Identification and validation of a six immune-related gene signature for prediction of biochemical recurrence in localized prostate cancer following radical prostatectomy

Jiaochen Luan[1#], Qijie Zhang[1#], Lebin Song[2#], Yichun Wang[1], Chengjian Ji[1], Rong Cong[1], Qitong Zheng[3], Zhenggang Xu[3], Jiadong Xia[1], Ninghong Song[1,4]

[1]Department of Urology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China; [2]Department of Dermatology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China; [3]Hepatobiliary Center, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China; [4]The Affiliated Kezhou People's Hospital of Nanjing Medical University, Kezhou, China

Contributions: (I) Conception and design: N Song, J Xia; (II) Administrative support: N Song, J Xia; (III) Provision of study materials or patients: N Song, J Xia; (IV) Collection and assembly of data: J Luan, Q Zhang, L Song; (V) Data analysis and interpretation: J Luan, Q Zhang, L Song; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Ninghong Song, MD, PhD. Department of Urology, The First Affiliated Hospital of Nanjing Medical University, 300 Guangzhou Road, Nanjing, China. Email: songninghong_urol@163.com; Jiadong Xia, MD, PhD. Department of Urology, The First Affiliated Hospital of Nanjing Medical University, 300 Guangzhou Road, Nanjing, China. Email: dongjiaxianjmu@sina.com.

Background: Prostate cancer (PCa) is the second lethal heterogeneous cancer among males worldwide, and approximately 20% of PCa patients following radical prostatectomy (RP) will undergo biochemical recurrence (BCR). This study is aimed to identify the immune-related gene signature that can predict BCR in localized PCa following RP.

Methods: Expression profile of genes together with clinical parameters from The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus database (GEO) and the immune-related genes from the Molecular Signatures Database v4.0 were applied to construct and validate the gene signature. The Cox regression analyses were conducted to identify the candidate genes and establish the gene signature. To estimate the prognostic power of the risk score, the time-dependent receiver operating characteristic (ROC) analysis and Harrell's index of concordance (C-index) were utilized. We also established a nomogram to forecast the probability of patients' survival.

Results: A total of 268 patients from the TCGA and 77 patients from GSE70770 and six immune-related genes (*SCIN*, *THY1*, *TBX1*, *NOTCH4*, *MAL*, *BNIP3L*) were eventually selected. The Kaplan-Meier analysis demonstrated that patients in the low-risk group had a significantly longer recurrence-free survival (RFS) compared to those in the high-risk group. In the multivariate Cox model, the signature was identified as an independent prognostic factor, which was significantly associated with RFS (TCGA: HR =5.232, 95% CI: 1.762–15.538, P=0.003; GSE70770: HR =2.158, 95% CI: 1.051–4.432, P=0.036). Moreover, the C-index got improved after incorporating the risk score into original clinicopathological parameters. In addition, the novel nomogram was constructed to better predict the 1-, 3- and 5-year RFS.

Conclusions: This signature could serve as an independent prognostic factor for BCR. Incorporation of our signature into traditional risk classification might further stratify patients with different prognosis, which could assist practitioners in developing clinical decision-making.

Keywords: Localized prostate cancer; biochemical recurrence (BCR); immune-related gene signature; prognosis

Introduction

In the light of the latest cancer statistics reported, Prostate cancer (PCa) is one of the most common aggressive malignant tumors and the second lethal heterogeneous cancer among males worldwide (1,2). Currently, the treatments of PCa include the surgical resection, chemoradiotherapy, endocrinotherapy and immunotherapy (3,4), and the radical prostatectomy (RP) is recommended as the primary therapy for localized PCa. However, approximately 20% of PCa patients following RP will undergo biochemical recurrence (BCR), and eventually progress into castration-resistant prostate cancer (CRPC) (5,6). In accordance with the definition of the European Association of Urology (EAU), BCR refers to consecutive rising prostate-specific antigen (PSA) values above 0.2 ng/mL twice (7). Clinicopathological parameters, including PSA, Gleason score (GS) and TNM stage, have been introduced to predict BCR in PCa patients following RP (8). Nevertheless, patients with the same clinicopathological characteristics may develop into opposite clinical outcomes eventually (9). Therefore, it's of great importance to construct molecular markers to better stratify these patients into high- and low-risk groups.

Recently, increasing evidence illustrated that the immune system, such as immune microenvironment and immune cells (macrophages, neutrophils, T cells, NK cells, DCs, B cells, iNKT cells) establish a critical role in the tumorigenesis and progression (10). PCa can be capable of escaping the immune system through downregulating human leukocyte antigen class I, inducing T-cell apoptosis and secreting immunosuppressive cytokines (11,12). Immunotherapies, such as autologous cellular immunotherapy (Sipuleucel-T) and experimental PCa vaccines (PROSTVAC, DCVAC/PCa, ProstAtak), have made dramatical progress in clinical studies (13). Considering the critical impacts of the immune factors in PCa initiation and progression, immune-related risk signature may serve as the potential candidate complement to original risk stratification so as to improve the prediction of BCR.

In the current study, we concentrated our efforts on performing an immune-related gene prognostic signature based on the whole list of immune genes downloaded from the Molecular Signatures Database v4.0. The RNA sequencing (RNA-seq) patient data of the open-access databases, including the Cancer Genome Atlas (TCGA, https:/portal.gdc.cancer.gov/) and the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.gov/geo/) database

were utilized for our analysis and prediction. Then, we employed the TCGA microarray as a discovery set and further constructed the immune-related gene signature for localized PCa patients undergoing RP, which was systematically identified in the validation set (GSE70770 microarray). Finally, we aimed to figure out the association between this signature and clinicopathological parameters in localized PCa patients. The study would help to provide a stable and eligible genomic prognostic signature, which might serve as a crucial supplement to traditional clinical prognostic features for BCR after RP. We present the following article in accordance with the TRIPOD reporting checklist (available at: http://dx.doi.org/10.21037/tau-20-1231).

Methods

Mining and processing of sample databases and immune-related genes

Expression profile of genes together with clinical follow-up parameters, including age, PSA, stage, GS and outcomes (BCR) for localized PCa patients were acquired from The Cancer Genome Atlas (TCGA) RNA-seq in our retrospective study. Moreover, the whole related candidate data sets were recognized and evaluated by thoroughly seeking another public cohort named Gene Expression Omnibus (GEO) data sets before April 10, 2020. Inclusion and exclusion criteria of data sets were as follows: The eligible data sets included: (I) biospecimens were concentrated from localized PCa patients following RP; (II) each data set contained more than 60 specimens; (III) both clinicopathological characteristics [GS, clinical stage (T) or PSA] and results (BCR) were contained. The exclusion criteria included: (I) databases lacking efficient and sufficient sample data; (II) duplicates of the previously qualified databases; (III) patients also undergoing radiotherapy or chemotherapy in addition to the surgical resection. The list of immune related-genes were extracted from the Molecular Signatures Database v4.0 (14) (Immune response M19817, Immune system process M13664, http://www.broadinstitute.org/gsea/msigdb/index.jsp).

Construction and validation of the risk score model

The analysis charts of our study were illustrated in *Figure 1*. To estimate the relationship between the recurrence-free survival (RFS) and the expression of immune-related genes in TCGA and GEO databases, the univariate Cox

1020

Luan et al. An immune-related gene signature predicting BCR
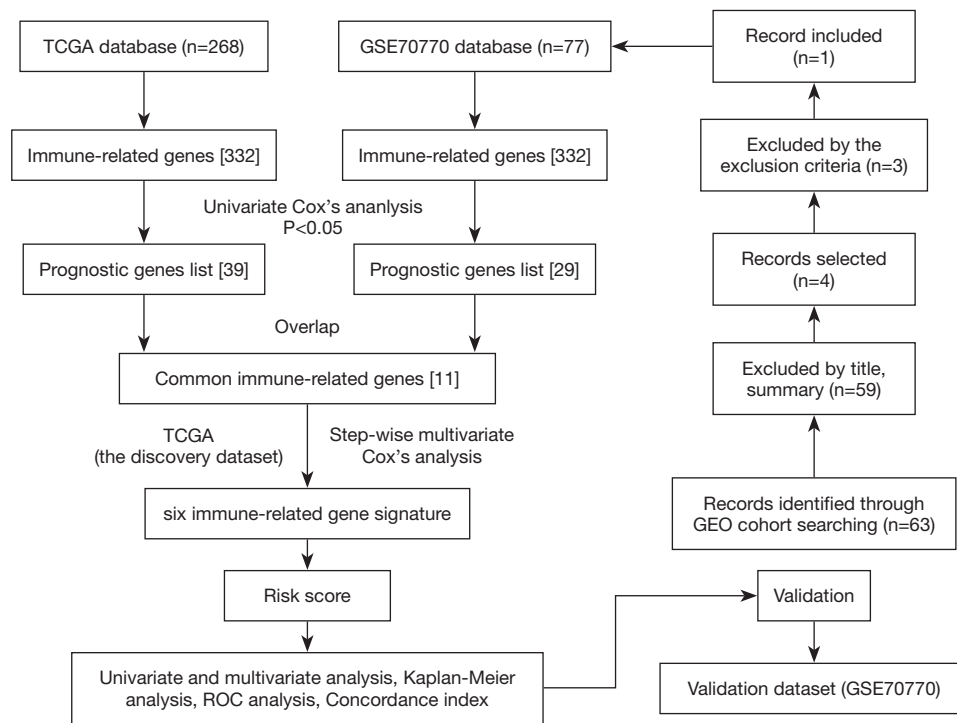
Figure 1 Workflow chart of data generation and analysis.

regression analysis was performed. P=0.05 was selected as a cut-off for the correlation analysis, and the eligible immune-related genes were identified for the construction of prognostic model. Only the immune-related genes which were common in both TCGA and GEO data sets were selected. Subsequently, each candidate gene was enrolled into the step-wise multivariate Cox's regression model to evaluate the independent prognostic contribution to RFS in the discovery data set (TCGA), and only the eligible genes were selected for further construction of the prognostic risk score model. Then, according to the regression coefficients calculated by the multivariate Cox' s model and the expression level of candidate immune-related genes, the following algorithm of a risk score predicting BCR in localized PCa patients undergoing RP was established as follows:

$$\text{risk score} = \sum_{i=1}^{n} \text{expi} * \beta i \qquad [1]$$

More specially, the risk score of each patient was figured out by a linear combination of the expression value of immune-related genes weighted by the regression coefficient (β). βi refers to the regression coefficient of gene i obtained from the discovery cohort, and expi is the expression value of prognostic immune-related genes. By employing the median risk score of the discovery database as the threshold, we separated localized PCa patients into high-risk and low-risk groups for the following analysis.

Establishment and verification of the immune-related gene nomogram

In accordance with the independent prognostic immune-related genes screened from the Cox proportional hazards regression analyses, a compound nomogram was established to forecast the probability of 1-, 3- and 5-year survival of localized PCa patients following RP in the TCGA set. Then, the time-dependent ROC curves were plotted, and the area under the ROC curve was calculated to verify the exactitude of the novel prognostic nomogram. Besides, we utilized the calibration curves to compare the predicted and observed outcomes of the immune-related gene nomogram, among which the 45° line represented the best prediction.

Statistical analysis

Statistical analyses were conducted utilizing SPSS Statistics 26.0 and R software 3.6.3. The Fisher's exact test or Chi-

squared test was applied for categorical variables, and the Wilcoxon rank-sum test or Student *t*-test for continuous variables. The comparison of RFS between high- and low-risk groups was evaluated by the Kaplan-Meier curve utilizing the log-rank test. The univariate Cox regression analysis was performed to estimate the relationships between RFS and the risk score, and the multivariate Cox proportional regression analysis was utilized to evaluate whether the six immune-related gene model was independent of other clinicopathological parameters. The hazard ratio and the regression coefficient were figured out by the Cox proportional hazards regression model. To further assess the predicted accuracy and sensitivity of the risk score, the time-dependent receiver operating characteristic (ROC) analysis was utilized, and the area under curve (AUC) was correspondingly acquired. Harrell's index of concordance (C-index) was applied to calculate and estimate the prognostic ability of the risk score. Two-sided P values less than 0.05 were considered statistically significant for the whole statistical analyses.

Results

Acquisition of the eligible databases

The method of database selection was revealed in *Figure 1*, and eventually, the TCGA data set and the GSE70770 database from the GEO cohort were chosen in our study. The discovery set (TCGA) and the validation set (GSE70770) included 268 and 77 localized PCa patients undergoing RP, separately. The clinicopathological features of both databases were demonstrated in *Table 1*, and there were no obvious differences of clinical characteristics, such as age, clinical stage (T), Gleason score and PSA, between the high-risk and low-risk groups.

Identification and verification of the prognostic immune-related gene signature risk score

A total of 332 immune-related genes were extracted from the Molecular Signatures Database v4.0. Then, the Cox's Analyses were performed in 332 immune-related genes for two data sets. After the univariate Cox's analysis and log-rank test (P<0.05), we singled out 39 and 29 immune-related prognostic genes from the TCGA and the GSE70770 database, respectively, and only eleven common immune-related genes were found in both cohorts (Tables S1,S2 and Figure S1A). Further, to obtain the better prognostic genes, we applied the step-wise multivariate Cox's analysis

among eleven immune-related genes. In the end, six immune-related genes comprising of SCIN, THY1, TBX1, NOTCH4, MAL and BNIP3L were identified, which were associated with RFS of localized PCa patients following RP.

In our study, we developed a six immune-related gene prognostic signature by utilizing the risk score model. The risk score algorithm was calculated based on the regression coefficients and expression values of genes: Risk score = (–0.92247 × expression level of SCIN) + (2.54278 × expression level of THY1) + (1.13792 × expression level of TBX1) + (1.45694 × expression level of NOTCH4) + (–1.45727 × expression level of MAL) + (–3.29561 × expression level of BNIP3L) (Figure S1B and Table S3). Then, patients in the discovery set were divided into high- and low-risk groups and the median risk score was used as a cutoff value. Kaplan-Meier survival curves demonstrated that patients in the low-risk group had a significantly longer RFS compared to those in the high-risk one (P<0.0001) (*Figure 2A*). Those findings were subsequently verified in the validation data set (GSE70770) according to the same algorithm and regression coefficient (β), which displayed the similar results (P=0.0016) as expected (*Figure 2B*). In line with the coefficient value of the six immune-related genes, the risk scores for samples were calculated and ranked in both cohorts. The survival condition of localized PCa patients following RP was signed as the dot plot. As shown in *Figure 3A* and *B*, with the increase of the risk scores, the localized PCa patients in the high-risk group increased, and the number of recurrent persons grew. The heatmap of six immune-related genes expression profiles between high- and low-risk patients was also detailed in this figure.
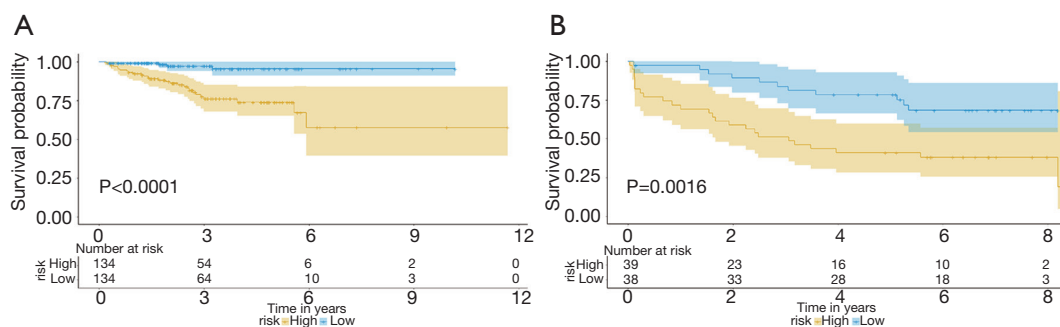
The immune-related gene signature is an independent prognostic factor of RFS in localized PCa patients

The Cox's regression analyses were applied to assess whether the prognostic value of the six immune-related gene signature was an independent risk factor in the discovery and validation sets compared to several clinicopathological factors. The results were shown in *Tables 2* and *3*. In the univariate Cox model of RFS, the high-risk group revealed a 7.453-fold increased and 2.973-fold increased risk of death compared with the low-risk one, respectively (discovery set: 95% CI: 2.613–21.261, P<0.001; validation set: 95% CI: 1.458–6.064, P=0.003) (*Table 2*). In the multivariate Cox regression model, after adjustment for age, PSA, Gleason score and clinical stage (T), the six immune-related gene signature was an independent prognostic factor, which was

1022

Luan et al. An immune-related gene signature predicting BCR

Table 1 Associations between Risk score and clinicopathologic features in the discovery and validation datasets

Variable	Risk score					
	Discovery set (TCGA)			Validation set (GSE70770)		
	Low	High	P	Low	High	P
Age (years)	NA	NA	0.079	NA	NA	NA
≤65	110 (41.0%)	98 (36.6%)	NA	NA	NA	NA
>65	24 (9.0%)	36 (13.4%)	NA	NA	NA	NA
PSA	NA	NA	NA	NA	NA	0.255
<10	NA	NA	NA	28 (36.4%)	24 (31.1%)	NA
10–20	NA	NA	NA	10 (13.0%)	15 (19.5%)	NA
GS	NA	NA	<0.001	NA	NA	0.011
≤6	28 (10.4%)	5 (1.9%)	NA	15 (19.5%)	4 (5.2%)	NA
7	83 (31.0%)	60 (22.4%)	NA	20 (26.0%)	29 (37.6%)	NA
≥8	23 (8.6%)	69 (25.7%)	NA	3 (3.9%)	6 (7.8%)	NA
Clinical stage (T)	NA	NA	0.603	NA	NA	0.175
< T2b	92 (34.3%)	88 (32.8%)	NA	37 (48.0%)	35 (45.5%)	NA
≥ T2b	42 (15.7%)	46 (17.2%)	NA	1 (1.3%)	4 (5.2%)	NA

PSA, prostate-specific antigen; GS, Gleason score; NA, not available.

Figure 2 Kaplan-Meier survival curves of RFS among localized PCa patients following RP from high- and low-risk groups stratified by the six immune-related gene signature from (A) the discovery database (TCGA), and (B) the validation database (GSE70770). RFS, recurrence free survival; PCa, prostate cancer; RP, radical prostatectomy.

significantly associated with RFS (discovery set: HR =5.232, 95% CI: 1.762–15.538, P=0.003; validation set: HR =2.158, 95% CI: 1.051–4.432, P=0.036) (*Table 3*).

Supplementary to original clinical risk stratification with the risk score

To further evaluate the prognostic ability of the six immune-related gene prognostic signature, the time-dependent ROC analyses were applied, and the AUC (area under curve) values were calculated and estimated. The risk score curve value for RFS was 0.834 in the discovery set, which was much higher compared with clinical stage (T) (AUC =0.610) and the Gleason score curve value (AUC =0.687). Similarly, the risk score AUC value in the validation set (AUC =0.798) was also remarkably higher than those connected with PSA (AUC =0.673), clinical stage (T) (AUC =0.553), and Gleason score (AUC =0.679) (*Figure 4*). The
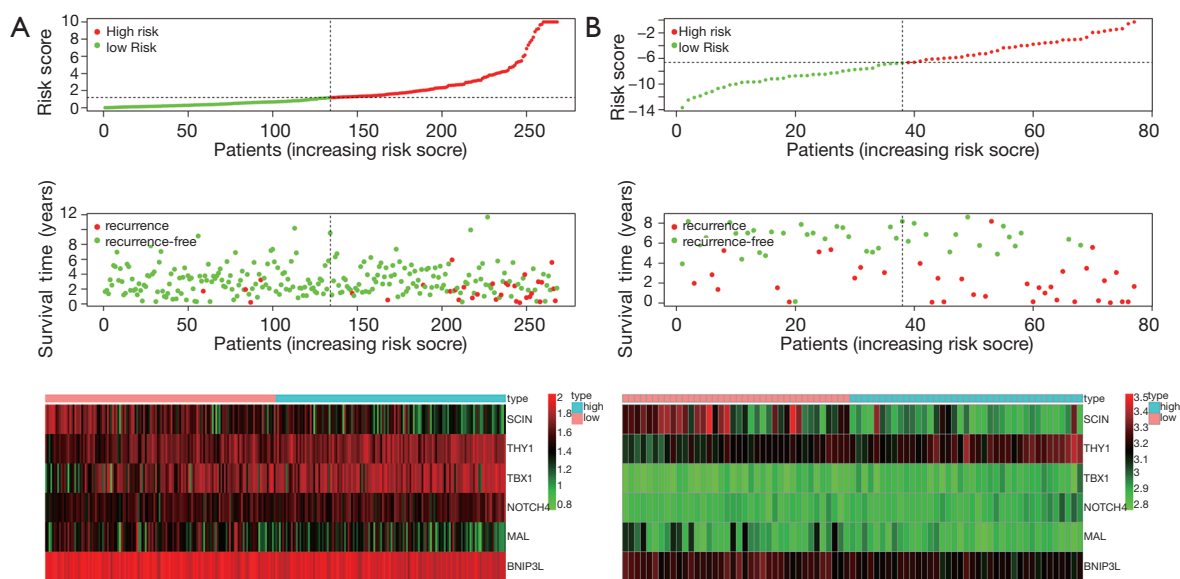
Figure 3 The distribution of risk scores for samples, patients' survival status and the expression heatmap of six different immune-related prognostic genes in both cohorts. (A) The discovery database (TCGA); (B) the validation database (GSE70770).

Table 2 Univariate Cox regression analysis of clinicopathologic factors and Risk score for RFS in the discovery and validation databases

| Variable | Recurrence-free survival | | | |
| | Discovery set (TCGA) | | Validation set (GSE70770) | |
	HR (95% CI)	P	HR (95% CI)	P
Age (years)				
≤65	1.000 (reference)	NA	NA	NA
>65	0.230 (0.055–0.961)	0.044	NA	NA
PSA				
<10	NA	NA	1.000 (reference)	NA
10–20	NA	NA	1.558 (0.796–3.048)	0.195
GS				
≤6	1.000 (reference)	NA	1.000 (reference)	NA
7	1.855 (0.235–14.665)	0.558	13.933 (1.893–102.549)	0.010
≥8	9.661 (1.297–71.969)	0.027	26.333 (3.198–216.853)	0.002
Clinical stage (T)				
< T2b	1.000 (reference)	NA	1.000 (reference)	NA
≥ T2b	2.382 (1.184–4.790)	0.015	1.712 (0.521–5.622)	0.375
Risk score				
Low	1.000 (reference)	NA	1.000 (reference)	NA
High	7.453 (2.613–21.261)	<0.001	2.973 (1.458–6.064)	0.003

Italic P values indicate that the variables were statistically significant. RFS, recurrence free survival; PSA, prostate-specific antigen; GS, Gleason score; NA, not available.

1024

Luan et al. An immune-related gene signature predicting BCR

Table 3 Multivariate Cox regression analysis of clinicopathologic factors and Risk score for RFS in the discovery and validation databases

Variable	Recurrence-free survival			
	Discovery set (TCGA)		Validation set (GSE70770)	
	HR (95% CI)	P	HR (95% CI)	P
Age (years)				
≤65	1.000 (reference)	NA	NA	NA
>65	0.159 (0.038–0.669)	*0.012*	NA	NA
PSA				
<10	NA	NA	NA	NA
10–20	NA	NA	NA	NA
GS				
≤6	1.000 (reference)	NA	1.000 (reference)	NA
7	1.014 (0.124–8.274)	0.990	11.335 (1.525–84.271)	*0.018*
≥8	4.150 (0.527–32.655)	0.176	19.794 (2.363–165.828)	*0.006*
Clinical stage (T)				
< T2b	1.000 (reference)	NA	NA	NA
≥ T2b	1.912 (0.948–3.856)	0.070	NA	NA
Risk score				
Low	1.000 (reference)	NA	1.000 (reference)	NA
High	5.232 (1.762–15.538)	*0.003*	2.158 (1.051–4.432)	*0.036*

Italic P values indicate that the variables were statistically significant. RFS, recurrence free survival; PSA, prostate-specific antigen; GS, Gleason score; NA, not available.
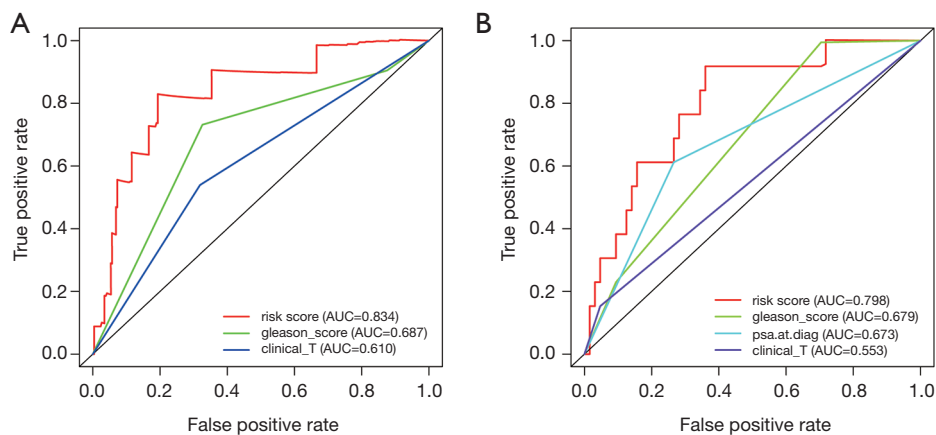
Figure 4 Time-dependent ROC curve analysis of the risk score and clinicopathological characteristics based on the discovery and validation databases. (A) The discovery database (TCGA); (B) the validation database (GSE70770). ROC, receiver operating characteristic; AUC, area under curve.

aforementioned results illustrated that the six immune-related gene signature was a better prospective independent prognostic risk score of RFS in localized PCa patients than other clinical features.

Table 4 Comparison of the prognostic accuracies of risk score and original risk factors [PSA, clinical stage (T) and GS]

Model	C-index
Discovery set (TCGA)	
Risk score	0.692
GS + clinical stage (T)	0.748
GS + clinical stage (T) + risk score	0.784
Validation set (GSE70770)	
Risk score	0.645
PSA + GS + clinical stage (T)	0.718
PSA + GS + clinical stage (T) + risk score	0.748

C-index, Harrell's Concordance index; GS, Gleason score; PSA, prostate-specific antigen.

Additionally, to furnish a more accurate and valid prediction model for BCR in PCa patients after RP, the risk score was integrated with original clinical prognostic factors [clinical stage (T), PSA and Gleason score]. Hence, we constructed the multivariate prognostic modules, including GS + clinical stage (T) + risk score in the discovery set and PSA + GS + clinical stage (T) + risk score in the validation set. In Both data sets, the Concordance index (C-index) of the multivariate prognostic module was not only higher than that of the gene risk model but also was much higher than that of traditional clinical risk factors (*Table 4*). The results underscored that the risk score was a robust complement to clinical risk factors for BCR.

Establishment and verification of the prognostic nomogram based on six hub immune-related genes

In order to provide a clinically quantitative tool for prognosis of localized PCa patients, we integrated the six immune-related genes (SCIN, THY1, TBX1, NOTCH4, MAL, BNIP3L) signature to establish a nomogram plot
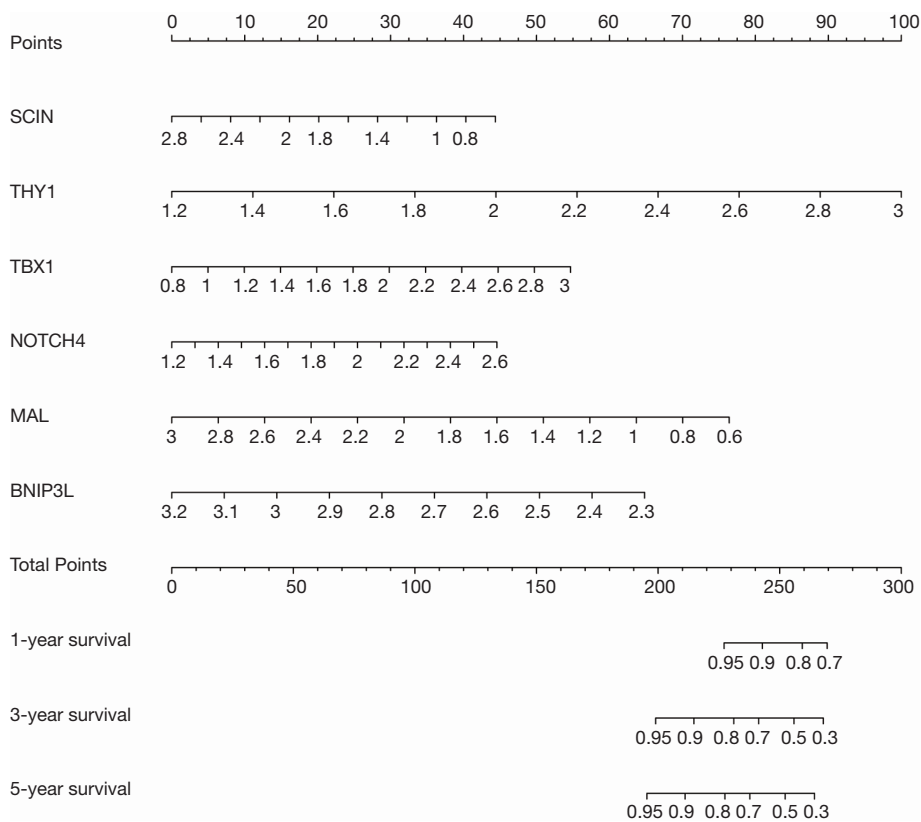
Figure 5 Nomogram based on our established six prognostic immune-related gene signature.
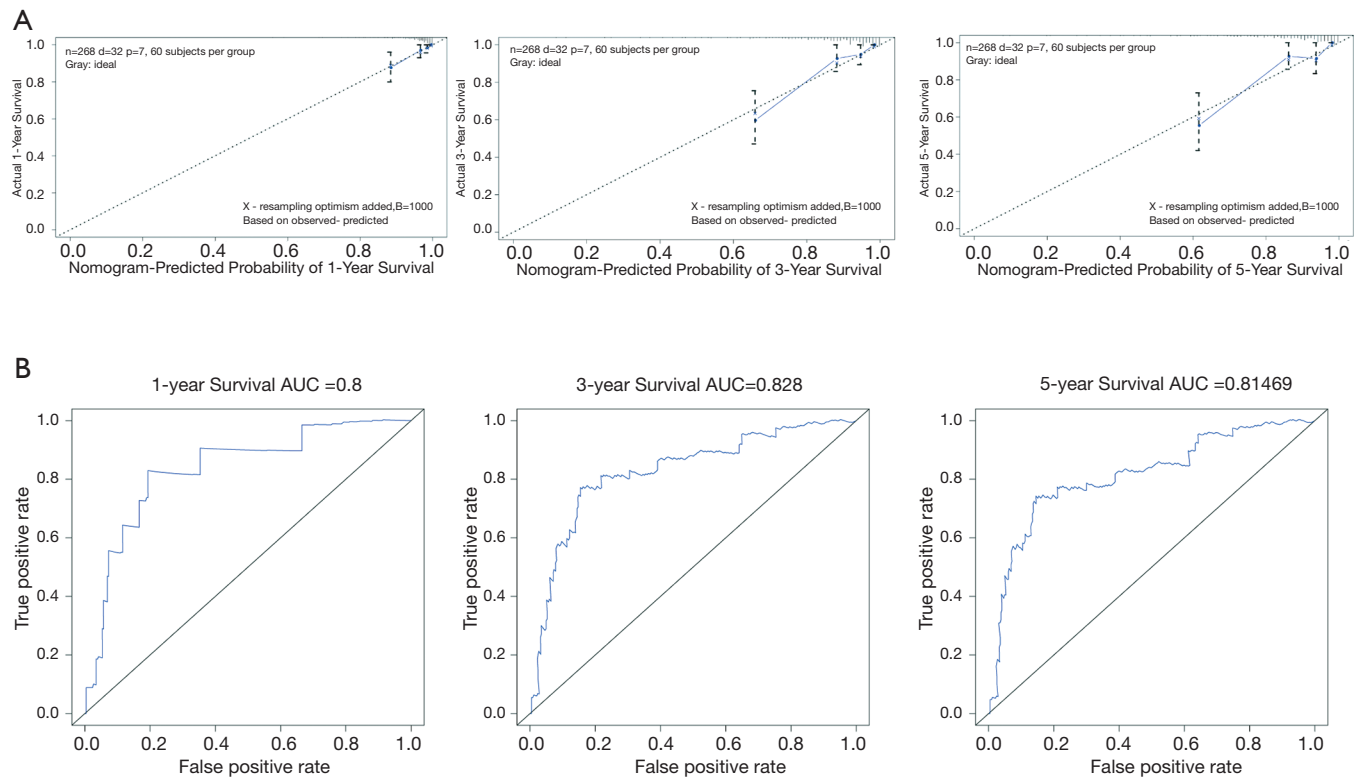
Figure 6 Verification of the nomogram. (A) The calibration curves of the novel nomogram between the predicted survival and actual observed survival; (B) the time-dependent ROC curves of the novel nomogram. ROC, receiver operating characteristic.

(*Figure 5*). A horizontal line was drawn to determine the point of each variable. Then we calculated the total points for each patient by summing the points of all variables and normalize it to a distribution of 0 to 100. This allowed us to calculate the survival probabilities of PCa patients at 1-, 3- and 5-year by plotting a vertical line between the total point axis and each prognosis axis. The calibration curves of the nomogram indicated the strong consistency between the predicted survival and actual observed survival (*Figure 6A*). In addition, the time-dependent ROC analyses were utilized to further assess the power of our novel nomogram, which displayed the sensitivity and specificity of a clinical model. The X-axis meant false positive rate and the Y-axis meant sensitivity. The 1-, 3- and 5-year AUC values of our nomogram were 0.800, 0.828 and 0.815, respectively, suggesting an excellent prognostic ability (*Figure 6B*). The results illustrated that our prognostic nomogram could better predict patients' survival status, which might assist relevant practitioners in developing clinical decision-making for localized PCa patients with BCR.

Discussion

Evidence illustrated that several localized PCa patients undergoing RP would eventually progress into BCR, and the study showed that 10-year survival probability of PCa patients BCR-free was 68% (15). Early BCR had significant association with PCa recurrence, metastasis, and mortality. Therefore, active surveillance of BCR for PCa patients following RP was quite meaningful for further therapies.

So far, surveillance of BCR in PCa patients has been based on clinicopathological characteristics with no strong integrity, so it would be beneficial and valuable to search for several new biomarkers, such as lncRNA, microRNA, mRNA for a better prediction of BCR. For instance, the long non-coding RNAs (lncRNAs) were of great importance for the molecular mechanism of PCa (16). Shao *et al.* (17) found that the five lncRNAs (CRNDE, PRKAG2-AS1, RP11-783K16.13, RP11-727F15.11 and AC013460.1) could serve as a novel prognostic gene signature for predicting BCR in PCa patients after RP, and these lncRNAs were

likely to help PCa patients in targeted molecular therapies. The functions of microRNAs in PCa could not be denied. Zhao *et al.* (18) constructed the BCR scoring system based on the five-microRNA molecular biomarkers, and the combination of these microRNAs remarkably enhanced the ability of prediction for PCa patients experiencing BCR after RP. As for mRNA, Li *et al.* (19) identified that SAMD5 was overexpressed in PCa patients, which was feasible for BCR risk assessment, and the survival analysis gave a clear illustration that the increased expression of SAMD5 mRNA was related to a worse prognosis for PCa patients following RP. Nevertheless, few studies concentrated on the immune-related genes and their effects on predicting BCR in localized PCa patients underwent RP.

In recent years, immune genes established a remarkable role in tumorigenesis, development and metastasis. Moreover, the immune response and immunotherapy showed great significance in tumor biological processes (10). With regard to PCa, the relationship between the immune system and its anti-tumor activity was quite complicated (20). For instance, mastocyte showed negative association with clinical features and BCR, while tumor-associated macrophage (TAM) played a positive role in cancer invasiveness and recurrence following RP (21,22). Referring to the immunotherapy, Sipuleucel-T was the first autologous cellular therapy for PCa, which was recommended by US Food and Drug Administration (FDA) (23).

In our study, we singled out six immune-related genes (*SCIN*, *THY1*, *TBX1*, *NOTCH4*, *MAL*, *BNIP3L*), which was significantly connected with BCR in localized PCa patients undergoing RP. The regression coefficient of three genes (*THY1*, *TBX1*, *NOTCH4*) were additive, while the other three genes (*SCIN*, *MAL*, *BNIP3L*) were subtractive. According to the algorithm, the additive genes were positively corrected with the risk score. Our study showed that the higher the risk scores, the more the patients in high-risk group, and the higher the numbers of recurrent persons. Therefore, the additive genes (*THY1*, *TBX1*, *NOTCH4*) might contribute to the development and progression of PCa. On the contrary, SCIN, MAL, BNIP3L were negatively corrected with the risk score, which might play an antioncogenic role in PCa. Then, we established a six immune-related gene signature that could distinguish samples into the high- and low-risk groups. Next, the survival curves demonstrated that high immune-related genes risk scores were related to poor RFS in both the training cohort (TCGA) and the testing database (GSE70770). Based on the univariate and multivariate Cox

regression models, the six immune-related gene signature was considered as an independent prognostic parameter for BCR in localized PCa patients. The AUC (area under curve) values of the risk score for survival prediction were much higher than those of other clinicopathological prognostic factors, including PSA, clinical stage (T) and GS in both cohorts. Therefore, our risk score was superior to original clinical features and more sensitive and specific of predicting BCR in PCa. Furthermore, we combined the risk score with traditional clinicopathological parameters, and the C-index of the multivariate prognostic module got remarkably improved. The results illustrated that our risk score was supplementary to current clinical risk stratification, which enhanced the veracity of prognostic module for BCR. In addition, a novel nomogram was created to better predict the 1-year RFS, 3-year RFS and 5-year RFS of localized PCa patients after RP based on six immune-related gene signature. The presentation of calibration plot illustrated that there was great conformity between the predicted and observed results, and the time-dependent ROC analysis of our nomogram demonstrated the excellent prognostic power in predicting RFS.

Several studies indicated that these immune-related genes played an important role in tumorigenesis, even in PCa. For example, silencing of SCIN reduced the expression of EGFR and slowed the growth of PCa cells (24). Furthermore, by inhibiting the expression of SCIN, the PCa cell proliferation was suppressed and cell cycle arrested in G0/G1 phase, which was likely to be a promising therapy for PCa patients in the future. As for THY1, it was reported that the overexpression of CD90/THY1 was linked to PCa stromal fibromuscular cells and the tumor microenvironment, which could serve as a tumor marker in the diagnosis and treatment of PCa (25,26). Originating from the T-box family, TBX1 served as a tumor inhibitor in thyroid cancer by regulating PI3K/AKT and MAPK/ERK pathways (27). As a biomarker in many tumors, NOTCH4 established significant roles in cell growth, invasion and apoptosis. Zhang *et al.* (28) revealed that suppressing the expression of NOTCH4 could inhibit the PCa progression and epithelial-mesenchymal transition by stimulating NF-κB. Regarding MAL (29), it was reported that low expression of MAL was associated with poor failure-free survival and it might serve as an independent prognostic signature when estimating the tumor immune infiltrate in classical Hodgkin lymphoma. Moreover, Cheng *et al.* (30) demonstrated that BNIP3Lwas significantly associated with PCa stage via the cell apoptosis way, and this made great contributions to the clarification of

1028

Luan et al. An immune-related gene signature predicting BCR

genomic alterations of PCa aggressiveness. In conclusion, studies as mentioned above revealed the reasonability and accuracy of our six immune-related gene signature in cancer initiation and progression.

It is worth noting that several limitations should be mentioned in spite of our efforts to avert the defect in our study. First, this study was retrospective, thus the accuracy and effectiveness of our six immune-related gene prognostic signature were required to be further verified in some other public cohorts, even in related clinical trials. Second, two databases were selected in our study with no complete clinicopathological parameters, such as PSA or age. Third, experimental validation was needed to pay attention to evaluate the biological functions. Therefore, multi-institutional and well-designed studies are required to verify this immune-related gene signature in the future. Despite the mentioned deficiencies, the predictive value of the risk score for BCR in localized PCa patients following RP cannot be ignored.

Conclusions

In summary, we singled out six prognostic immune genes that could predict BCR in localized PCa patients following RP, namely *SCIN*, *THY1*, *TBX1*, *NOTCH4*, *MAL*, *BNIP3L*. Combining the immune-related gene signature with original clinicopathological parameters might further stratify these patients for diverse prognosis and treatment. However, further prospective studies are required to verify the efficiency and accuracy of this signature and to have a better understanding of the molecular mechanisms in PCa.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at http://dx.doi.org/10.21037/tau-20-1231

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/tau-20-1231). The authors have no other

conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.
2. Ferlay J, Colombet M, Soerjomataram I, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer 2019;144:1941-53.
3. Gunjur A. Personalising the treatment of prostate cancer. Lancet Oncol 2015;16:e592.
4. Denmeade SR, Isaacs JT. A history of prostate cancer treatment. Nat Rev Cancer 2002;2:389-96.
5. Luo HW, Chen QB, Wan YP, et al. Protein regulator of cytokinesis 1 overexpression predicts biochemical recurrence in men with prostate cancer. Biomed Pharmacother 2016;78:116-20.
6. Simmons MN, Stephenson AJ, Klein EA. Natural history of biochemical recurrence after radical prostatectomy: risk assessment for secondary therapy. Eur Urol 2007;51:1175-84.
7. Cornford P, Bellmunt J, Bolla M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part II: Treatment of Relapsing, Metastatic, and Castration-Resistant Prostate Cancer. Eur Urol 2017;71:630-42.
8. Mottet N, Bellmunt J, Bolla M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. Eur

Urol 2017;71:618-29.

9. Li D, Lv H, Hao X, et al. Prognostic value of serum alkaline phosphatase in the survival of prostate cancer: evidence from a meta-analysis. Cancer Manag Res 2018;10:3125-39.

10. Gonzalez H, Hagerling C, Werb Z. Roles of the immune system in cancer: from tumor initiation to metastatic progression. Genes Dev 2018;32:1267-84.

11. Lu X, Horner JW, Paul E, et al. Effective combinatorial immunotherapy for castration-resistant prostate cancer. Nature 2017;543:728-32.

12. Drake CG, Jaffee E, Pardoll DM. Mechanisms of immune evasion by tumors. Adv Immunol 2006;90:51-81.

13. Bilusic M, Madan RA, Gulley JL. Immunotherapy of Prostate Cancer: Facts and Hopes. Clin Cancer Res 2017;23:6764-70.

14. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.

15. Roehl KA, Han M, Ramos CG, et al. Cancer progression and survival rates following anatomical radical retropubic prostatectomy in 3,478 consecutive patients: long-term results. J Urol 2004;172:910-4.

16. Shao N, Zhu Y, Wan FN, et al. Identification of seven long noncoding RNAs signature for prediction of biochemical recurrence in prostate cancer. Asian J Androl 2019;21:618-22.

17. Shao N, Tang H, Qu Y, et al. Development and validation of lncRNAs-based nomogram for prediction of biochemical recurrence in prostate cancer by bioinformatics analysis. J Cancer 2019;10:2927-34.

18. Zhao Z, Weickmann S, Jung M, et al. A Novel Predictor Tool of Biochemical Recurrence after Radical Prostatectomy Based on a Five-MicroRNA Tissue Signature. Cancers 2019;11:1603.

19. Li F, Xu Y, Liu RL. SAMD5 mRNA was overexpressed in prostate cancer and can predict biochemical recurrence after radical prostatectomy. Int Urol Nephrol 2019;51:443-51.

20. Zhao SG, Lehrer J, Chang SL, et al. The Immune Landscape of Prostate Cancer and Nomination of PD-L2 as a Potential Therapeutic Target. J Natl Cancer Inst 2019;111:301-10.

21. Strasner A, Karin M. Immune Infiltration and Prostate

Cancer. Front Oncol 2015;5:128.

22. Lanciotti M, Masieri L, Raspollini MR, et al. The role of M1 and M2 macrophages in prostate cancer in relation to extracapsular tumor extension and biochemical recurrence after radical prostatectomy. Biomed Res Int 2014;2014:486798.

23. Handy CE, Antonarakis ES. Sipuleucel-T for the treatment of prostate cancer: novel insights and future directions. Future Oncol 2018;14:907-17.

24. Wang D, Sun S Q, Yu YH, et al. Suppression of SCIN inhibits human prostate cancer cell proliferation and induces G0/G1 phase arrest. Int J Oncol 2014;44:161-6.

25. Pascal LE, Goo YA, Vêncio RZ, et al. Gene expression down-regulation in CD90+ prostate tumor-associated stromal cells involves potential organ-specific genes. BMC Cancer 2009;9:317.

26. True LD, Zhang H, Ye M, et al. CD90/THY1 is overexpressed in prostate cancer-associated fibroblasts and could serve as a cancer biomarker. Mod Pathol 2010;23:1346-56.

27. Wang N, Li Y, Wei J, et al. TBX1 Functions as a Tumor Suppressor in Thyroid Cancer Through Inhibiting the Activities of the PI3K/AKT and MAPK/ERK Pathways. Thyroid 2019;29:378-94.

28. Zhang Z, Bu X, Yang J, et al. NOTCH4 regulates colorectal cancer proliferation, invasiveness, and determines clinical outcome of patients. J Cell Physiol 2018;233:6975-85.

29. Kelley TW, Pohlman B, Elson P, et al. The ratio of FOXP3+ regulatory T cells to granzyme B+ cytotoxic T/NK cells predicts prognosis in classical Hodgkin lymphoma and is independent of bcl-2 and MAL expression. Am J Clin Pathol 2007;128:958-65.

30. Cheng I, Levin AM, Tai YC, et al. Copy number alterations in prostate tumors and disease aggressiveness. Genes Chromosomes Cancer 2012;51:66-76.

A

B
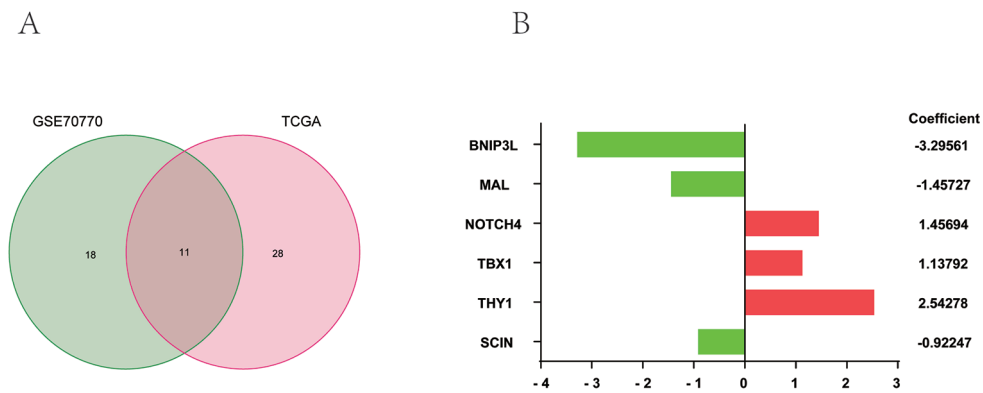
Figure S1 Identification of the candidate prognostic immune-related genes. (A) Venn diagram to identify the common immune-related genes in the TCGA and GSE70770 cohorts; (B) the regression coefficients of six hub immune-related genes based on the step-wise multivariate Cox's analysis.

Table S1 The individual HR and P value of 39 immune-related genes in the TCGA dataset according to the univariate Cox's analysis

ID	HR	HR.95L	HR.95H	P value
TGFB2	0.42708	0.191306	0.953432	0.037862
DPP4	0.23281	0.097197	0.557638	0.001074
CDK6	0.149373	0.061495	0.362833	2.68E-05
SOD1	317.5978	1.209557	83392.86	0.042672
GLMN	38.54605	2.66657	557.1944	0.00737
JAG2	19.95717	2.540245	156.7915	0.004422
SCIN	0.28306	0.130213	0.615323	0.001444
TNFRSF4	3.778192	1.253386	11.38894	0.018219
APLN	3.098408	1.155949	8.304977	0.024573
TAZ	298.5792	21.55416	4136.072	2.14E-05
IFITM2	18.4038	1.561187	216.9502	0.020676
ANXA11	0.001027	3.04E-05	0.034741	0.000128
THY1	7.040337	1.591029	31.15363	0.010113
HDAC7	101.8181	2.921243	3548.81	0.010722
VTN	2.140595	1.033689	4.432808	0.040445
SEMA3C	0.13419	0.04845	0.371664	0.000111
TBX1	3.408151	1.155886	10.049	0.026246
TCF7	0.24632	0.073607	0.824295	0.022996
LTB4R	11.55419	2.866285	46.57573	0.000581
C2	0.137101	0.043855	0.428609	0.000634
POU2F2	0.13759	0.033967	0.557342	0.005453
IFI6	4.727716	1.062073	21.04498	0.041449
TREM2	7.267333	1.752814	30.13105	0.006268
DMBT1	0.254605	0.083137	0.779721	0.016589
RFX1	22.85753	1.329324	393.0321	0.031075
FCGR3B	0.321115	0.113371	0.909539	0.03248
HRH2	0.345915	0.141123	0.847894	0.020305
NOTCH4	8.838301	1.738605	44.93004	0.008623
IFITM3	40.09343	1.837964	874.6001	0.018927
MR1	0.281967	0.098429	0.80774	0.018394
MAL	0.241944	0.08877	0.659423	0.005539
BLNK	0.142822	0.034461	0.591926	0.0073
MADCAM1	4.463989	1.217414	16.36847	0.024025
BNIP3L	0.032574	0.002548	0.416418	0.008444
IL31RA	0.200838	0.067398	0.598471	0.003958
TRAF2	70.48369	2.574432	1929.727	0.011737
LTF	0.641518	0.412332	0.998091	0.049018
INHA	21.77129	6.368357	74.42876	9.02E-07
SART1	230.0323	1.205395	43898.36	0.042389

Italic ID indicate that the genes are candidate.

Table S2 The individual HR and P value of 29 immune-related genes in the GSE70770 dataset according to the univariate Cox's analysis

ID	HR	HR.95L	HR.95H	P value
ANXA11	6.64E-05	4.34E-08	0.101519	0.010129861
APOA1	5.03E+08	46.00215	5.5E+15	0.015394713
BCAR1	2.86E-06	2.55E-10	0.0321	0.007302514
BNIP3	3.69E-05	4.30E-08	0.031681	0.003060305
BNIP3L	0.003247	1.06E-05	0.993626	0.049744588
C2	0.001285	1.50E-05	0.110186	0.003378015
CD276	7954.55	1.386208	45645999	0.041959825
CD79B	2.09E-06	8.12E-11	0.053814	0.011605038
CDK6	0.01913	0.001313	0.278674	0.003793904
CHUK	5.96E-05	1.37E-08	0.260445	0.022932327
CXCL13	0.002251	1.37E-05	0.370275	0.019204061
DPP4	0.056733	0.005577	0.577094	0.015330703
FCGRT	747.5088	3.590495	155624.6	0.015129004
HAMP	9.27E-07	5.22E-10	0.001644	0.000273126
IL6ST	4.79E-07	4.20E-13	0.544654	0.04081911
MADCAM1	9.93E+11	8.158414	1.21E+23	0.0339101
MAFB	253.7712	1.755904	36676.17	029122244
MAL	0.000689	3.20E-06	0.148253	0.007896253
MBP	0.000254	7.04E-07	0.091466	0.005847227
MLF1	54481.75	23.49861	1.26E+08	0.005806979
MMP9	0.045893	0.002969	0.709438	0.027406092
NOTCH4	248801.9	123.8624	5E+08	0.001365132
POU2AF1	7.51E-10	7.33E-16	0.000769	0.002926535
S1PR4	3.72E-05	4.94E-09	0.281013	0.025181618
SCG2	7.482214	1.093579	51.19295	0.040253725
SCIN	0.139569	0.019794	0.984131	0.048153013
TBX1	7468.459	6.030822	9248801	0.014108295
THY1	67.63122	1.327771	3444.857	0.035611906
TNFSF13	1.30E-06	2.94E-12	0.573864	0.040983306

Italic ID indicate that the genes are candidate.

Table S3 The regression coefficients and HR of six hub immune-related genes

ID	Coefficient	HR
SCIN	−0.92247	0.397535
THY1	2.54278	12.715
TBX1	1.13792	3.120277
NOTCH4	1.45694	4.292792
MAL	−1.45727	0.232871
BNIP3L	−3.29561	0.037045