gdc.cancer.gov/>). Qualified samples for training cohort were screened as follows: (1) a

pathological diagnosis of MIBC. (2) intact clinical features and corresponding expression profiles. (3) intact follow up time (over 30 days). As information of M stage was absent in nearly half (n = 200) of the 411 patients, metastasis status of patients was not taken into consideration. After filtration, 333 samples with intact clinical features including age, gender, overall stage, T stage, N stage were screened out and used as the training cohort. To assess the robustness of our results, GSE13507 (n = 165), GSE32548 (n = 127) and GSE48075 (n = 72) were downloaded from Gene Expression Omnibus (GEO) database and samples with intact follow up time (over 30 days) were used for external validation. GSE13507 was based on GPL6102 platform while GSE32548 and GSE48075 were based on GPL6947 platform. Patients in GSE48075 were all MIBCs while GSE13507 and GSE32548 contained patients with both MIBCs and NMIBCs. We kept NMIBC samples for external validation due to the limited sample size of MIBC samples in two cohorts. The median overall survival (OS) of patients in TCGA training cohort, GSE13507, GSE32548 and GSE48075 was 15.6 (1.2 - 168.3) months, 36.57 (1.03 - 137.00) months, 53.80 (1.08 - 110.3) months, and 28.71 (1.50 - 179.20) months respectively. The detailed clinical characteristics of patients in four cohorts were shown in Table 1 and Additional file 1: Table S1. The comprehensive list of 166 cancer stemness-related genes was downloaded from the CancerSEA database (<http://biocc.hrbmu.edu.cn/CancerSEA/>).

Table 1 Clinical characteristics of the training cohorts.

Variable	Category	Number	Percentage
Age	≤65	128	38.44%
	>65	205	61.56%
Gender	Female	87	26.13%
	Male	246	73.87%
Clinical stage	Stage II	83	24.92%
	Stage III	125	37.54%
	Stage IV	125	37.54%
T stage	T2	98	29.43%
	T3	181	54.35%
	T4	54	16.22%
N stage	N0	211	63.36%
	N1	43	12.91%
	N2	73	21.92%
	N3	6	1.80%
Survival status	Dead	130	39.04%
	Alive	203	60.96%

Development of stemness-related gene signature

To develop a multi-gene risk signature, we firstly conducted univariate Cox survival analysis to identify the prognostic stemness-related genes for the overall survival (OS) of MIBC patients in training cohort.

Then, we used two steps of regression analysis to select out the optimal gene signature, which were the least absolute shrinkage and selector operation (LASSO) regression and stepwise multivariate Cox regression. In LASSO regression, the penalty parameter “lambda” was adjusted by 100000 times cross validations [14]. Then a risk model was generated using a linear combination of expression values and Cox regression coefficients of signature genes. Patients were categorized into high-and low-risk group by the median risk score. The survival differences between two groups was plotted in Kaplan-Meier survival curve and compared by log-rank test. We also compared prognostic values of stemness-related gene signature and clinical factors using time-dependent receiver operating characteristic (ROC) curve and area under the curve (AUC).

Independence of stemness-related gene signature

To assess whether the stemness-related gene signature was independently associated with OS, univariate as well as further multivariate Cox proportional hazards regression analyses were performed. Age was divided into two subgroups (≤ 65 years old and > 65 years old). The risk score of stemness-related gene signature was used as a continuous variable while clinical parameters including age, gender, and clinical stage were used as categorical variables. We further performed stratification analysis to test the prognostic significance of stemness-related gene signature in different clinical subgroups.

Validation of stemness-related gene signature

To determine the robustness of the stemness-related gene signature, we tested it in three GEO validation cohorts. Risk scores of patients in these cohorts were also generated using a linear combination of expression values and regression coefficients of signature genes. We used the median risk score in training cohort as a cut-off value for classifying low-risk and high-risk group. The performance of the gene signature in validation cohorts were also evaluated by Kaplan-Meier curve and Time-dependent ROC curve.

Association between stemness-related risk signature and clinical features

We further analyzed the correlation between risk score and clinical features including age, gender, overall stage, T stage, and N stage. In TCGA training cohort, treatment response to primary chemotherapy and radiotherapy of 185 patients were recorded, which included 121 patients with complete response (CR), 12 patients with partial response (PR), 20 patients with stable disease (SD), and 32 patients with progression disease (PD). Therefore, we also investigated whether the gene signature-based risk score could predict treatment response in MIBC patients.

Comparison between stemness-related gene signature and mRNA stemness indices

In 2018, Malta et al. put forward a novel mRNA expression-based stemness indices (mRNAsi), which ranges from 0 to 1, to measure cancer stemness and tumor purity [15]. They applied one-class logistic regression (OCLR) machine learning algorithm on pluripotent stem cell samples from the Progenitor Cell Biology Consortium dataset and developed a mRNA expression-based stemness gene signature, which

comprises 11774 genes. With the 11774-gene stemness signature and mRNA sequencing data, Malta et al. computed the mRNAsi of various tumor samples in TCGA and concluded that higher mRNAsi was closely associated with poor clinical outcomes in most cancer types. Therefore, we compared our stemness-related gene signature with mRNAsi in TCGA training cohort. The prognostic value of mRNAsi was evaluated by Kaplan-Meier survival curve and time-dependent ROC curve. We also explored the correlation between the mRNAsi and our stemness-related gene signature.

Integrated nomogram by combining the stemness-related gene signature with clinical factors

In order to improve the predicative accuracy for OS, a prognostic nomogram was established basing on the independent prognostic factors. A stepwise multivariate Cox regression method was used to determine the weight score and coefficients of each prognostic factor. The calibration curve was used to evaluate the differences in 1-, 3-, 5-, 10-years survival possibilities between nomogram prediction and the actual observation. The time-dependent ROC curve and Harrell's concordance index (C-index) were utilized to assess the predicative accuracy of nomogram.

Gene Set Enrichment Analysis

In TCGA training cohort, Gene Set Enrichment Analysis (GSEA) was performed using GSEA 4.0.1 software and referring to the Molecular Signatures Database (MSigDB) so as to identify the significantly enriched biological processes in high-risk and low-risk groups. Gene sets with a normalized p value < 0.05 and a false discovery rate (FDR) q value < 0.05 were considered to be significantly enriched. The most enriched biological processes were selected out by ranking of the normalized enrichment score (NES).

Statistical analysis

All statistical analyses in our study were performed using R language, version 3.6.0 and Perl language, version 5.28.2. R packages used in our study are as follows: (1) "limma" R package (for normalization of TCGA data). (2) "survival" and "survminer" R package (for univariate Cox analysis, drawing Kaplan-Meier curve, and calculating Hazard ratio and C-index). (3) "glmnet" R package (for LASSO Cox regression analysis). (3) "timeROC" R package (for plotting time-dependent ROC curve). (4) "beeswarm" R package (for clinical correlation analysis). Survival analyses were performed using the Kaplan-Meier method and compared using a log-rank test by the 'survival' package. Differences of risk scores among different variables were tested using two samples t-test. A p value < 0.05 was considered statistically significant in our study.

Results

Development of stemness-related gene signature

Firstly, a total of 25 prognostic genes were identified by univariate survival analysis (Additional file 2: Table S2). Among these genes, six genes were regarded as favorable genes for OS (HR < 1.00) while other 19 genes were unfavorable genes (HR > 1.00). With LASSO regression (Fig. 1a/b) and stepwise

multivariate Cox regression, an optimal seven-gene risk signature (Table 2) was constructed, which included five unfavorable genes (EGFR, FOXA2, MME, SMOC2, and TFRC) and two favorable genes (HES1, RBM6). The risk score of each patient was generated as follow: risk score = (EGFR*0.988) + (FOXA2*1.721) + (HES1*-1.817) + (MME*0.797) + (RBM6*-3.190) + (SMOC 2*0.916) + (TFRC*1.711). In training cohort, 333 patients were classified into high-risk group (n = 167) and low-risk group (n = 166) by the median risk score 1.051. Kaplan-Meier curve exhibited a significant difference in OS between the two groups as high-risk group of patients had poorer clinical outcome than low-risk group (p < 0.001, Fig. 1c). The time-dependent ROC curve suggested that the AUCs of the stemness-related signature were 0.673, 0.732, 0.760 at 1-year, 3-years, and 5-years respectively (Fig. 1d). In addition, the seven-gene signature had better sensitivity and specificity than clinical factors in predicting 1-year, 3-years, and 5-years OS in MIBC patients (Additional file 3: Figure S1).

Table 2 The stemness-related seven-gene risk signature.

Gene	Coefficient	HR	HR.95L	HR.95H	p
EGFR	0.988	4.677	1.778	12.302	0.002
FOXA2	1.721	2.873	1.013	8.146	0.047
HES1	-1.817	0.155	0.044	0.548	0.004
MME	0.797	2.993	1.383	6.477	0.005
RBM6	-3.190	0.192	0.047	0.780	0.021
SMOC2	0.916	2.280	1.081	4.809	0.030
TFRC	1.711	6.038	1.423	25.625	0.015

HR: hazard ratio. HR.95L/H: lower /upper limit of the 95% confidence interval of hazard ratio.

Independence of stemness-related gene signature

In univariate Cox regression analysis, the seven-gene risk signature was a strong risk factor for OS (Hazard ration (HR) = 1.919, p < 0.001, Table 3). Clinical factors including age (HR = 1.962, p < 0.001), overall stage (HR = 1.958, p < 0.001), T stage (HR = 1.652, p < 0.001), and N stage (HR = 1.589, p < 0.001) were also risk factors for OS (Table 3). Moreover, the seven-gene risk signature (HR = 1.925, p < 0.001, Table 3) was an independent prognostic indicator for OS in MIBC patients when adjusted for age, gender, and clinical stage. In training cohort, Kaplan-Meier survival analysis was conducted for patients stratified by age, gender, overall stage, T stage, and N stage. High-risk group of patients always suffered worse prognosis when compared with their low-risk counterparts in all clinical subgroups and the results were significant (Fig. 2). Taken together, the novel gene signature could independently predict OS in MIBC patients.

Table 3 Univariate and multivariate Cox regression analysis of the gene signature and clinical factors in training cohort.

Variable	Univariate analysis		Multivariate analysis	
	HR (95%CL)	<i>p</i>	HR (95%CL)	<i>p</i>
age	1.962 (1.326–2.901)	<0.001	1.969 (1.327–2.922)	<0.001
gender	0.871 (0.596–1.271)	0.473	0.922 (0.627–1.355)	0.677
stage	1.958 (1.526–2.513)	<0.001	1.582 (1.009–2.481)	0.046
T	1.652 (1.268–2.152)	<0.001	1.164 (0.839–1.615)	0.363
N	1.589 (1.330–1.899)	<0.001	1.140 (0.830–1.567)	0.418
Risk score	1.919 (1.593–2.312)	<0.001	1.925 (1.564–2.370)	<0.001

HR: hazard ratio, CL: confidence interval.

Validation of stemness-related gene signature

We further verified the performance of the seven-gene signature in three validation cohorts. Patients in GSE13507 (high risk = 69, low risk = 96) and GSE32548 (high risk = 62, low risk = 65) were categorized by the cut-off value 1.051 in training cohort. While in GSE48075, only 6 patients were low-risk and the remaining 66 patients were all high-risk according to the cut-off value 1.051. Therefore, we reassigned the 72 patients in GSE48075 into equal groups (n = 36, 36 respectively) by their own median risk score 5.850. Kaplan-Meier survival curves showed that high-risk patients also suffered reduced overall survival when compared with low-risk patients, which was observed in GSE13507 ($p < 0.001$, Fig. 3a), GSE32548 ($p < 0.001$, Fig. 3b) and GSE48075 ($p < 0.001$, Fig. 3c). In GSE13507, the AUCs of seven-gene signature were 0.732, 0.645, 0.629 at 1, 3, 5 years, respectively (Fig. 3d). In GSE32548, the AUCs of seven-gene signature were 0.773, 0.803, 0.709 at 1, 3, 5 years respectively (Fig. 3e). In GSE48075, the AUCs at 1, 3, 5 years were 0.729, 0.694, 0.608 respectively (Fig. 3f).

Association between stemness-related gene signature and clinical features

The relationship between the stemness-related gene signature and clinical factors was analyzed. As expected, the risk score was significantly elevated along with cancer progression, including advanced T stage ($p < 0.001$, Fig. 4a), N stage ($p = 0.018$, Fig. 4b), and overall stage ($p < 0.001$, Fig. 4c). However, no difference in risk score was observed between different subgroups of age and gender. Moreover, the seven-gene risk signature was closely linked to response to primary chemotherapy and radiotherapy as the risk score was sequentially increasing in patients with CR, PR, SD, and PD ($p = 0.002$, Fig. 4d). These findings suggested that high risk score based on seven-gene signature was generally associated with unfavorable clinical manifestations.

Comparison between stemness-related gene signature and mRNA stemness indices

The mRNAsi of TCGA bladder cancer patients were obtained from Malta et al.'s study. We matched the samples between clinical data and mRNAsi by sample id number. As a result, a total of 401 patients were provided with mRNAsi. We used the median mRNAsi to divide 401 patients into two groups. Kaplan-Meier survival curve showed that patients with higher mRNAsi had significantly poor long-term survival than patients with lower mRNAsi ($p < 0.001$, Fig. 4e). The time-dependent ROC curve demonstrated that

the AUCs of mRNAsi were 0.559, 0.548, 0.597 at 1-year, 3-years, and 5-years respectively (Fig. 4f). Furthermore, we also identified a significant correlation between mRNAsi and our seven-gene signature. The high-risk group patients identified by our seven-gene signature had significantly higher mRNAsi (Fig. 4g).

Integrated nomogram by combining the stemness-related gene signature with clinical factors

We subsequently developed a nomogram in training cohort to achieve better predictive performance (Fig. 5a). Three independent risk factors for OS including age (HR = 1.969, $p < 0.001$), overall stage (HR = 1.582, $p = 0.046$), and the seven-gene risk signature (HR = 1.925, $p < 0.001$) were included in the nomogram (Table 3). No collinearity was observed between three risk factors. The AUCs of the nomogram were 0.761, 0.785, 0.776 at 1-, 3-, 5-years respectively (Fig. 5b). The calibration plot exhibited high degree of agreement between predicted and actual survival probabilities (Fig. 5c-e).

Gene Set Enrichment Analysis

Using GSEA, we explored the signaling pathways which were associated with our gene signature-based risk classification. Under the criteria of $p < 0.05$ and $q < 0.05$, a total of 1825 biological processes were significantly altered in high risk group. The most altered biological processes included extracellular structure organization (NES = 2.400), wound healing (NES = 2.390), cell matrix adhesion (NES = 2.367), ossification (NES = 2.340), angiogenesis (NES = 2.333), cartilage development (NES = 2.328), cell substrate adhesion (NES = 2.325), connective tissue development (NES = 2.318), vasculature development (NES = 2.318), and regulation of cell shape (NES = 2.317), which were dominantly related with extracellular adhesion and angiogenesis (Figure 6; Additional file 4: Table S3).

Discussion

Bladder cancer is one of the most common malignancies in urinary system. MIBC, which account for about 20% of BC, presents with frequent metastasis and dismal prognosis. Accumulating evidences suggested that CSCs are partly responsible for the aggressive biological behavior and treatment failure and could be potential therapeutic targets. Therefore, a better understanding of CSCs in bladder cancer could help to provide insights into the mechanisms responsible for metastasis and treatment resistance as well as identifying novel therapeutic targets.

Recently, the availability of public, large-scale sequencing datasets has made it possible to identify gene expression signatures associated with patient's clinical outcome. In BC, researchers have reported several gene expression signatures which could reliably estimate survival. Abudurexiti et al. made a meta-analysis of published gene signature studies in BC patients and collected 274 candidate genes to develop a 12-gene signature to predict OS in MIBC patients [16]. Cao et al. developed a seven-gene signature related with epithelial-mesenchymal transition (EMT) process, which could stratify patients into different risk groups with distinct prognosis [17]. Jiang et al. analyzed genes associated with immune cells infiltration in MIBC and developed a nine-gene signature which could distinguish patients with different

prognoses and immunotherapeutic response [18]. However, genetic biomarkers associated with cancer stemness features were seldom studied in bladder cancer.

In the present study, we developed and validated a novel stemness-related seven-gene signature for predicting prognosis and treatment response in MIBC patients, which is different from previously reported gene signatures. Firstly, the 166 stemness-related genes in CancerSEA database were derived from single cell RNAseq data. Therefore, the resource of candidate genes is novel and reliable. Secondly, we used one TCGA patient cohorts to construct the gene signature and three GEO cohorts to verify our results. The seven-gene risk signature had robust performance in recognizing MIBC patients with unfavorable outcome. Meanwhile, two of the three validation cohorts are composed of both NMIBC and MIBC patients, indicating that the seven-gene signature might be also applicable in NMIBC patients. Then we investigated the relationship between the seven-gene risk signature and clinical factors. Higher risk score indicates not only aggressive clinical manifestations but also treatment failure. Patients with high risk scores are more likely to be chemoresistant and radioresistant and develop disease progression after treatment, when compared with those patients with lower risk scores. Additionally, we also compared our seven-gene risk signature with previously reported pan cancer mRNA stemness indices. Prognostic value of our seven-gene risk signature is better than mRNAsi in BC. High risk group of patients, which was identified by the seven-gene signature, had significantly higher mRNAsi. In Malta et al.'s study, higher mRNAsi suggests poorer clinical outcome. Taken together, the novel seven-gene risk signature is distinct, robust, and reliable.

Among seven genes, several genes have been uncovered to maintain cancer stem cells. EGFR, which encodes the epidermal growth factor receptor (EGFR) family of receptors, is frequently overexpressed or mutated in solid tumors and acts as driver gene and therapeutic target. EGFR plays an important role in bladder cell dedifferentiation as it maintains a less differentiated state of basal layer urothelial cells [19]. In bladder cancer, EGFR is widely reported to be overexpressed, which is more evident in MIBC than in NMIBC [20, 21]. As members of EGFR families, ERBB2 and RBB2 expression correlates with tumor recurrence and metastasis in MIBC [21]. The roles of EGFR signaling pathway in CSCs have been implicated in breast cancer [22], and glioblastoma [23]. However, the roles of EGFR in BCSCs are still unclear. FOXA2, a member of fork head box (FOX) superfamily, is considered as a urothelial differentiation marker in human bladder development as co-expression of FOXA2 and uroplakins (UPK) initiate the transformation of epithelial cells into a urothelial lineage. Expression of FOXA2 also activates the expression of bladder-specific genes, which are required for developing urothelium to give rise to other cell types [24]. In ovarian cancer, FOXA2 is responsible for maintaining CSCs properties, and knockdown of FOXA2 impairs the self-renewal ability of ovarian CSCs [25]. HES1 encodes the Hes1 protein which is an important molecular component for maintaining the undifferentiated state of normal stem cells and CSCs due to its transcriptional repression ability [26, 27]. Hes1 and its upstream Notch signaling pathway have been implicated to be associated with CSCs in colon cancer [28], pancreatic cancer [29], and breast cancer [30]. In oral cancer, loss of Hes1 decreases the tumor sphere-forming and self-renewal capacities of cancer cells [31]. In BC, absences of Notch-Hes1 pathway promotes epithelial-mesenchymal transition

(EMT) and tumor cells with low levels of Hes1 present more mesenchymal features and more aggressive phenotype [32].

MME, also known as CD10, encodes a cell surface zinc-dependent metallopeptidase that helps bladder cancer cells invade the adjacent matrix [33]. High CD10 expression in tumor cells was significantly associated with advanced tumor stage and poorer prognosis in BC [34, 35]. CD10 is reported to maintain CSCs in breast cancer and head and neck squamous cell carcinoma (HNSC) [36, 37]. In breast cancer, expression of CD10 is involved in progression and dissemination and correlates with the expression of cancer stem cell markers such as CD44 and ALDH1 significantly [36]. In HNSC, upregulated expression of CD10 is related with resistance to cisplatin, fluorouracil and radiation and an increased ability to form spheres and higher level of expression of the CSC marker OCT3/4 [37]. RBM6 (RNA binding motif protein 6) is an RNA-binding protein involved in nucleic acid binding and RNA binding. RBM6 mRNA expression level has shown to be upregulated in breast cancer [38], and pancreatic cancer [39] when compared with normal tissues. However, the role of RBM6 in CSCs and bladder cancer is seldom reported. Another member of RBM family, RBM5, is an RNA-binding protein that modulates cell cycle and apoptosis and acts as tumor suppressor in cancers [40]. In laryngeal carcinoma, expression of RBM6 is downregulated in tumor tissues and cancer cell lines. RBM6 could repress tumor growth and progression, which is possibly mediated by reducing the expression of EGFR, ERK and downregulation of EGFR-associated signaling pathways [41]. SMOC2 encodes the secreted modular calcium binding protein-2, which is a matricellular protein [42]. Overexpression of SMOC2 could significantly arrest cell at the G0/G1 phase so as to repress cell growth [43]. In HCC, downregulated expression SMOC2 was significantly associated with aggressive biological behaviors including advanced TNM stage and distant metastasis and reduced overall survival and disease-free survival [43]. TFRC (transferrin receptor), also known as CD71, is an important regulator for iron metabolism in cancers. CD71 is frequently upregulated and indicate aggressive phenotype as well as poor clinical outcome in esophageal squamous cell carcinoma [44] and breast cancer [45]. Expression of CD71 was closely related to the staining of proliferative marker Ki-67 and CD71-suppressed tumor cells exhibit cell growth inhibition and cell cycle arrest at S phase [44]. CD71 is also regarded as a phenotypic marker of cancer stem-like cells in cervical cancer [46].

GSEA analysis uncovered that the most enriched biological processes in high risk group were dominantly associated with extracellular adhesion, development of soft tissues, and angiogenesis, indicating that high risk group of patients might possess more prominent stemness characteristics than low risk group. CSCs are believed to reside in niches, which constitute the tumor environment and maintain stem-like properties in cancer cells [47]. The extracellular matrix is a major component of the tumor microenvironment and plays crucial roles in cell differentiation and determination of the balance between self-renewal and differentiation in stem cells [48]. The presence of CSCs are also associated with angiogenesis and tumor vasculature as CSCs could transdifferentiate into endothelial cells [49] or produce several cytokines and chemokines [50] to regulate the tumor vasculature and from their own vascular niches.

Despite the promising results, limitations of our study should be noted. Firstly, although the novel seven-gene signature, which was built in training cohort, could be confirmed in three validation cohorts. clinical features of patients in three validation cohorts were not intact and we were therefore unable to verify the independence of novel gene signature in three validation cohorts. Secondly, our seven-gene signature could effectively predicate treatment resistance. However, the detailed information of chemotherapy and radiotherapy were not recorded. Last but not least, our study was based on retrospective view of transcriptome data and clinical data. Prospective clinical trials with large cohorts are required before application of the seven gene signature into diagnosis and treatment. Biological and molecular experiments are also needed to explain the relationship between these stemness-related genes and cancer stem cell biology in bladder cancer.

Conclusion

In conclusion, we developed and validated a novel and robust seven-gene risk signature related with cancer stemness that could predict prognosis and treatment response in MIBC patients. Our findings could contribute to better understanding of CSCs and identification of novel therapeutic targets in bladder cancer.

Declarations

Acknowledgements

The authors are grateful to the public databases TCGA, GEO, CancerSEA for the availability of the data.

Authors' contributions

MT and JLH conceived and designed the study. MT and JQH performed statistical analyses. MT wrote the manuscript. JLH and HZ reviewed the manuscript.

Funding

None.

Availability of data and materials

The datasets used and/or analyzed during the current study are all available from public database and can be retrieved from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Head and Neck Oncology and Department of Radiation Oncology, Cancer Center and State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, No.37 Guo-Xue Road, Chengdu 610041, China. ²Department of Medical Oncology, West China Hospital, Sichuan University, No.37 Guo-Xue Road, Chengdu 610041, China.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.** *CA Cancer J Clin* 2018, **68**:394-424.
2. Khandelwal P, Abraham SN, Apodaca G: **Cell biology and physiology of the uroepithelium.** *Am J Physiol Renal Physiol* 2009, **297**:F1477-1501.
3. Shadpour P, Emami M, Haghdani S: **A Comparison of the Progression and Recurrence Risk Index in Non-Muscle-Invasive Bladder Tumors Detected by Narrow-Band Imaging Versus White Light Cystoscopy, Based on the EORTC Scoring System.** *Nephrourol Mon* 2016, **8**:e33240.
4. Ploussard G, Shariat SF, Dragomir A, Kluth LA, Xylinas E, Masson-Lecomte A, Rieken M, Rink M, Matsumoto K, Kikuchi E, et al: **Conditional survival after radical cystectomy for bladder cancer: evidence for a patient changing risk profile over time.** *Eur Urol* 2014, **66**:361-370.
5. Ishii H, Iwatsuki M, Ieta K, Ohta D, Haraguchi N, Mimori K, Mori M: **Cancer stem cells and chemoradiation resistance.** *Cancer Sci* 2008, **99**:1871-1877.
6. Easwaran H, Tsai HC, Baylin SB: **Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance.** *Mol Cell* 2014, **54**:716-727.
7. Chan KS, Espinosa I, Chao M, Wong D, Ailles L, Diehn M, Gill H, Presti J, Jr., Chang HY, van de Rijn M, et al: **Identification, molecular characterization, clinical prognosis, and therapeutic targeting of human bladder tumor-initiating cells.** *Proc Natl Acad Sci U S A* 2009, **106**:14016-14021.
8. Bryan RT: **Bladder cancer and cancer stem cells: basic science and implications for therapy.** *ScientificWorldJournal* 2011, **11**:1187-1194.
9. Cordon-Cardo C: **Molecular alterations associated with bladder cancer initiation and progression.** *Scand J Urol Nephrol Suppl* 2008:154-165.
10. Yang YM, Chang JW: **Bladder cancer initiating cells (BCICs) are among EMA-CD44v6+ subset: novel methods for isolating undetermined cancer stem (initiating) cells.** *Cancer Invest* 2008, **26**:725-733.

11. Su Y, Qiu Q, Zhang X, Jiang Z, Leng Q, Liu Z, Stass SA, Jiang F: **Aldehyde dehydrogenase 1 A1-positive cell population is enriched in tumor-initiating cells and associated with progression of bladder cancer.** *Cancer Epidemiol Biomarkers Prev* 2010, **19**:327-337.
12. Yang Z, Li C, Fan Z, Liu H, Zhang X, Cai Z, Xu L, Luo J, Huang Y, He L, et al: **Single-cell Sequencing Reveals Variants in ARID1A, GPRC5A and MLL2 Driving Self-renewal of Human Bladder Cancer Stem Cells.** *Eur Urol* 2017, **71**:8-12.
13. Yuan H, Yan M, Zhang G, Liu W, Deng C, Liao G, Xu L, Luo T, Yan H, Long Z, et al: **CancerSEA: a cancer single-cell state atlas.** *Nucleic Acids Res* 2019, **47**:D900-d908.
14. Wang H, Lengerich BJ, Aragam B, Xing EP: **Precision Lasso: accounting for correlations and linear dependencies in high-dimensional genomic data.** *Bioinformatics* 2019, **35**:1181-1187.
15. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kamińska B, Huelsken J, Omberg L, Gevaert O, et al: **Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation.** *Cell* 2018, **173**:338-354.e315.
16. Abudurexiti M, Xie H, Jia Z, Zhu Y, Zhu Y, Shi G, Zhang H, Dai B, Wan F, Shen Y, Ye D: **Development and External Validation of a Novel 12-Gene Signature for Prediction of Overall Survival in Muscle-Invasive Bladder Cancer.** *Front Oncol* 2019, **9**:856.
17. Cao R, Yuan L, Ma B, Wang G, Qiu W, Tian Y: **An EMT-related gene signature for the prognosis of human bladder cancer.** *J Cell Mol Med* 2020, **24**:605-617.
18. Jiang W, Zhu D, Wang C, Zhu Y: **An immune relevant signature for predicting prognoses and immunotherapeutic responses in patients with muscle-invasive bladder cancer (MIBC).** *Cancer Med* 2020.
19. Messing E, Gee JR, Saltzstein DR, Kim K, diSant'Agnese A, Kolesar J, Harris L, Faerber A, Havighurst T, Young JM, et al: **A phase 2 cancer chemoprevention biomarker trial of isoflavone G-2535 (genistein) in presurgical bladder cancer patients.** *Cancer Prev Res (Phila)* 2012, **5**:621-630.
20. Lipponen PK, Liukkonen TJ: **Reduced expression of retinoblastoma (Rb) gene protein is related to cell proliferation and prognosis in transitional-cell bladder cancer.** *J Cancer Res Clin Oncol* 1995, **121**:44-50.
21. Chow NH, Chan SH, Tzai TS, Ho CL, Liu HS: **Expression profiles of ErbB family receptors and prognosis in primary transitional cell carcinoma of the urinary bladder.** *Clin Cancer Res* 2001, **7**:1957-1962.
22. Steelman LS, Fitzgerald T, Lertpiriyapong K, Cocco L, Follo MY, Martelli AM, Neri LM, Marmioli S, Libra M, Candido S, et al: **Critical Roles of EGFR Family Members in Breast Cancer and Breast Cancer Stem Cells: Targets for Therapy.** *Curr Pharm Des* 2016, **22**:2358-2388.
23. Liffers K, Lamszus K, Schulte A: **EGFR Amplification and Glioblastoma Stem-Like Cells.** *Stem Cells Int* 2015, **2015**:427518.
24. Gandhi D, Molotkov A, Batourina E, Schneider K, Dan H, Reiley M, Laufer E, Metzger D, Liang F, Liao Y, et al: **Retinoid signaling in progenitors controls specification and regeneration of the urothelium.** *Dev Cell* 2013, **26**:469-482.

25. Peng Q, Qin J, Zhang Y, Cheng X, Wang X, Lu W, Xie X, Zhang S: **Autophagy maintains the stemness of ovarian cancer stem cells by FOXA2.** *J Exp Clin Cancer Res* 2017, **36**:171.
26. Tian C, Zheng G, Cao Z, Li Q, Ju Z, Wang J, Yuan W, Cheng T: **Hes1 mediates the different responses of hematopoietic stem and progenitor cells to T cell leukemic environment.** *Cell Cycle* 2013, **12**:322-331.
27. Sang L, Roberts JM, Collier HA: **Hijacking HES1: how tumors co-opt the anti-differentiation strategies of quiescent cells.** *Trends Mol Med* 2010, **16**:17-26.
28. Gao F, Zhang Y, Wang S, Liu Y, Zheng L, Yang J, Huang W, Ye Y, Luo W, Xiao D: **Hes1 is involved in the self-renewal and tumorigenicity of stem-like cancer cells in colon cancer.** *Sci Rep* 2014, **4**:3963.
29. Abel EV, Kim EJ, Wu J, Hynes M, Bednar F, Proctor E, Wang L, Dziubinski ML, Simeone DM: **The Notch pathway is important in maintaining the cancer stem cell population in pancreatic cancer.** *PLoS One* 2014, **9**:e91983.
30. Wong NK, Fuller M, Sung S, Wong F, Karsan A: **Heterogeneity of breast cancer stem cells as evidenced with Notch-dependent and Notch-independent populations.** *Cancer Med* 2012, **1**:105-113.
31. Lee SH, Hong HS, Liu ZX, Kim RH, Kang MK, Park NH, Shin KH: **TNF α enhances cancer stem cell-like phenotype via Notch-Hes1 activation in oral squamous cell carcinoma cells.** *Biochem Biophys Res Commun* 2012, **424**:58-64.
32. Maraver A, Fernandez-Marcos PJ, Cash TP, Mendez-Pertuz M, Dueñas M, Maietta P, Martinelli P, Muñoz-Martin M, Martínez-Fernández M, Cañamero M, et al: **NOTCH pathway inactivation promotes bladder cancer progression.** *J Clin Invest* 2015, **125**:824-830.
33. Jang TJ: **CD10 Is Again Expressed at a Certain Stage during the Neoplastic Process of Bladder Transitional Cell Carcinomas.** *Cancer Res Treat* 2012, **44**:262-266.
34. Kumagai-Togashi A, Uozaki H, Kikuchi Y, Watabe S, Numakura S, Watanabe M: **Tumorous CD10 Is More Strongly Related to the Progression of Urothelial Carcinoma than Stromal CD10.** *Anticancer Res* 2019, **39**:635-640.
35. Abdou AG: **CD10 expression in tumour and stromal cells of bladder carcinoma: an association with bilharziasis.** *Apmis* 2007, **115**:1206-1218.
36. Louhichi T, Saad H, Dhiab MB, Ziadi S, Trimeche M: **Stromal CD10 expression in breast cancer correlates with tumor invasion and cancer stem cell phenotype.** *BMC Cancer* 2018, **18**:49.
37. Fukusumi T, Ishii H, Konno M, Yasui T, Nakahara S, Takenaka Y, Yamamoto Y, Nishikawa S, Kano Y, Ogawa H, et al: **CD10 as a novel marker of therapeutic resistance and cancer stem cells in head and neck squamous cell carcinoma.** *Br J Cancer* 2014, **111**:506-514.
38. Rintala-Maki ND, Goard CA, Langdon CE, Wall VE, Traulsen KE, Morin CD, Bonin M, Sutherland LC: **Expression of RBM5-related factors in primary breast tissue.** *J Cell Biochem* 2007, **100**:1440-1458.
39. Duan B, Hu X, Fan M, Xiong X, Han L, Wang Z, Tong D, Liu L, Wang X, Li W, et al: **RNA-Binding Motif Protein 6 is a Candidate Serum Biomarker for Pancreatic Cancer.** *Proteomics Clin Appl* 2019, **13**:e1900048.

40. Mourtada-Maarabouni M, Sutherland LC, Meredith JM, Williams GT: **Simultaneous acceleration of the cell cycle and suppression of apoptosis by splice variant delta-6 of the candidate tumour suppressor LUCA-15/RBM5.** *Genes Cells* 2003, **8**:109-119.
41. Wang Q, Wang F, Zhong W, Ling H, Wang J, Cui J, Xie T, Wen S, Chen J: **RNA-binding protein RBM6 as a tumor suppressor gene represses the growth and progression in laryngocarcinoma.** *Gene* 2019, **697**:26-34.
42. Bloch-Zupan A, Jamet X, Etard C, Laugel V, Muller J, Geoffroy V, Strauss JP, Pelletier V, Marion V, Poch O, et al: **Homozygosity mapping and candidate prioritization identify mutations, missed by whole-exome sequencing, in SMOC2, causing major dental developmental defects.** *Am J Hum Genet* 2011, **89**:773-781.
43. Huang XQ, Zhou ZQ, Zhang XF, Chen CL, Tang Y, Zhu Q, Zhang JH, Xia JC: **Overexpression of SMOC2 Attenuates the Tumorigenicity of Hepatocellular Carcinoma Cells and Is Associated With a Positive Postoperative Prognosis in Human Hepatocellular Carcinoma.** *J Cancer* 2017, **8**:3812-3827.
44. Chan KT, Choi MY, Lai KK, Tan W, Tung LN, Lam HY, Tong DK, Lee NP, Law S: **Overexpression of transferrin receptor CD71 and its tumorigenic properties in esophageal squamous cell carcinoma.** *Oncol Rep* 2014, **31**:1296-1304.
45. Habashy HO, Powe DG, Staka CM, Rakha EA, Ball G, Green AR, Aleskandarany M, Paish EC, Douglas Macmillan R, Nicholson RI, et al: **Transferrin receptor (CD71) is a marker of poor prognosis in breast cancer and can predict response to tamoxifen.** *Breast Cancer Res Treat* 2010, **119**:283-293.
46. Tyagi A, Vishnoi K, Mahata S, Verma G, Srivastava Y, Masaldan S, Roy BG, Bharti AC, Das BC: **Cervical Cancer Stem Cells Selectively Overexpress HPV Oncoprotein E6 that Controls Stemness and Self-Renewal through Upregulation of HES1.** *Clin Cancer Res* 2016, **22**:4170-4184.
47. Hanahan D, Coussens LM: **Accessories to the crime: functions of cells recruited to the tumor microenvironment.** *Cancer Cell* 2012, **21**:309-322.
48. Yamashita YM, Yuan H, Cheng J, Hunt AJ: **Polarity in stem cell division: asymmetric stem cell division in tissue homeostasis.** *Cold Spring Harb Perspect Biol* 2010, **2**:a001313.
49. Ricci-Vitiani L, Pallini R, Biffoni M, Todaro M, Invernici G, Cenci T, Maira G, Parati EA, Stassi G, Larocca LM, De Maria R: **Tumour vascularization via endothelial differentiation of glioblastoma stem-like cells.** *Nature* 2010, **468**:824-828.
50. Folkins C, Shaked Y, Man S, Tang T, Lee CR, Zhu Z, Hoffman RM, Kerbel RS: **Glioma tumor stem-like cells promote tumor angiogenesis and vasculogenesis via vascular endothelial growth factor and stromal-derived factor 1.** *Cancer Res* 2009, **69**:7243-7251.

Figures

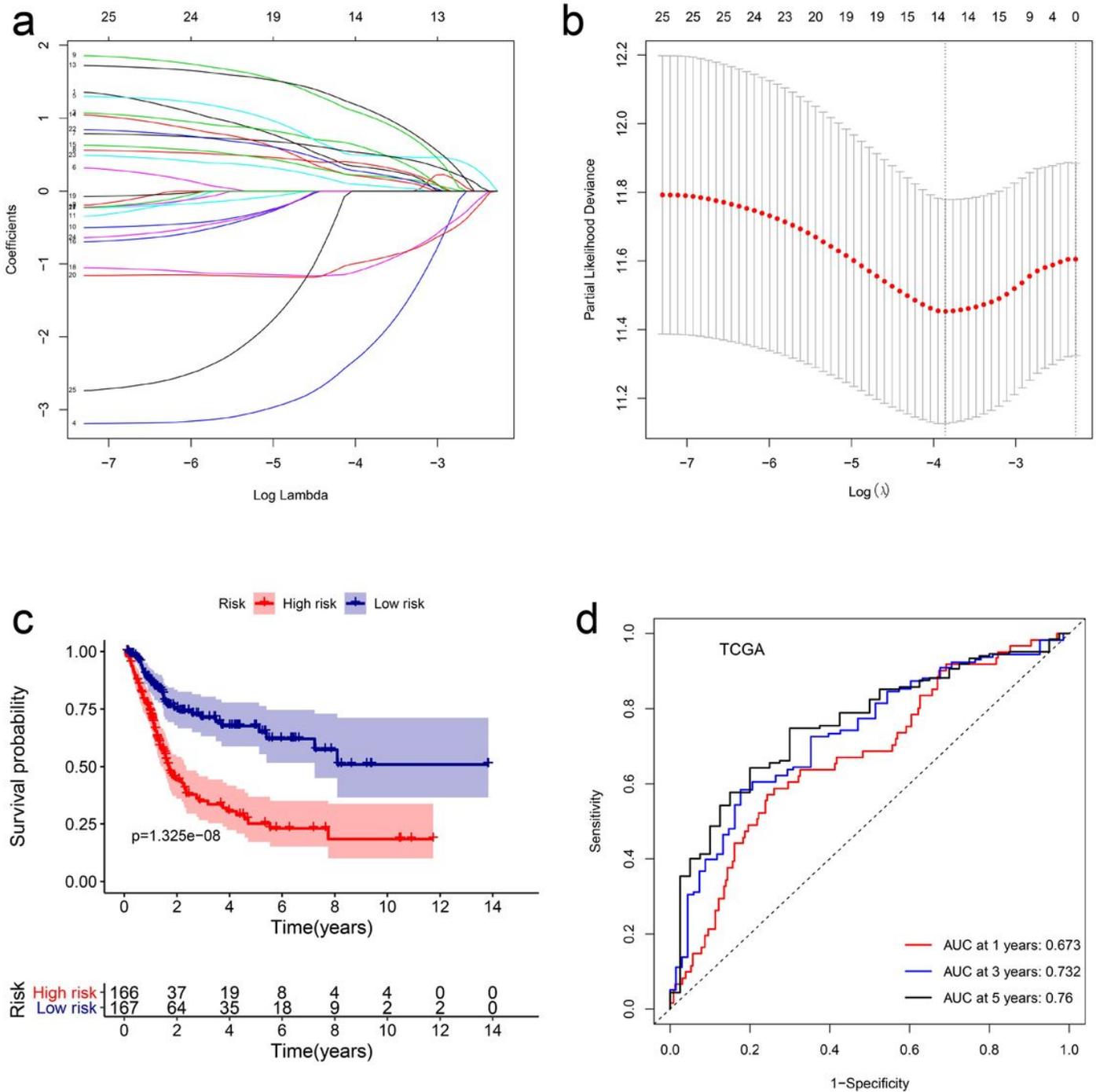


Figure 1

Development of the seven-gene signature in training cohort. a The least absolute shrinkage and selection operator (LASSO) regression of 25 prognostic genes. b The selection of best lambda to construct the gene signature. c Kaplan–Meier survival curve of low risk and high risk groups of patients. d The time-dependent ROC curve of the seven-gene signature.

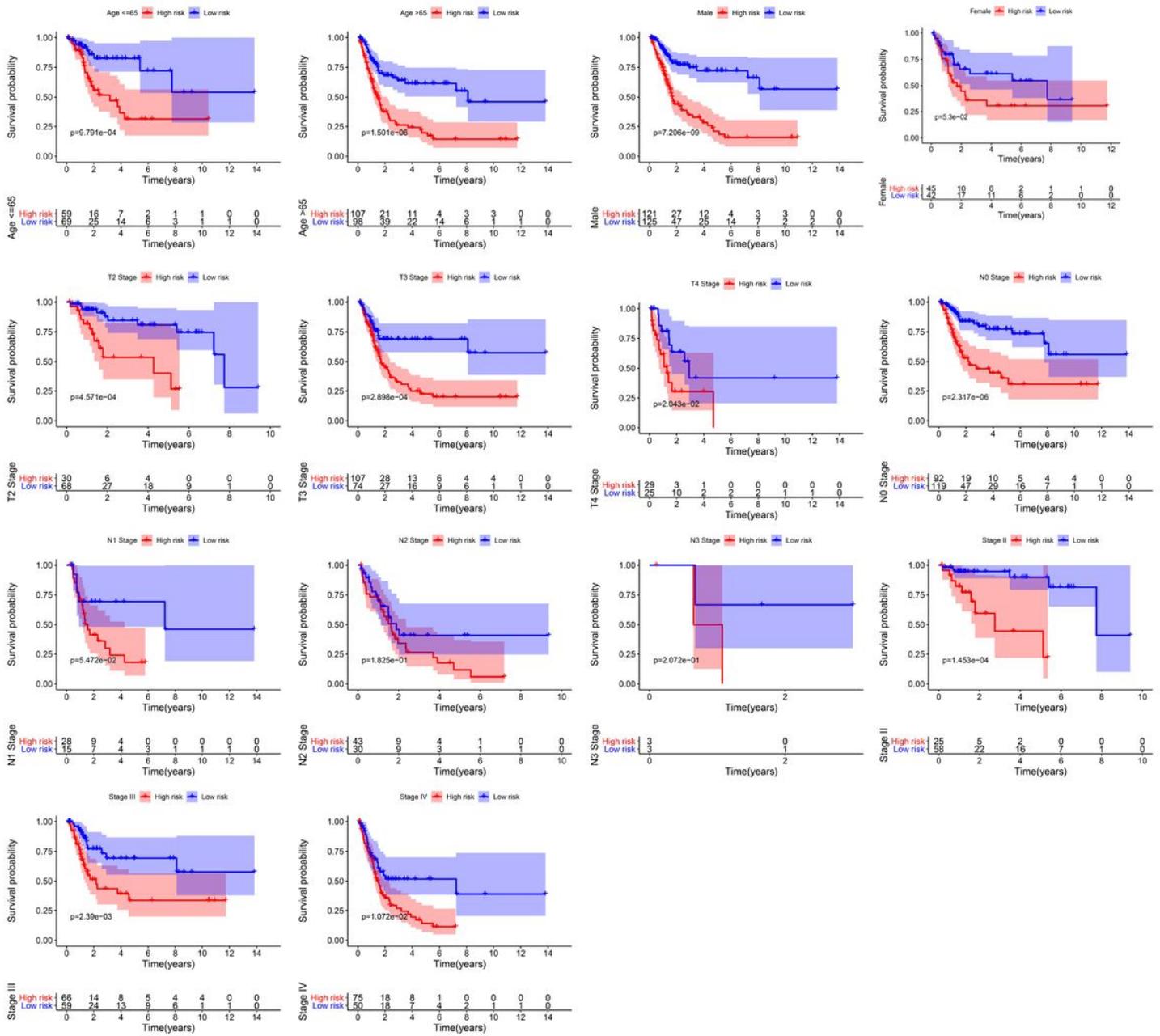


Figure 2

Kaplan-Meier analysis of seven-gene risk signature when bladder cancer patients were stratified by age, gender, and clinical stage.

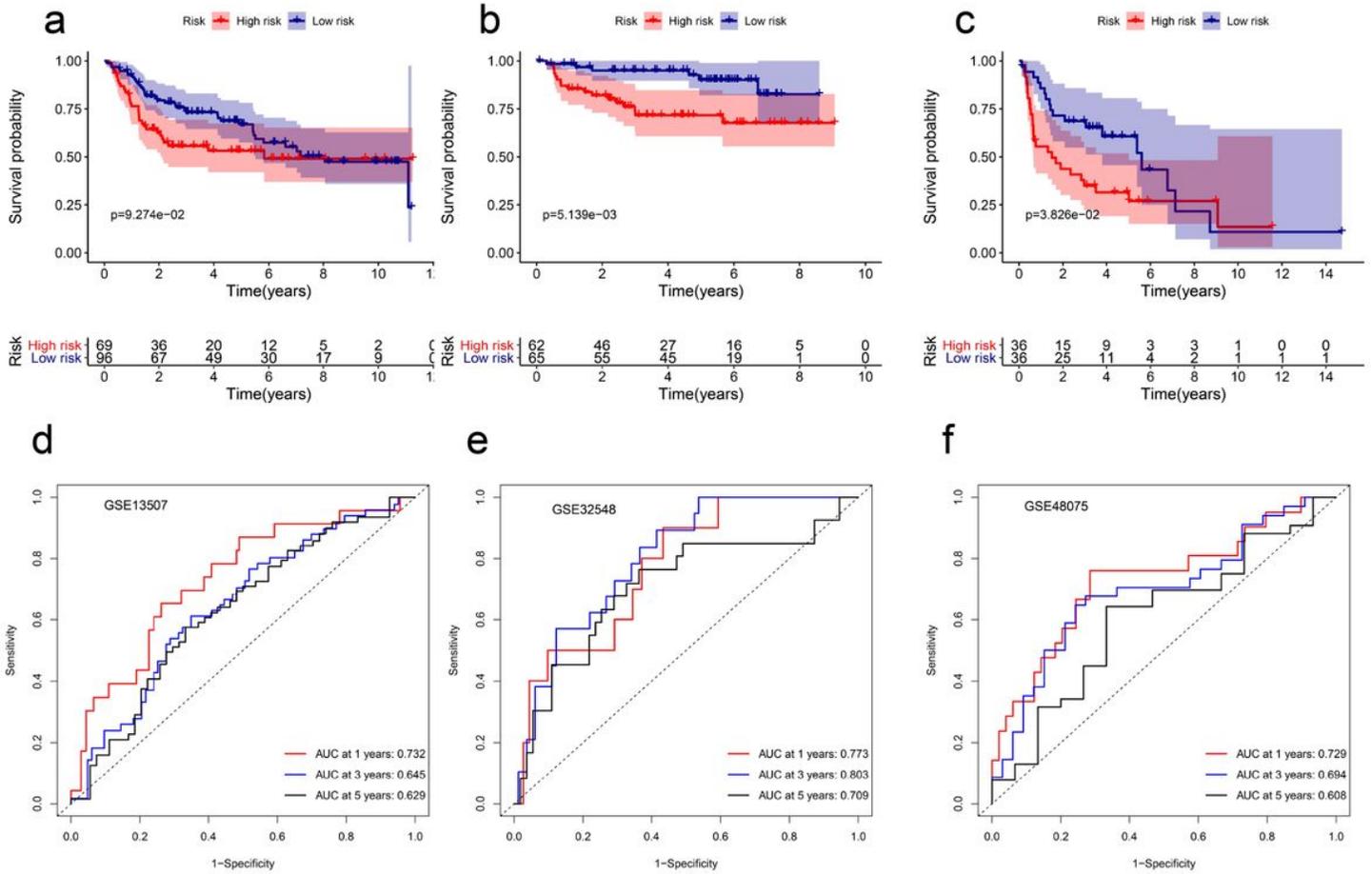


Figure 3

External validation of the seven-gene signature in three datasets. Patients with high risk scores exhibited poor 5-year OS than those with low risk scores in GSE13507 (a), GSE32548 (b), GSE38075 (c). Time-dependent ROC curves suggested that the seven-gene signature had robust predictive power in GSE13507 (d), GSE32548 (e), GSE38075 (f).

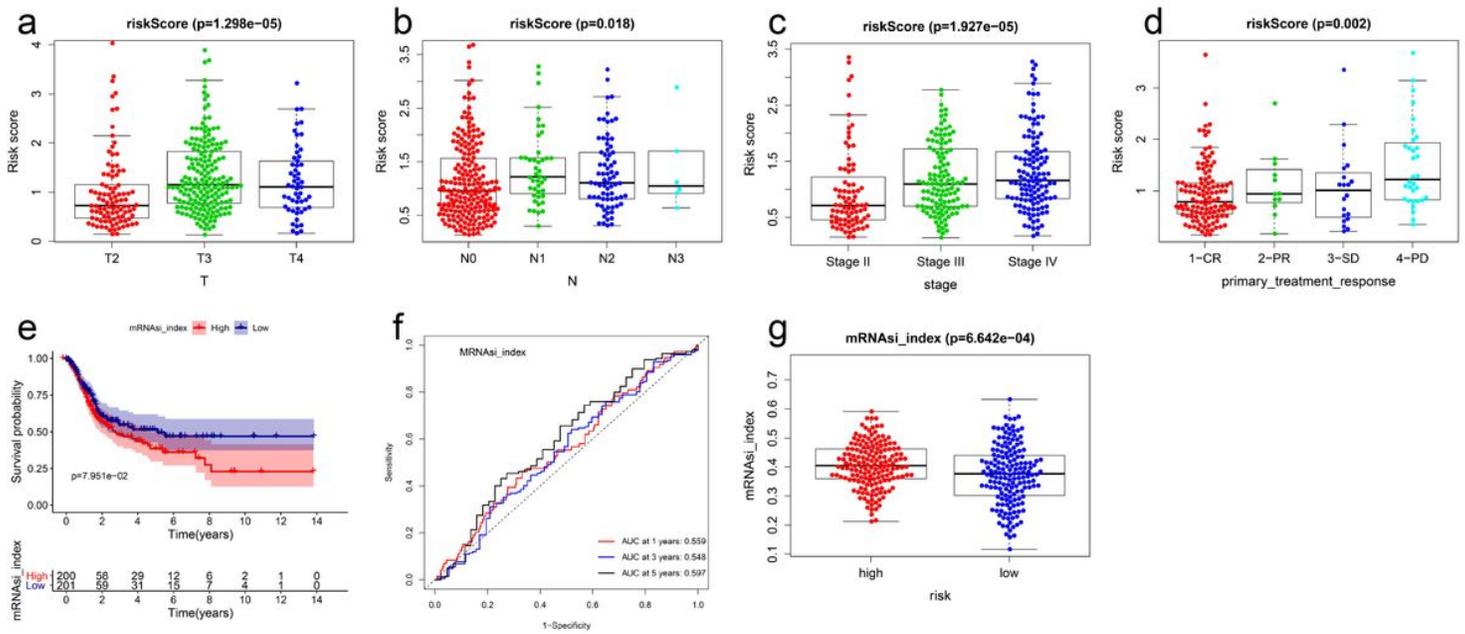


Figure 4

Correlation between the seven-gene risk signature and clinical parameters, treatment response, and mRNAsi in bladder cancer. Bladder cancer patients with advanced T stage (a), N stage (b), and overall stage (c) had significantly higher risk scores than patients with low risk scores. e Kaplan–Meier analysis of high and low risk groups stratified by mRNAsi index. f Time-dependent ROC curve of mRNAsi in training cohort. g Patients in the seven-gene signature generated high risk group had higher mRNAsi than patients in low risk group. mRNAsi: mRNA stemness indices.

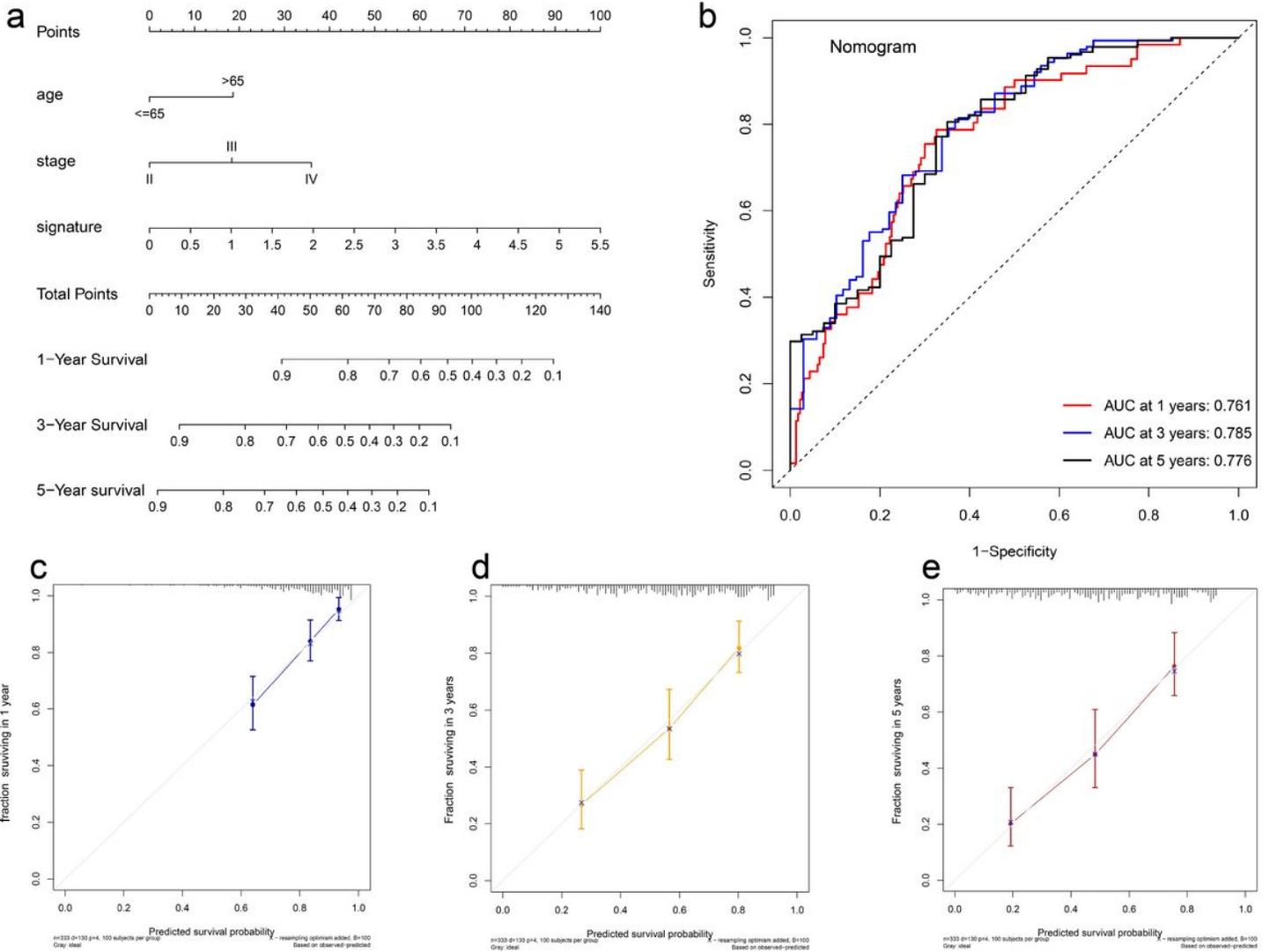


Figure 5

Establishment of a genetic and clinical nomogram. a The nomogram integrating age, overall stage and gene signature. b Time-dependent ROC curve of the prognostic nomogram. c-e The calibration curve of the nomogram in predicting 1-, 3-, 5-, 10-years survival possibilities.

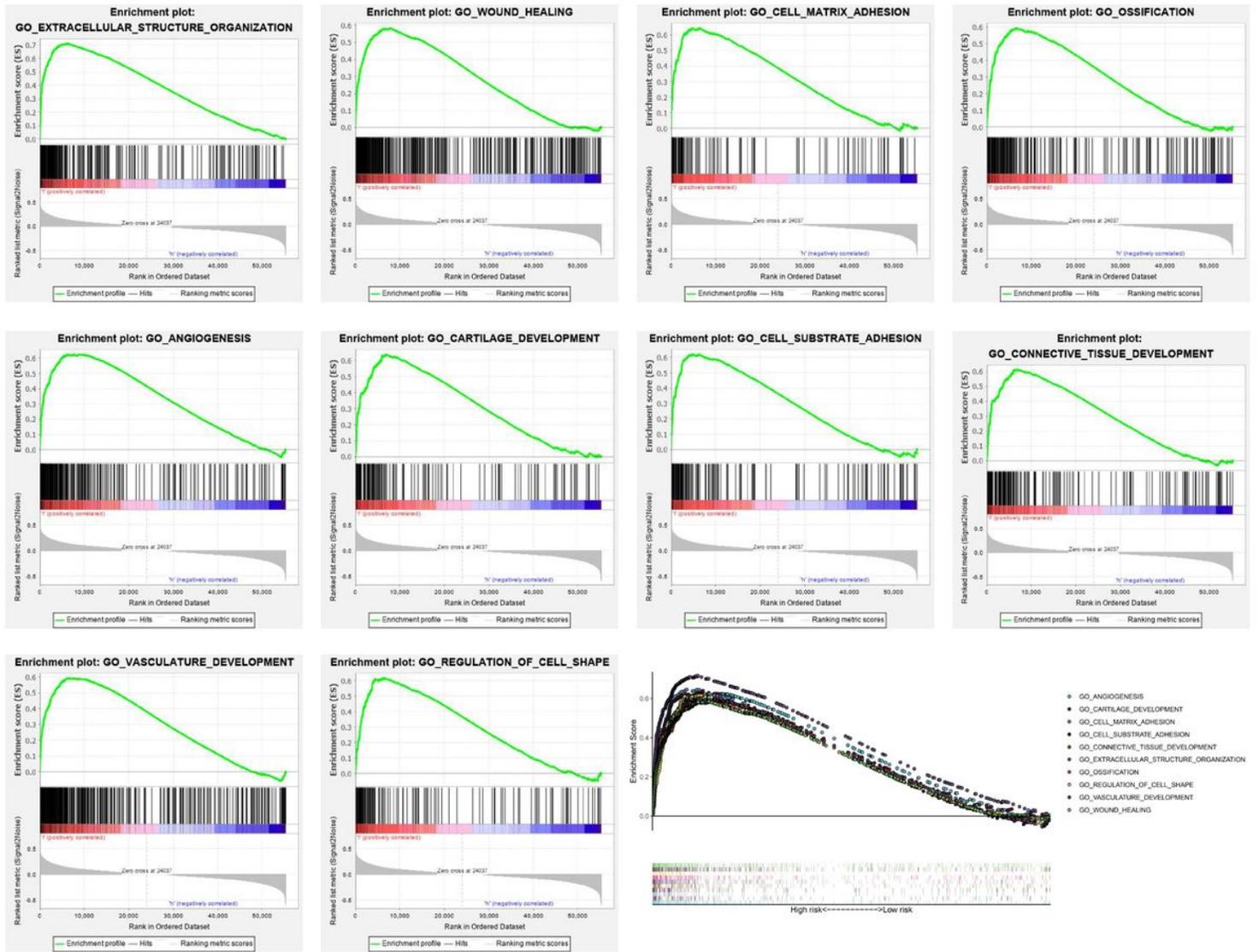


Figure 6

Enrichment plots of top ten altered biological processes in high risk group patients.

Supplementary Files

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

- [Additionalfile2.doc](#)
- [Additionalfile4.doc](#)
- [Additionalfile3.doc](#)
- [Additionalfile1.docx](#)