An Immune-Related Clinical-Gene-Risk-Score (CGRS) prognosis Model for Bladder Cancer

Sanhe Liu (lshwdl@163.com)
Huazhong Agriculture University: Huazhong Agricultural University https://orcid.org/0000-0001-7682-053X

Liqun Duan
Hubei Cancer Institute: Hubei Cancer Hospital

Yiqi Wang
Huazhong Agriculture University: Huazhong Agricultural University

Diansheng Cui
Hubei Cancer Institute: Hubei Cancer Hospital

Kangli Deng
Hubei Cancer Institute: Hubei Cancer Hospital

Shaozhong Wei
Huazhong Agriculture University: Huazhong Agricultural University

Zhiqiang Dong
Huazhong Agriculture University: Huazhong Agricultural University

Research Article

Keywords:

Posted Date: May 6th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1397810/v2

Abstract

Background Immunotherapy for bladder cancer is prevalent, while the efficacy is uncertain. To predict the outcomes of immunotherapy in bladder cancer patients, an immune-related prognosis model was conducted in this study.

Methods Single sample Gene Set Enrichment Analysis was used to delineate the landscape of the tumor immune microenvironment based on two independent public databases. Hierarchical clustering analysis and t-distributed Stochastic Neighbor Embedding cluster map were used to classify the bladder cancers in the training cohort as well as the validating cohort into the immune high and immune low subtypes. Differentially expression analysis and The least absolute shrinkage and selection operator regression analysis had identified 23 differentially expressed immune-related genes between the two subtype patients to build a prognostic gene risk score model. Further combining the model with other clinicopathologic signatures, a Clinical-Gene-Risk-Score (CGRS) prognosis model was developed.

Results Validations of the two models in independent cohorts revealed good accuracy.

Conclusions Our model could provide prognostic indicators and potential immunotherapeutic targets for bladder cancers.

Background

Bladder cancer is the most common malignant tumor of the urogenital tract with significant morbidity and mortality, the majority of which are transitional cell carcinomas derived from the uroepithelium[1-4]. Although substantial progress has been made in bladder cancer treatment, the therapeutic prognosis remains unsatisfactory in patients with progressed and metastatic bladder cancer[1, 5]. In recent years, immunotherapy has emerged as an effective therapeutic approach against cancer that relies on dynamic interactions between the tumor cells and the immune cells in the tumor microenvironment (TME)[6]. Encouraging advancements have been made in cancer immunotherapy which represents a revolution in the paradigm for cancer treatment by harnessing the power of the host innate and adaptive immune system to control cancer progression and metastasis[7-11]. Bladder cancer carries one of the highest overall mutation loads across various cancer types[12] and is known to be immunogenic. Bacille Calmette–Guérin (BCG), injected intravesically into the bladder cavity, was the first effective immunotherapy for bladder cancer[13]. However, despite recent advances in cancer immunotherapy, the efficacy of this strategy remains limited[14], and only a small portion of cancer patients can benefit from the treatment[15, 16].

Diverse players involved in cancer immunotherapy include tumor cells, immune cells, stromal cells, and other tumor microenvironmental factors[17]. The composition of the tumor immune microenvironment is a critical determinant of tumor-immune interactions and can determine the response to immunotherapy[18]. Although immunotherapy for bladder cancer is widely used, few studies have addressed the immune microenvironment of bladder cancer. Thus, it will be of great importance to

delineate the landscape of the tumor immune microenvironment, classify the bladder cancers into the immune sensitive and nonsensitive subtype, and identify the differential expression genes between sensitive and nonsensitive patients to build a prognostic model for predicting the outcomes of immunotherapy in bladder cancer patients.

Here we conducted an integrated bioinformatic analysis of the public bladder cancer datasets from the GEO and TCGA databases and characterized the tumor immune landscapes which were further divided into immune high and immune low subtypes individually, differentially expression analysis between the two groups demonstrated 23 key immune-related prognostic genes that constructed a gene risk score model, combined with other clinicopathologic signatures to develop a Clinical-Gene-Risk-Score (CGRS) model, validations of the two models in independent cohort revealed good accuracy.

Methods

1. Data collection and preparation

The GEO dataset was downloaded from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/gds) which is composed of three chip datasets (gse13507, gse32894, gse48276) with full follow-up pieces of information. The three datasets were normalized using the normalizeBetweenArrays function in R package limma[19] and further merged as the training cohort. The TCGA-BLCA dataset as the validation cohort was downloaded from The Cancer Genome Atlas (TCGA, https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga), converted from FPKM to TPM, then normalized with normalizeBetweenArrays. The cancer samples in gse13507, gse32894, gse48276, and TCGA-BLCA datasets were 256, 308, 116, and 405, respectively. The immune-related genes were downloaded from ImmPort[20] (https://www.immport.org/home) and InnateDB[21] (https://www.innatedb.com/), and 1793 genes were incorporated after removing duplicates. The transcription factors were downloaded from Cistrome Project[22] (http://cistrome.com/). The tumor mutation burden (TMB) data with 411 samples were also downloaded from TCGA.

2. Single sample Gene Set Enrichment Analysis (ssGSEA)

ssGSEA was implemented to quantify activity profiles of the 28 immune cell types infiltrating each cancer type, and the ssGSEA scores were computed by the gsva function of R package GSVA[23]. Hierarchical clustering of samples was done with the hclust function of the R package and patients were stratified into immune high subtype and immune low subtype according to the median split. A clustering heatmap was generated using the ColorDendrogram function of the R package sparcl (https://CRAN.R-project.org/package=sparcl). The t-distributed Stochastic Neighbor Embedding (tSNE) visualization was used to confirm our appropriate clustering (R package Rtsne (https://CRAN.R-project.org/package=Rtsne)). The estimateScore function of the R package estimate[24] was used to calculate a TME immune score. On representing the relative abundance of 28 immune cell types, the

ssGSEA scores of each clinicopathologic signature and TME score for every sample were presented with a heatmap using the pheatmap (https://CRAN.R-project.org/package=pheatmap) R package. The ssGSEA enrichment scores of the anti-tumor and immune-suppressive immune cells were compared between the immune high and immune low subtypes by Wilcox t-test via ggviolin (R package ggplot2 (https://ggplot2.tidyverse.org). Gene signature for each immune cell type was obtained from previously published data[25, 26]. The TME scores and the expressions of the Human leukocyte antigen (HLA) genes of the two subtypes were also compared by the Wilcox t-test. The exprssions of the immune checkpoint genes between the two subtypes were compared by the Wilcox t-test via ggboxplot (https://www.rdocumentation.org/packages/ggpubr/versions/0.4.0/topics/ggboxplot). Gene set enrichment analysis (GSEA) was performed with omicShare (https://www.omicshare.com/tools/Home/Soft/gsea) and the immune-related Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were selected for enrichment analysis. Aditionally, the correlation of the immune cells were assessed by the Spearman correlation test (R package correplot (https://CRAN.R-project.org/package=corrplot)).

3. Independent prognostic immune cells

Survival analysis according to the ssGSEA enrichment scores of immune cells and the immune subtypes was performed by the log-rank test and displayed with the Kaplan-Meier curves, using the R package survival (https://CRAN.R-project.org/package=survival) and survminer (https://CRAN.R-project.org/package=survminer). The surv_cutpoint function was used to determine the optimal cutoff for the ssGSEA scores of the 28 immune cell types. Univariate and multivariate Cox regression analyses were shown with a forest plot. Significant parameters with p-value < 0.05 in univariate analysis were included in a multivariate Cox analysis, and variables with P < 0.05 in multivariate analysis were considered as independent prognostic factors for overall survival (OS) of bladder cancers.

4. Construction and validation of gene risk score model

Differential expression analysis was performed with the R package limma by comparing the immune high subtype samples with the immune low subtype samples. The results were presented with a volcano map. The intersection of the differential expressed mRNAs and the immune-related genes were extracted with the R package venn (https://CRAN.R-project.org/package=venn). The GEO dataset (the training cohort) and the TCGA-BLCA dataset (the validation cohort) were subjected to univariate cox regression analysis separately and the p-value < 0.05 was considered statistically significant. The least absolute shrinkage and selection operator (Lasso) regression analysis of the training cohort was conducted with R package glmnet (https://CRAN.R-project.org/package=glmnet). Sankey diagram was employed to describe the correlations of the prognostic genes based on the univariate result and the downloaded transcription factors, and the Protein-Protein Interaction (PPI) Networks between them was derived from

the STRING database (https://string-db.org/)[27, 28]. Heatmaps of differential expression genes and differential expression immune-related genes were performed with the R package pheatmap.

The Lasso-Cox regression was used to construct the gene risk score model. Calibration of the model in the training set as well as the validation set, which compared predicted with actual survival, was evaluated with a calibration curve. The predictive accuracy of the model was assessed with the area under the curve (AUC) of the receiver operating curve (ROC), using the timeROC (https://CRAN.R-project.org/package=timeROC)[28] and rms (https://CRAN.R-project.org/package=rms) R packages. The correlation of the immune cells and the model genes was assessed by the Spearman correlation test. The risk score, vital status, and their correlations were visualized with risk curve, scatter plot, and correlation scatter plot, respectively. Based on the median risk score, all populations of the two cohorts were divided into high- and low-risk groups. Survival analysis and the comparison boxplots of the immune checkpoint genes between high- and low-risk groups were performed as previously described.

5. Clinical-Gene-Risk-Score (CGRS) model construction and validation

Univariate and multivariate Cox regression analysis, which incorporated the clinicopathologic and gene risk score predictors accounting for both statistical and clinical significance, was applied to the construction of a nomogram which we named as a Clinical-Gene-Risk-Score (CGRS) model. The nomogram model was built with the R package regplot (https://CRAN.R-project.org/package=regplot). The validation of the CGRS model was completed as described previously.

All statistical analysis was performed using R (version 4.1.0) and R studio for windows (version 1.4.1103). Significance levels were defined as follows: ns=not significant, * = P-value < 0.05; ** = P-value < 0.01; *** = P-value < 0.001.

Results

1. Hierarchical clustering identified immune high and immune low subtypes

Hierarchical clustering analysis identified two clusters, the immune high subtype and the immune low subtype, among 680 samples from three merged GEO datasets with 405 samples from the TCGA-BLCA dataset. In the GEO dataset, 241 samples were clustered into the immune high subtype and 439 samples were defined as the immune low subtype. While in the TCGA-BLCA dataset, the immune high and immune low samples were 340 and 65, respectively. Both the two clusters could be well-separated in the two datasets (Supplementary Fig 1).

2. The landscape of tumor immune microenvironment in the GEO and TCGA datasets.

SsGSEA analysis was performed to estimate the abundance of the 28 immune cell types in each sample. The ssGSEA enrichment score of each sample was calculated to evaluate the landscape of the tumor immune microenvironment in the GEO dataset and the TCGA dataset. In the GEO dataset, the immune scores, stromal scores, and ESTIMATE scores of the immune high subtype samples were higher than the scores in the immune low subtype samples. However, the tumor purity scores were the opposite (Fig 1 a, Fig 2 b). The results of the TCGA dataset were consistent with those of the GEO dataset (Fig 1 b, Fig 2 g). The clinicopathologic characters were also presented in Fig 1. Most of the scores of the anti-tumor immunity cells and immune-suppressive cells of the immune high subtype were higher than those of the immune low subtype in the GEO dataset, except for the type 17 T helper cell, CD56bright natural killer cell, and immature dendritic cell (Fig 2 a). And all the scores of the anti-tumor immunity cells and immune-suppressive cells of the immune high subtype were higher than those of the immune low subtype in the TCGA dataset (Fig 2 f). Moreover, most of the expressions of HLA genes were higher in the immune high subtype than the immune low subtype in the GEO dataset (Fig 2 c), except for *HLA-DPB2*. And all the expressions of HLA genes were higher in the immune high subtype than the immune low subtype in the TCGA dataset (Fig 2 h). The expressions of the immune checkpoint genes such as *CD8A*, *CTLA4*, *CXCL9*, *HAVCR2*, *LAG3*, *PDCD1LG2*, and *PRF1* were higher in the immune high subtype than the immune low subtype in the GEO dataset (Fig 2 d). And the expressions of the immune checkpoint genes such as *CD8A*, *TIGIT*, *CD274*, *GZMB*, *IDO1*, *HAVCR2*, *CXCL9*, *IFNG*, *PDCD1LG2*, *CTLA4*, *TNF*, *GZMA*, *LAG3*, *CXCL10*, *PRF1*, and *PDCD1* were higher in the immune high subtype than in the immune low subtype in the TCGA dataset, while the expressions of the *SIGLEC15* and *TBX2* were lower in the immune high subtype than in the immune low subtype in the TCGA dataset (Fig 2 i). Finally, the immune-related KEGG pathways like antigen processing and presentation, autoimmune thyroid disease, B cell receptor signaling pathway, FC gamma r mediated phagocytosis, the intestinal immune network for IgA production, leukocyte transendothelial migration, natural killer cell mediated cytotoxicity, primary immunodeficiency, and T cell receptor signaling pathway were enriched in the GEO dataset by GSEA analysis (Fig 2 e). While the immune-related KEGG pathways such as antigen processing and presentation, autoimmune thyroid disease, B cell receptor signaling pathway, FC gamma r mediated phagocytosis, the intestinal immune network for IgA production, natural killer cell mediated cytotoxicity, primary immunodeficiency, and T cell receptor signaling pathway, TGF beta signaling pathway, and VEGF signaling pathway were enriched in the TCGA dataset (Fig 2 j). Interestingly, most of the immune cells were highly positively correlated with each other in the two datasets (Supplementary Fig 2).

3. Survival analysis based on immune cells and immune subtypes.

In the GEO dataset, survival analysis indicated that 22 immune cell types had a prognostic value with bladder cancers (Fig 3 a-v). The higher scores of the effector memory CD4 T cell, effector memory CD8 T cell, Immature dendritic cell, monocyte, T follicular helper cell, and type 17 T helper cell in the immune high subtype were associated with a good prognosis of bladder cancers, and the p-value was 0.011, 0.014, 0.003, 0.002, 0.044, and 0.019, respectively. Conversely, The higher scores of the activated CD4 T cell, activated CD8 T cell, activated dendritic cell, CD56dim natural killer cell, central memory CD4 T cell, central memory CD8 T cell, gamma delta T cell, MDSC (Myeloid-derived suppressor cell), memory B cell, natural killer cell, natural killer T cell, neutrophil, plasmacytoid dendritic cell, regulatory T cell, type 1 T helper cell, and type 2 T helper cell in the immune high subtype were related to a shorter survival probability, and the p-value was 0.002, 0.016, 0.003, 0.032, <0.001, 0.002, <0.001, 0.002, 0.036, 0.002, <0.001, <0.001, 0.007, 0.043, 0.029 and 0.013, respectively. Additionally, 16 immune cell types had a prognostic impact on the survival of bladder cancer patients in the TCGA dataset (Fig 3 w-l1). The higher scores of the activated CD4 T cell, activated CD8 T cell, CD56bright natural killer cell, effector memory CD8 T cell, immature B cell, and monocyte in the immune high subtype predicted a longer survival ability, and the p-value was 0.011, 0.004, 0.001, 0.009, 0.018, and 0.028, respectively. In contrast, patients with higher scores of the central memory CD8 T cell, effector memory CD4 T cell, eosinophil, Immature dendritic cell, mast cell, memory B cell, natural killer cell, neutrophil, plasmacytoid dendritic cell, and regulatory T cell had a poor outcome on bladder cancers, and the p-value was 0.002, 0.001, 0.047, 0.030, 0.003, <0.001, 0.042, 0.049, 0.006, and 0.022, respectively. Besides, in the GEO dataset, the percentage of alive and dead patients was 70% and 30% among the immune high subgroup, while among the immune low subgroup it was 73% and 27%, respectively (Fig 3 m1). And in the TCGA dataset, the percentage of alive and dead patients was 55% and 45% among the immune high subgroup, while among the immune low subgroup it was 63% and 37%, respectively (Fig 3 n1). However, the ESTMATEScores of alive and dead patients among the immune high and immune low subgroups were not statistically different in either the GEO or TCGA dataset (Supplementary Fig 3). Finally, Kaplan-Meier survival analysis revealed poorer survival (P=0.011) among the immune high subtype in the gse32894 dataset (Fig 3 o1), while it showed prolonged survival (P=0.039) among the immune high subtype in the gse48276 dataset (Fig 3 p1). However, there were no significant differences in the survival probability between the immune high and immune low subtype in the merged GEO dataset, the gse13507 dataset, and the TCGA-BLCA dataset (Supplementary Fig 4). Similarly, the survival probability of the female and male patients in both the two datasets among the immune high and immune low subgroup was not significantly different (Supplementary Fig 5).

4. Independent prognostic immune cell types in the GEO dataset and the TCGA-BLCA dataset

Univariate Cox regression analysis of the GEO dataset for the OS revealed that activated CD4 T cell, central memory CD8 T cell, gamma delta T cell, type 2 T helper cell, activated dendritic cell, MDSC (myeloid-derived suppressor cells), monocyte, natural killer cell, natural killer T cell, neutrophil, and

plasmacytoid dendritic cell were statistically significant prognostic immune cell types for the OS of bladder cancer patients (P<0.05) (Fig 4 a). On multivariate Cox regression analysis, monocyte was revealed as the independent prognostic immune cell type for the OS of bladder cancer patients, for which the hazard ratio (HR) was 0 (0-0.4) and the p-value was 0.001 (Fig 4 b). Moreover, univariate Cox regression analysis of the TCGA-BLCA dataset for the OS demonstrated that activated CD8 T cell, central memory CD8 T cell, effector memory CD4 T cell, memory B cell, CD56bright natural killer cell, mast cell, and plasmacytoid dendritic cell were statistically significant prognostic immune cell types for the OS of bladder cancer patients (P<0.05) (Fig 4 c). By multivariate Cox regression analysis, activated CD8 T cell (P<0.001, HR 0.005(0-0.054)), central memory CD8 T cell (P=0.011, HR 2988.975(6.236-1432610.99)), memory B cell (P=0.025, HR 241.642(1.993-29303.584)), CD56bright natural killer cell (P=0.02, HR 0(0-0.282)) were revealed as the independent prognostic immune cell types for the OS of bladder cancer patients (Fig 4 d).

5. Construction and validation of gene risk score model

Differential expression analysis between the immune high and immune low subtype of the TCGA-BLCA dataset revealed 2403 differentially expressed genes (DEGs), among which 1963 were up-regulated and 440 were down-regulated (Fig 5 a). Heatmap of DEGs was shown in Supplementary Fig 6 a. Then these 2403 DEGs were intersected with 1793 immune-related genes, and 504 genes were found to be overlapped, among which 221 genes were shared by both datasets (Fig 5 b). Heatmap of differentially expressed immune-related genes was shown in Supplementary Fig 6 b. We next screened the prognostic genes (P<0.05) individually within the 221 shared genes in the GEO dataset and the TCGA-BLCA dataset using Univariate Cox regression analysis. As shown in Fig 5 c and d, 56 genes were identified as prognostic genes in the GEO dataset and 51 genes in the TCGA-BLCA dataset. Using Lasso regression analysis of the univariate model (Fig 5 e, f), which were based on the GEO dataset as the training set, we constructed the final gene risk score model, with the following parameters: the lambda.min=0.0225, and the Riskscore=(0.0153)×*S100A9*+(0.0026)×*TNFRSF11B*+(0.0045)×*SLPI*+(-0.3779)×*CXCR6*+(0.0116)×*FCGR3A*+(0.0353)×*CD70*+(0.2527)×*IFI30*+(0.0126)×*PDGFC*+(0.0211)×*EDN1*+(0.0565)×*SPP1*+(0.0927)×*SFTPD*+(0.2113)×*IL18*+(-0.1092)×*SLC40A1*+(-0.1691)×*CD3G*+(0.0308)×*S100A8*+(0.0845)×*IFITM1*+(0.023)×*FCGR3B*+(-0.2586)×*IL15*+(0.0508)×*VEGFC*+(0.0232)×*PTX3*+(0.0454)×*AQP9*+(0.0958)×*GRP*+(0.0481)×*CCL8* (Fig 6 a). The correlations between the 51 prognostic genes in the TCGA-BLCA dataset and corresponding transcription factors were presented in Fig 5 g and h.

Calibration curves of the training cohort and the validation cohort both performed well for the 1-, 3-, and 4-year time points. ROC curves revealed that AUC for the 1-, 3-, and 4-year time points in the training cohort were 0.768, 0.780, and 0.766, respectively. And the AUC for the 1-, 3-, and 4-year time points in the validation cohort were 0.610, 0.577, and 0.561, respectively. The correlations of the 23 risk score genes and the 28 immune cell types in the training set and the validation set were displayed in Fig 6 f-g. The risk curve of the training cohort illustrated with higher risk scores indicating a higher risk of death and the risk score was inversely correlated with OS (the coefficient=-0.25, P=6.3e-08) (Fig 6 h-j). However, in the

validation cohort, the negative correlation was not significant (Supplementary Fig 7). Survival analysis indicated poorer OS in the high-risk group both in the training cohort (P<0.001) and the validation cohort (P=0.045) (Fig 6 k-i). Besides, survival analysis indicated that higher TMB predicted a favorable survival outcome in the validation cohort (P<0.001) (Fig 6 m). Furthermore, the high-TMB & low-risk group showed the best survival, the high-TMB & high-risk group was the second-best, followed by the low-TMB & low-risk group, while the low-TMB & high-risk group showed the worst survival (P<0.001) (Fig 6 n). However, the TMB between the high- and low-risk groups was not significantly different. In the training cohort, The immune checkpoint genes like *CTLA4*, *CXCL9*, *HAVCR2*, *LAG3*, *PDCD1LG2*, and *PRF1* displayed a higher expression in the high-risk group, except *CD8A* (Fig 6 u). Whereas in the validation cohort, the immune checkpoint genes like *TIGIT*, *CD274*, *GZMB*, *IDO1*, *HAVCR2*, *CXCL9*, *PDCD1LG2*, *CTLA4*, *TNF*, *GZMA*, *LAG3*, *CXCL10*, and *PRF1* displayed a higher expression in the high-risk group, while *SIGLEC15* and *TBX2* showed a lower expression in the high-risk group, and the expressions of *CD8A*, *TIGIT*, *IFNG*, and *PDCD1* were not significantly different between the high- and low-risk groups (Fig 6 v).

6. Construction and validation of CGRS model

Univariate cox regression analysis of the GEO dataset (training cohort) demonstrated that age, T stage, and gene risk score model were statistically significant prognostic factors (P<0.05) for OS (Fig 7 a). Multivariate cox regression analysis of the training cohort demonstrated that age (P<0.001, HR 1.039(1.019-1.059)), T stage (P<0.001, HR 1.847(1.555-2.194)), and gene risk score model (P<0.001, HR 2.492(1.800-3.451)) were independent prognostic factors for the OS of bladder cancer patients (Fig 7 b). The nomogram for predicting the probability of OS at 1, 3, and 5 years was presented in Fig 7 c. Each predictor corresponded to a point, and all points summed up to a total point. The total point then corresponded to the survival probabilities. For example, for one patient with a total point of 226, he/she's survival probability for an OS less than 1, 3, and 5 years was 31%, 60.2%, and 75.5%, respectively. The calibration curves of both the training cohort (Fig 7 d) and the validation cohort (Fig 7 f) for 1, 3, and 4 years illustrated that our nomogram model performed well because the calibration curves showed good consistency between the predictions by the nomogram (the solid lines) and actual observations (the dotted line). The ROC curves of the training cohort (Fig 7 e) and the validation cohort (Fig 7 g) also proved the accuracy of our nomogram model. The AUC for 1, 3, and 4 years in the training cohort was 0.850, 0.744, and 0.756, respectively. The AUC for 1, 3, and 4 years in the validation cohort was 0.719, 0.696, and 0.695, respectively.

Discussion

In recent several years, dividing the tumors into immune-hot and immune-cold phenotypes according to their immune infiltrations has attracted considerable attention. Some studies[29] were pursuing to find a therapy that could convert immune "cold" tumors into "hot" ones to be more sensitive to immunotherapy. Therefore, identifying the tumor as an immune "hot" or "cold" category would be of great importance. Here, based on the ssGSEA scores, we had separately classified the bladder cancer samples in the GEO

and TCGA datasets into the immune high and immune low subtypes using hclust and tSNE hierarchical clustering algorithm. Upon the clustering, most scores of immune cell types, as well as the expression levels of HLA genes or immune checkpoint genes, were higher in the immune high group than the immune low one. Likewise, most of the immune-related KEGG pathways overlapped within the two datasets. Therefore, the results in the GEO cohort and the TCGA cohort highly corroborated each other, which could help maximumly distinguish the differences between the immune high and immune low subtypes.

Based on the optimal cutoff for the ssGSEA scores of the 28 immune cell types, each cell type was classified into the immune high and immune low subgroups. From the survival results, we could see that most of the cell types were associated with prognosis (22/28 in the GEO cohort and 16/28 in the TCGA cohort). unexpectedly, most of the cell types within the high score subgroup showed worse prognosis compared with the low score subgroup (16/22 in the GEO cohort and 10/16 in the TCGA cohort). What's more, the percentages of alive patients in the immune low subtype were higher than that in the immune high subtype in both the GEO cohort and the TCGA cohort, although the ESTIMATEScors of alive and dead patients were not significantly different. And the survival probabilities between the immune high and immune low subgroup had no significant difference in both the two cohorts. Upon multivariate analysis, the ssGSEA score of monocyte in the GEO cohort, the ssGSEA scores of activated CD8 T cell, central memory CD8 T cell, memory B cell, and CD56bright natural killer cell in the TCGA cohort were revealed as the independent prognostic factors for the OS of bladder cancer patients. In several solid tumors and hematologic malignancies, absolute monocyte count in peripheral blood has been used as an indicator of risk or a worse prognosis before treatment[30, 31]. Some evidence had suggested that scoring both total T cell and CD8+ T cell counts is a better prognostic factor for colorectal cancer patients[32]. Memory CD8+ T cells are a critical component of protective immunity in many different types of diseases including cancer[33]. The number of memory B cells in immune therapy responders increased greatly and played a role in inhibiting tumor cell proliferation and metastasis[34]. Other evidence had also shown that CD56bright NK cells restrain T cell responses during various autoimmune diseases[35]. All the above researches were consistent with what was obtained in our recent study: HR >1 with central memory CD8 T cell and memory B cell; HR<1 with monocyte, activated CD8 T cell, and CD56bright natural killer cell.

A series of immune-related prognostic gene signatures for bladder cancers have been constructed based on the public datasets obtained from TCGA and GEO databases[36-43]. They usually employed the TCGA-BLCA dataset to build the model and then validated it with a single external GEO dataset. While such results are subject to the limitations of the data used because TCGA is known to be one of the largest available sets[44] and not feasibly matched by the small size of GEO. Thus, we had conducted a Lasso-UniCox regression analysis to build a 23 immune-related gene signature risk score model with a balanced sample size between the TCGA-BLCA and three merged GEO datasets (405 vs 680). The gene signature was further proved to be an independent prognostic factor. What's more, we had evaluated the landscape of the tumor immune microenvironment of bladder cancers in the two cohorts separately. The established gene signature demonstrated good performance and accuracy in the validation cohort and

was well able to successfully classify the patients from both two cohorts into high- and low-risk subgroups. In the TCGA-BLCA cohort, the survival of different stratification combinations of TMB level and risk group was in line with both actual and expected outcomes. Finally, the immune checkpoint is known to be involved in the maintenance of immunologic homeostasis and helps to maintain peripheral tolerance of self-molecules to prevent excess autoimmunity[45]. In our study, all the immune checkpoint proteins in the GEO dataset and most of the immune checkpoint proteins in the TCGA dataset were expressed higher in the high-risk group, which is in line with its increased expression relating to the progression of cancer patients. These integrated bioinformatics analyses could identify independent immune cell types as well as independent gene signatures or clinicopathology characters for the OS of bladder cancer patients, which would be more comprehensive than previous studies.

Ultimately, a nomogram named CGRS model consisting of immune gene risk score signature, age, gender, tumor grade, and tumor stage was established. It showed good and robust predictive ability on prognosis because the predicted survival estimates of the model were upheld in a separate patient cohort. Given the growing popularity of immune checkpoint inhibitors for a variety of cancer treatments, our prognostic signatures and nomogram model might provide prognostic indicators and potential immunotherapeutic targets for bladder cancers.

Conclusions

In summary, our integrated bioinformatic analysis research revealed the prognostic values of a 23 immune-related genes' signature as well as a CGRS model composed of it for bladder cancers. Our findings identify a new prognostic indicator and potential immunotherapeutic target for bladder cancers.

Abbreviations

CGRS: Clinical-Gene-Risk-Score; TME: tumor microenvironment; BCG: Bacille Calmette–Guérin; GEO: Gene Expression Omnibus; TCGA: The Cancer Genome Atlas; TMB: tumor mutation burden; ssGSEA: single sample Gene Set Enrichment Analysis; t-SNE: t-distributed Stochastic Neighbor Embedding; HLA: Human leukocyte antigen; GSEA: Gene set enrichment analysis; OS: overall survival; KEGG: Kyoto Encyclopedia of Genes and Genomes; lasso: Least absolute shrinkage and selection operator; PPI: Protein-Protein Interaction; AUC: area under the curve; ROC: receiver operating curve; HR: hazard ratio; MDSC: myeloid-derived suppressor cells; DEGs: differentially expressed genes.

Declarations

Ethical Approval and Consent to participate

The study was approved by the Ethics Committee of Huazhong Agricultural University (HZAUHU-2020-0013).

Consent for publication

Not applicable.

Authors' contributions

SL, DC, and SW designed, SL, LD, and KD wrote the manuscript. SL and YW conducted the bioinformatics analysis. ZD and SW reviewed, edited and approved the final version of the manuscript. The authors read and approved the final manuscript.

Availability of data and materials

The dataset(s) supporting the conclusions of this article were derived from open public databases TCGA and GEO.

Acknowledgements

We gratefully acknowledge TCGA and GEO databases for the public access to their data.

Funding

This work was supported by the Cancer Research Program of the National Cancer Center (NCC201817B054).

Conflict of interest statement

Not applicable.

References

1. Liang Z, Wang X, Xu X, Xie B, Ji A, Meng S, Li S, Zhu Y, Wu J, Hu Z *et al*: **MicroRNA-608 inhibits proliferation of bladder cancer via AKT/FOXO3a signaling pathway**. *Mol Cancer* 2017, **16**(1):96-96.
2. Liu Y, Huang W, Cai Z: **Synthesizing AND gate minigene circuits based on CRISPReader for identification of bladder cancer cells**. *Nat Commun* 2020, **11**(1):5486-5486.
3. Su H, Tao T, Yang Z, Kang X, Zhang X, Kang D, Wu S, Li C: **Circular RNA cTFRC acts as the sponge of MicroRNA-107 to promote bladder carcinoma progression**. *Mol Cancer* 2019, **18**(1):27-27.
4. Liang Y, Zhu J, Huang H, Xiang D, Li Y, Zhang D, Li J, Wang Y, Jin H, Jiang G *et al*: **SESN2/sestrin 2 induction-mediated autophagy and inhibitory effect of isorhapontigenin (ISO) on human bladder cancers**. *Autophagy* 2016, **12**(8):1229-1239.
5. Grayson M: **Bladder cancer**. *Nature* 2017, **551**(7679):S33.
6. Bahmani B, Gong H, Luk BT, Haushalter KJ, DeTeresa E, Previti M, Zhou J, Gao W, Bui JD, Zhang L *et al*: **Intratumoral immunotherapy using platelet-cloaked nanoparticles enhances antitumor immunity in solid tumors**. *Nat Commun* 2021, **12**(1):1999-1999.

7. Zhang L, Awadalla M, Mahmood SS, Nohria A, Hassan MZO, Thuny F, Zlotoff DA, Murphy SP, Stone JR, Golden DLA *et al*: **Cardiovascular magnetic resonance in immune checkpoint inhibitor-associated myocarditis**. *Eur Heart J* 2020, **41**(18):1733-1743.
8. Kersten K, de Visser KE, van Miltenburg MH, Jonkers J: **Genetically engineered mouse models in oncology research and cancer medicine**. *EMBO Mol Med* 2017, **9**(2):137-153.
9. Sharma M, Khong H, Fa'ak F, Bentebibel S-E, Janssen LME, Chesson BC, Creasy CA, Forget M-A, Kahn LMS, Pazdrak B *et al*: **Bempegaldesleukin selectively depletes intratumoral Tregs and potentiates T cell-mediated cancer therapy**. *Nat Commun* 2020, **11**(1):661-661.
10. Bian X, Xiao Y-T, Wu T, Yao M, Du L, Ren S, Wang J: **Microvesicles and chemokines in tumor microenvironment: mediators of intercellular communications in tumor progression**. *Mol Cancer* 2019, **18**(1):50-50.
11. Zhang C, Zeng Z, Cui D, He S, Jiang Y, Li J, Huang J, Pu K: **Semiconducting polymer nano-PROTACs for activatable photo-immunometabolic cancer therapy**. *Nat Commun* 2021, **12**(1):2934-2934.
12. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA *et al*: **Mutational heterogeneity in cancer and the search for new cancer-associated genes**. *Nature* 2013, **499**(7457):214-218.
13. Vance RE, Eichberg MJ, Portnoy DA, Raulet DH: **Listening to each other: Infectious disease and cancer immunology**. *Sci Immunol* 2017, **2**(7):eaai9339.
14. Kumar D, Gorain M, Kundu G, Kundu GC: **Therapeutic implications of cellular and molecular biology of cancer stem cells in melanoma**. *Mol Cancer* 2017, **16**(1):7-7.
15. Sharma P, Hu-Lieskovan S, Wargo JA, Ribas A: **Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy**. *Cell* 2017, **168**(4):707-723.
16. Chen DS, Mellman I: **Elements of cancer immunity and the cancer-immune set point**. *Nature* 2017, **541**(7637):321-330.
17. Zhou X, Qu M, Tebon P, Jiang X, Wang C, Xue Y, Zhu J, Zhang S, Oklu R, Sengupta S *et al*: **Screening Cancer Immunotherapy: When Engineering Approaches Meet Artificial Intelligence**. *Adv Sci (Weinh)* 2020, **7**(19):2001447-2001447.
18. Hirata E, Sahai E: **Tumor Microenvironment and Differential Responses to Therapy**. *Cold Spring Harb Perspect Med* 2017, **7**(7).
19. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies**. *Nucleic Acids Res* 2015, **43**(7):e47.
20. Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen J, Hu Z, Zalocusky KA, Shankar RD, Shen-Orr SS *et al*: **ImmPort, toward repurposing of open access immunological assay data for translational and clinical research**. *Sci Data* 2018, **5**:180015.
21. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL, Hancock REW, Brinkman FSL, Lynn DJ: **InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation**. *Nucleic Acids Res* 2013, **41**(Database issue):D1228-D1233.

22. Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y, Shin H, Wong SS, Ma J, Lei Y *et al*: **Cistrome: an integrative platform for transcriptional regulation studies**. *Genome Biol* 2011, **12**(8):R83-R83.
23. Hänzelmann S, Castelo R, Guinney J: **GSVA: gene set variation analysis for microarray and RNA-seq data**. *BMC Bioinformatics* 2013, **14**:7.
24. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA *et al*: **Inferring tumour purity and stromal and immune cell admixture from expression data**. *Nat Commun* 2013, **4**:2612.
25. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z: **Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade**. *Cell Rep* 2017, **18**(1):248-262.
26. Lin W, Qiu X, Sun P, Ye Y, Huang Q, Kong L, Lu JJ: **Association of IDH mutation and 1p19q co-deletion with tumor immune microenvironment in lower-grade glioma**. *Mol Ther Oncolytics* 2021, **21**:288-302.
27. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P *et al*: **The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets**. *Nucleic Acids Res* 2021, **49**(D1):D605-D612.
28. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P *et al*: **STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets**. *Nucleic Acids Res* 2019, **47**(D1):D607-D613.
29. Gómez-Aleza C, Nguyen B, Yoldi G, Ciscar M, Barranco A, Hernández-Jiménez E, Maetens M, Salgado R, Zafeiroglou M, Pellegrini P *et al*: **Inhibition of RANK signaling in breast cancer induces an anti-tumor immune response orchestrated by CD8+ T cells**. *Nat Commun* 2020, **11**(1):6335-6335.
30. Dosani T, Covut F, Beck R, Driscoll JJ, de Lima M, Malek E: **Significance of the absolute lymphocyte/monocyte ratio as a prognostic immune biomarker in newly diagnosed multiple myeloma**. *Blood Cancer J* 2017, **7**(6):e579.
31. Bento L, Díaz-López A, Barranco G, Martín-Moreno AM, Baile M, Martín A, Sancho JM, García O, Rodríguez M, Sánchez-Pina JM *et al*: **New prognosis score including absolute lymphocyte/monocyte ratio, red blood cell distribution width and beta-2 microglobulin in patients with diffuse large B-cell lymphoma treated with R-CHOP: Spanish Lymphoma Group Experience (GELTAMO)**. *Br J Haematol* 2020, **188**(6):888-897.
32. Galon J, Pagès F, Marincola FM, Thurin M, Trinchieri G, Fox BA, Gajewski TF, Ascierto PA: **The immune score as a new possible approach for the classification of cancer**. *J Transl Med* 2012, **10**:1.
33. Bae J, Hideshima T, Zhang GL, Zhou J, Keskin DB, Munshi NC, Anderson KC: **Identification and characterization of HLA-A24-specific XBP1, CD138 (Syndecan-1) and CS1 (SLAMF7) peptides inducing antigens-specific memory cytotoxic T lymphocytes targeting multiple myeloma**. *Leukemia* 2018, **32**(3):752-764.

34. Helmink BA, Reddy SM, Gao J, Zhang S, Basar R, Thakur R, Yizhak K, Sade-Feldman M, Blando J, Han G *et al*: **B cells and tertiary lymphoid structures promote immunotherapy response**. *Nature* 2020, **577**(7791):549-555.
35. Schleinitz N, Vély F, Harlé J-R, Vivier E: **Natural killer cells in human autoimmune diseases**. *Immunology* 2010, **131**(4):451-458.
36. Na L, Bai Y, Sun Y, Wang Z, Wang W, Yuan L, Zhao C: **Identification of 9-Core Immune-Related Genes in Bladder Urothelial Carcinoma Prognosis**. *Frontiers in Oncology* 2020, **10**.
37. Dong B, Liang J, Li D, Song W, Zhao S, Ma Y, Song J, Zhu M, Yang T: **Tumor Expression Profile Analysis Developed and Validated a Prognostic Model Based on Immune-Related Genes in Bladder Cancer**. *Frontiers in Genetics* 2021, **12**.
38. Wang Z, Tu L, Chen M, Tong S: **Identification of a tumor microenvironment-related seven-gene signature for predicting prognosis in bladder cancer**. *BMC Cancer* 2021, **21**(1):692.
39. Chen H, Pan Y, Jin X, Chen G: **An immune cell infiltration-related gene signature predicts prognosis for bladder cancer**. *Sci Rep* 2021, **11**(1):16679.
40. Yan Y, Huang Z, Cai J, Tang P, Zhang F, Tan M, Shen B: **Identification of a novel immune microenvironment signature predicting survival and therapeutic options for bladder cancer**. *Aging (Albany NY)* 2020, **13**(2):2780-2802.
41. Chen X, Jin Y, Gong L, He D, Cheng Y, Xiao M, Zhu Y, Wang Z, Cao K: **Bioinformatics Analysis Finds Immune Gene Markers Related to the Prognosis of Bladder Cancer**. *Front Genet* 2020, **11**:607.
42. Luo Y, Chen L, Zhou Q, Xiong Y, Wang G, Liu X, Xiao Y, Ju L, Wang X: **Identification of a prognostic gene signature based on an immunogenomic landscape analysis of bladder cancer**. *J Cell Mol Med* 2020, **24**(22):13370-13382.
43. 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, **9**(8):2774-2790.
44. Noorbakhsh J, Farahmand S, Foroughi Pour A, Namburi S, Caruana D, Rimm D, Soltanieh-Ha M, Zarringhalam K, Chuang JH: **Deep learning-based cross-classifications reveal conserved spatial behaviors within tumor histological images**. *Nat Commun* 2020, **11**(1):6367-6367.
45. Song W, Shen L, Wang Y, Liu Q, Goodwin TJ, Li J, Dorosheva O, Liu T, Liu R, Huang L: **Synergistic and low adverse effect cancer immunotherapy by immunogenic chemotherapy and locally expressed PD-L1 trap**. *Nat Commun* 2018, **9**(1):2237-2237.

Figures

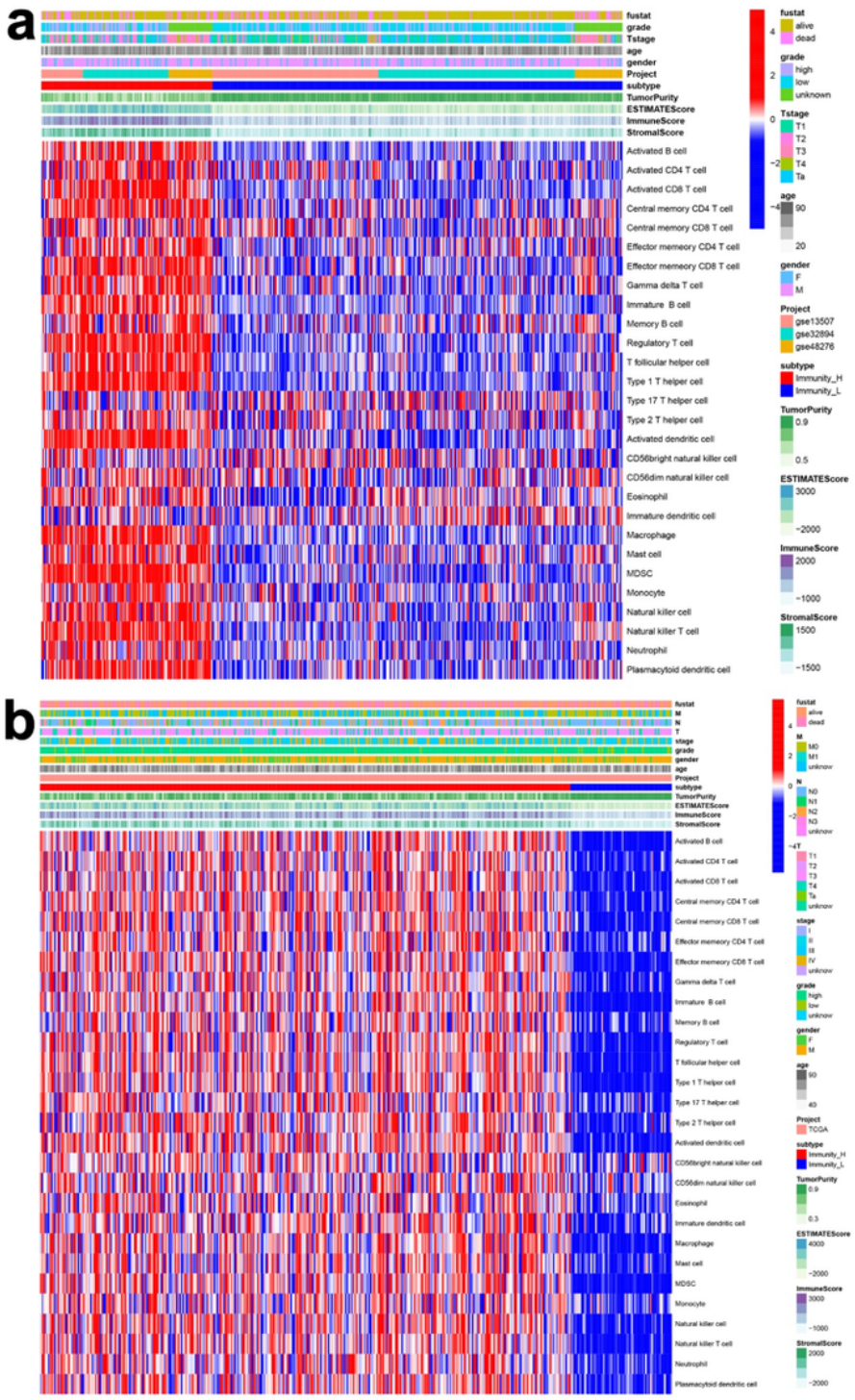

Figure 1

Heatmaps of ssGSEA analysis of GEO (a) and TCGA (b) datasets.

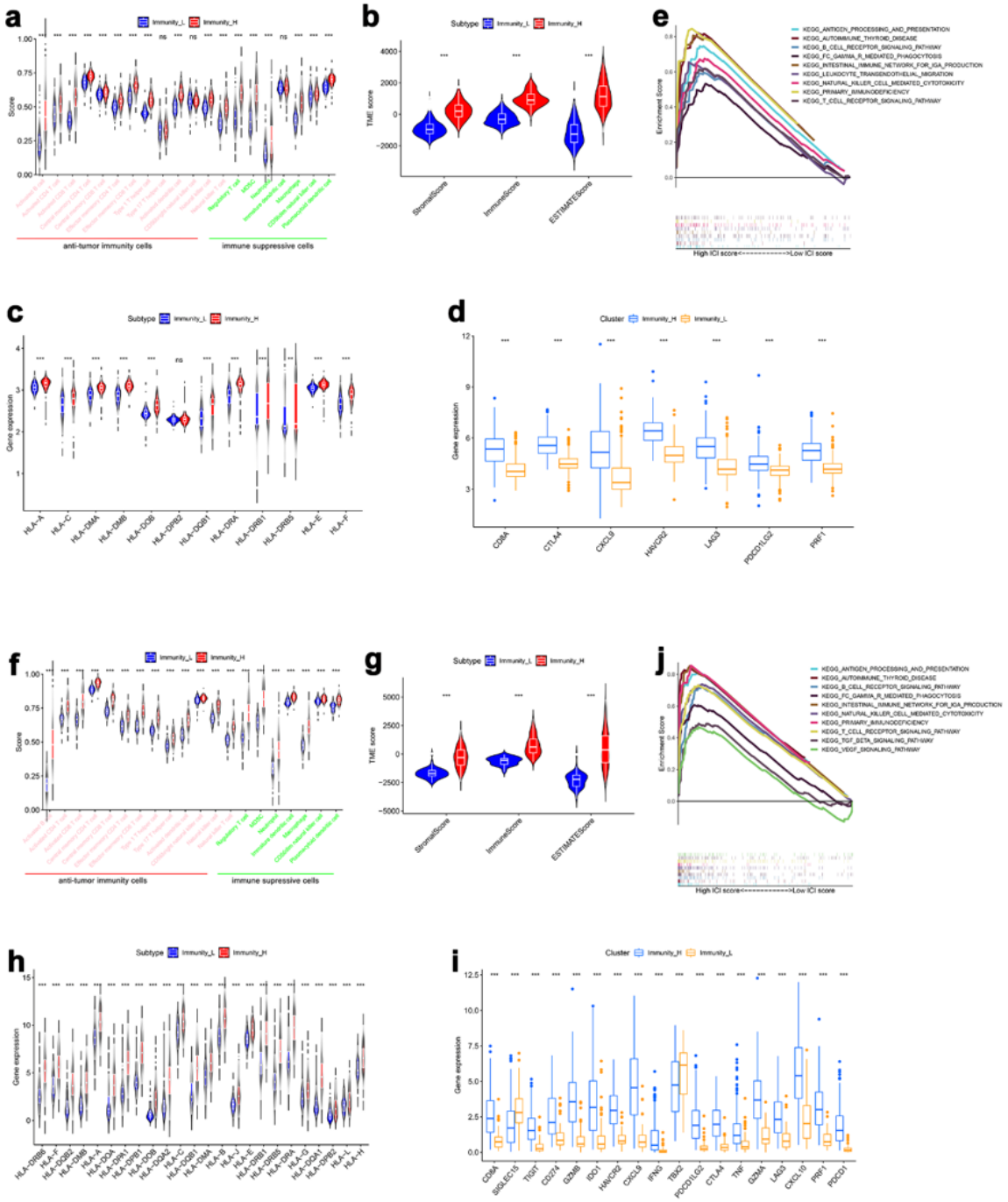

Figure 2

Analysis of the tumor immune microenvironment in the immune high (briefed to Immune_H) and immune low (briefed to Immune_L) subtypes of GEO (a, b, c, d, e) and TCGA datasets (f, g, h, i, j). (a, f) Violin plot of anti-tumor immunity cells and immune-suppressive cells between immune high and immune low subtypes. (b, g) Violin plot of the immune microenvironment scores between immune high and immune low subtypes. (c, h) Human leukocyte antigen (HLA) gene expressions between immune high and

immune low subtypes. (d, i) Boxplots of the immune checkpoint gene mRNA expression levels between immune high and immune low subtypes. (e, j) Gene set enrichment analysis (GSEA) enrichment plot of the immune-related KEGG pathways.

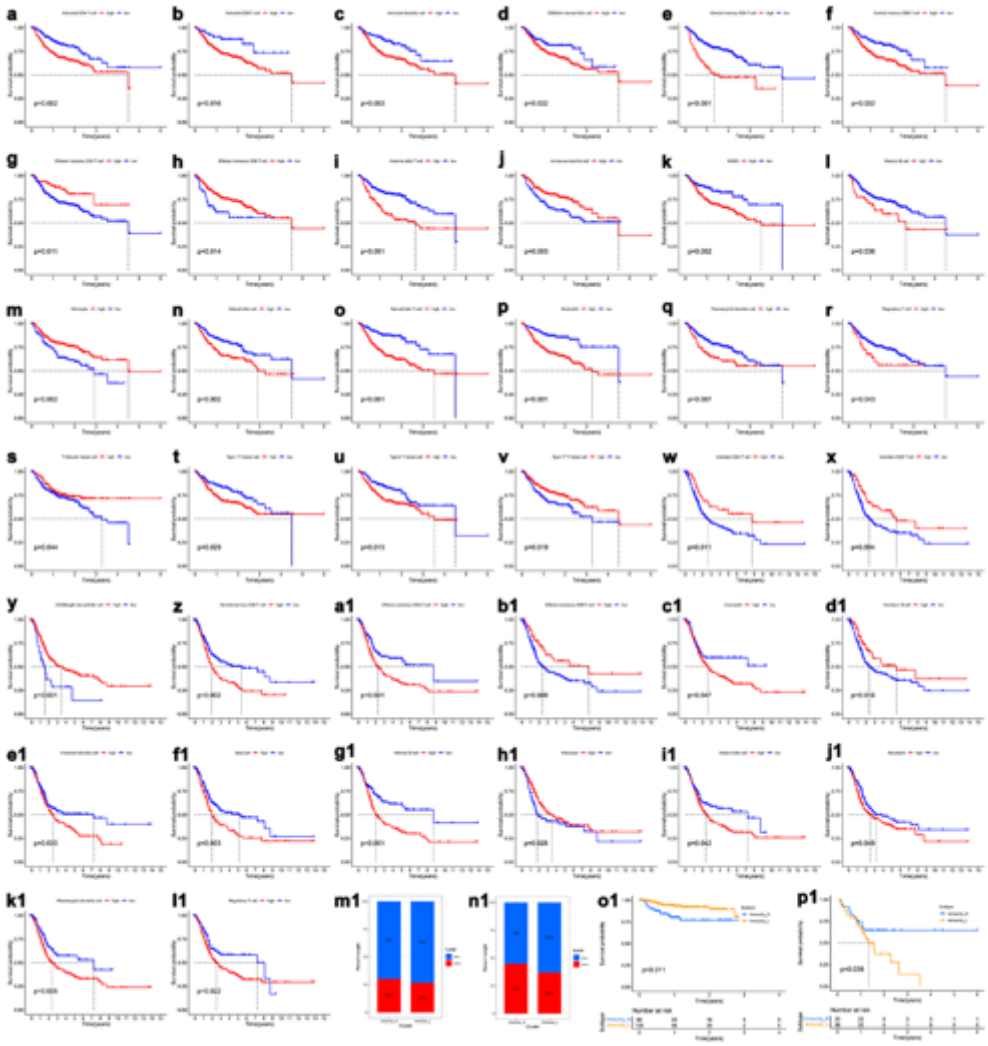

Figure 3

Survival analysis based on the immune cells and immune subtypes. Kaplan-Meier plots of GEO (a-v) and TCGA (w-l1) datasets classified by the median ssGESA scores of immune cells. Barplot of percentages of alive and dead patients between immune high and immune low subtypes in GEO (m1) and TCGA (n1) datasets. Kaplan-Meier plots of the immune high and immune low subtypes in gse32894 (o1) and gse48276 (p1) datasets.

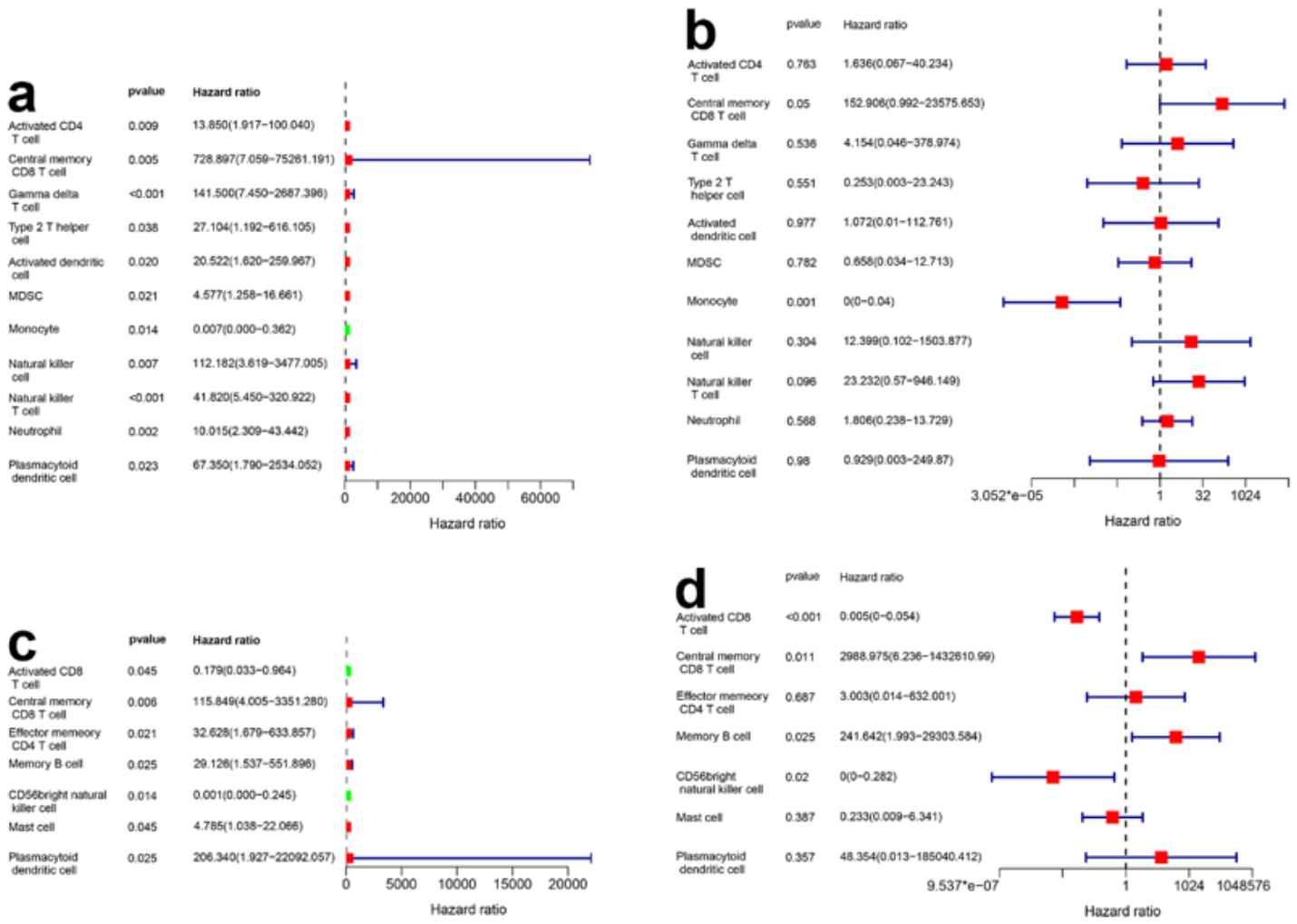

Figure 4

Univariate and multivariate cox regression analysis of GEO (a, b) and TCGA (c, d) datasets. Forest plots of uni- (a,c) and multivariate (b, d) analysis.

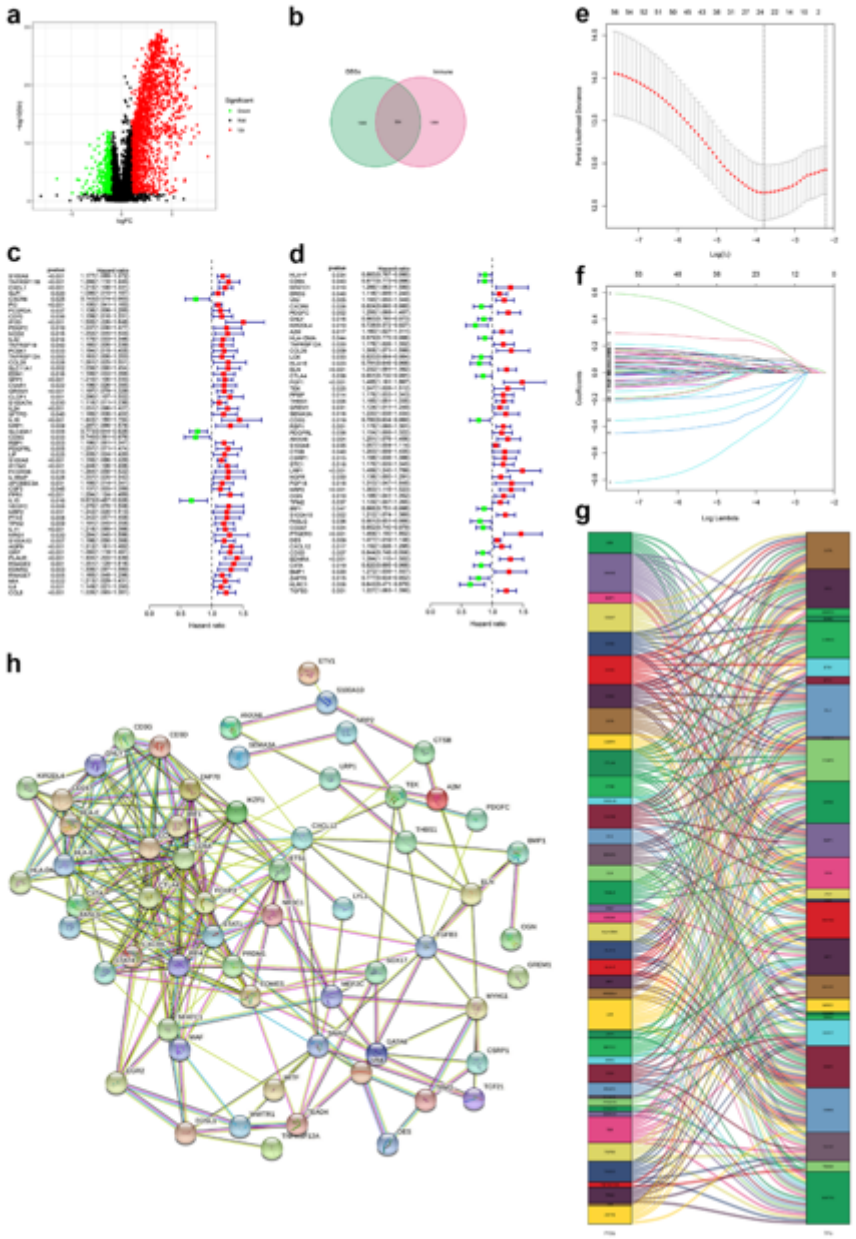

Figure 5

Lasso-cox regression analysis to screen the immune-related prognostic genes and the transcription factor coexpression network of TCGA dataset. Volcano map of differential expressed mRNAs in TCGA-BLCA dataset which compared between the immune high and immune low subgroups (a). Venn plot of The differential expressed genes and the immune-related genes (b). Forest plots of univariate cox regression analysis of the training cohort (c) and validating cohort (d). Lasso regression analysis based on the results of univariate analysis of the training cohort (e, f). Sankey diagram of the prognostic genes in the validating cohort and their correlational transcription factors (g). Protein-Protein Interaction (PPI) Networks of the prognostic genes in the validating cohort and their correlational transcription factors (h).

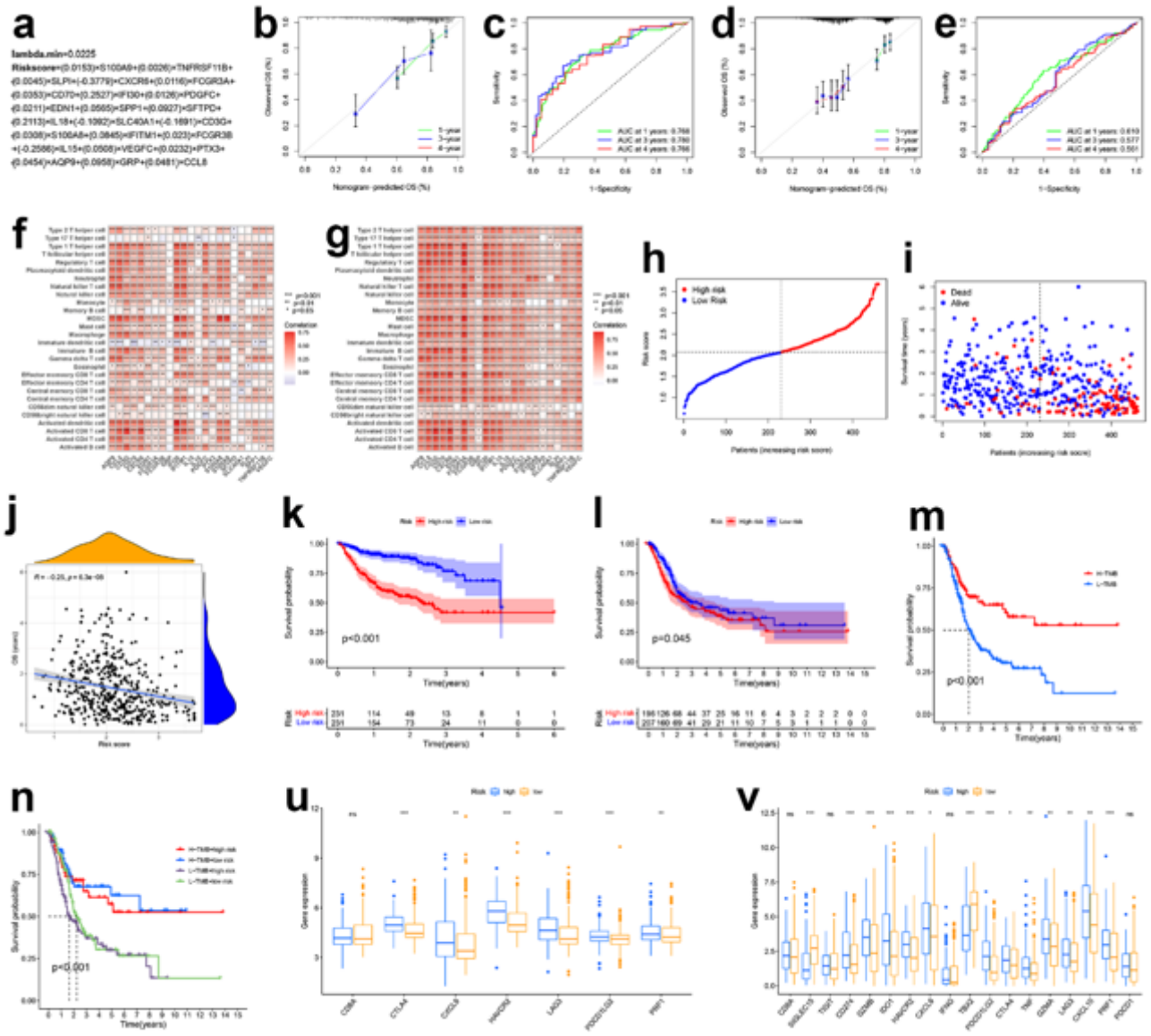

Figure 6

Gene risk score model construction and validation. Model genes and their coefficients (a). Calibration curve of the training cohort (b) and validating cohort (d). ROC curve of the training cohort (c) and validating cohort (e). Correlation heatmaps of model genes and immune cells in the training cohort (f) and validating cohort (g). Riskplot (h), scatterplot (i), and correlation scatterplot (j) of the training cohort. Kaplan-Meier curves of high-risk and low-risk patients in the training cohort (k) and validating cohort (l). Kaplan-Meier curves of high-TMB (briefed to H-TMB) and low-TMB (briefed to L-TMB) patients in the TCGA dataset (m). Kaplan-Meier curves of different TMB-risk subgroups in the TCGA dataset (n). Boxplots of the immune checkpoint gene mRNA expression levels in the training cohort (u) and validating cohort (v).

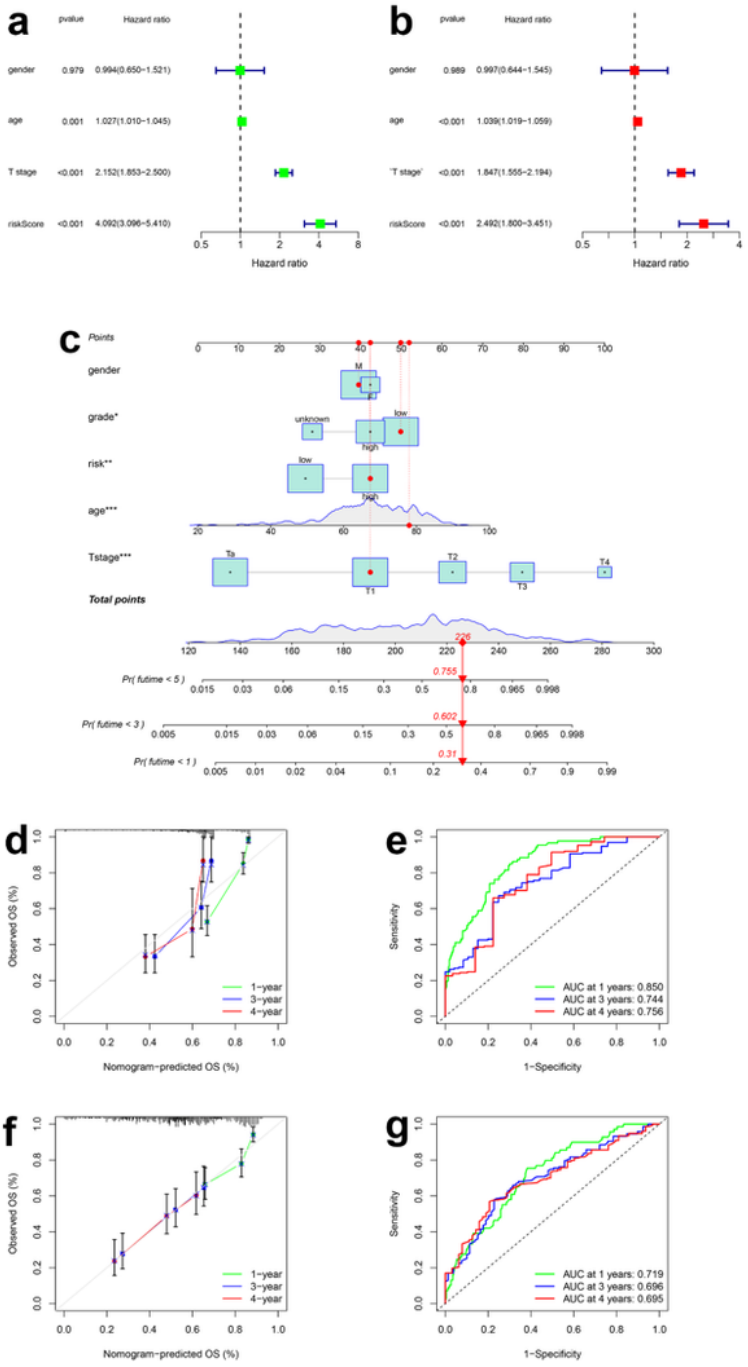

Figure 7

Clinical-Gene-Risk-Score (CGRS) model construction and validation. Forest plots of univariate (a) and multivariate (b) cox regression analysis of clinicopathologic and gene characteristics. Nomogram of the CGRS model (c). Calibration curve of the training cohort (d) and validating cohort (f). ROC curve of the training cohort (e) and validating cohort (g).

Supplementary Files

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

- supplementaryfigures.docx