Received: 6 February 2021 | Revised: 16 June 2021 | Accepted: 6 July 2021

DOI: 10.1111/cas.15062

ORIGINAL ARTICLE

Cancer Science WILEY

Novel immune-related signature for risk stratification and prognosis in prostatic adenocarcinoma

Hai-Bo Zhao[1,2,3,4] | Yan-Ru Zeng[5,6] | Zhao-Dong Han[6,7] | Yang-Jia Zhuo[6,7] | Ying-Ke Liang[6,7] | Chi Tin Hon[8] | Song Wan[9] | Shulin Wu[3,4] | Douglas Dahl[3] | Wei-De Zhong[1,3,6,7,10] | Chin-Lee Wu[3,4]

[1]Guangdong Provincial Institute of Nephrology, Nanfang Hospital, Southern Medical University, Guangzhou, China

[2]Department of Urology, The Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

[3]Department of Urology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA

[4]Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA

[5]Department of Anesthesiology, Guangzhou First People's Hospital, South China University of Technology, Guangzhou, China

[6]Department of Urology, Guangdong Key Laboratory of Clinical Molecular Medicine and Diagnostics, Guangzhou First People's Hospital, South China University of Technology, Guangzhou, China

[7]Department of Urology, The Second Affiliated Hospital, South China University of Technology, Guangzhou, China

[8]Macau Institute of Systems Engineering, Macau University of Science and Technology, Macau, China

[9]Department of Urology, Huadu District People's Hospital, Southern Medical University, Guangzhou, China

[10]Macau Institute for Applied Research in Medicine and Health, Macau University of Science and Technology, Macau, China

Correspondence
Wei-De Zhong, Guangdong Provincial Institute of Nephrology, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China.
Email: zhongwd2009@live.cn

Chin-Lee Wu, Departments of Urology and Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA.
Email: cwu2@mgh.harvard.edu

Funding information
This work was supported by grants from National Natural Science Foundation of China (Grant/Award Number: '81874099', '82072813'), Natural Science Foundation of Guangdong Province (Grant/Award Number: '2020A1515010473'), Science and Technology Project of Bureau of Health in Guangzhou Municipality (Grant/Award

Abstract

A substantial proportion of prostatic adenocarcinoma (PRAD) patients experience biochemical failure (BCF) after radical prostatectomy (RP). The immune microenvironment plays a vital role in carcinogenesis and the development of PRAD. This study aimed to identify a novel immune-related gene (IRG)-based signature for risk stratification and prognosis of BCF in PRAD. Weighted gene coexpression network analysis was carried out to identify a BCF-related module in a discovery cohort of patients who underwent RP at the Massachusetts General Hospital. The median follow-up time was 70.32 months. Random forest and multivariate stepwise Cox regression analyses were used to identify an IRG-based signature from the specific module. Risk plot analyses, Kaplan-Meier curves, receiver operating characteristic curves, univariate and multivariate Cox regression analyses, stratified analysis, and Harrell's concordance index were used to assess the prognostic value and predictive accuracy of the IRG-based signature in the internal discovery cohort; The Cancer Genome Atlas

Abbreviations: AUC, area under the time-dependent ROC curve; BCF, biochemical failure; CI, confidence interval; C-index, concordance index; DASL, cDNA-mediated annealing, selection, extension, and ligation; DFS, disease-free survival; FLRT, fibronectin leucine rich transmembrane; GO, Gene Ontology; HLA, human leukocyte antigen; IRG, immune-related gene; KEGG, Kyoto Encyclopedia of Genes and Genomes; PRAD, prostatic adenocarcinoma; PSA, prostate-specific antigen; RNA-seq, RNA sequencing; ROC, receiver operating characteristic; RP, radical prostatectomy; TCGA, The Cancer Genome Atlas; TIIC, tumor-infiltrating immune cell; TIMER, tumor immune estimation resource; TOM, topological overlap matrix; WGCNA, weighted gene coexpression network analysis.

Zhao and Zeng contributed equally to this work.

Number: '20171A011239'), Guangzhou Municipal Science and Technology Project (Grant/Award Number: '201803040001'), The Science Foundation of Guangzhou First People's Hospital (Grant/Award Number: 'Q2019007', 'Q2019020').

database was used as a validation cohort. Tumor immune estimation resource database analysis and CIBERSORT algorithm were used to assess the immunophenotype of PRAD. A novel IRG-based signature was identified from the specific module. Five IRGs (*BUB1B*, *NDN*, *NID1*, *COL4A6*, and *FLRT2*) were verified as components of the risk signature. The IRG-based signature showed good prognostic value and predictive accuracy in both the discovery and validation cohorts. Infiltrations of various immune cells were significantly different between low-risk and high-risk groups in PRAD. We identified a novel IRG-based signature that could function as an index for assessing tumor immune status and risk stratification in PRAD.

KEYWORDS

biochemical failure, immune-related risk signature, immunophenotype, prostatic adenocarcinoma, risk stratification

1 | INTRODUCTION

According to the IARC, PRAD ranks second in the global incidence of malignancies in men, and the mortality rate of PRAD ranks fifth.[1] With the aging of the population and development of diagnostic techniques, PRAD incidence has been increasing worldwide with a growing trend in younger men. Radical prostatectomy currently remains the primary PRAD therapy.[2] Rates of mortality and major complications have decreased with the improvement of radical prostate surgery techniques. However, 15%-40% of PRAD patients still experience BCF within 5 years after RP,[3] even though most of them are young and have a longer life expectancy.

Currently, the prognostic evaluation of PRAD relies on common indicators, such as Gleason score, PSA level, TNM stage, and surgical margins.[4] However, these existing indicators cannot be applied to accurately predict prognosis and to develop systematic therapeutic strategies for PRAD patients.[5] Postoperatively, the stratification of PRAD patients is of substantial benefit but remains challenging. Thus, identifying novel prognostic evaluation strategies is a critical unmet need to improve treatment outcomes. In this regard, advancements in sequencing technology have led to the growing use of prognostic genetic biomarkers.[6]

Recent studies suggest that the immune microenvironment plays a vital role in carcinogenesis and the development of PRAD[7-9] and is closely associated with their overall survival and DFS. Immunologic therapies for PRAD include chimeric antigen receptor-modified T cells, checkpoint inhibitors,[10] and therapeutic cancer vaccines.[11] However, these therapies are not effective in all patients. Previous studies have suggested that PRAD is not sensitive to immune checkpoint inhibitors, possibly due to the low expression of antigens that promote an immune response in PRAD, and a negative regulatory mechanism that inhibits the role of immune cells in PRAD might exist. However, clinical studies have shown that some prostate cancer patients still benefit from other immunotherapies. Therefore, undiscovered immune mechanisms could affect the prognosis of PRAD. The occurrence of these mechanisms might also be limited to some special sensitive groups. Indeed, genetic signatures have been proposed as useful tools to assess the BCF rate of PRAD patients. Nevertheless, the relationship between immunophenotype and prognosis in PRAD is relatively understudied. Consequently, there is an urgent need for more comprehensive studies focusing on IRGs, which could improve the prognostic value and predictive accuracy of the current assessment system. Risk stratification of patients using immune categories can more effectively screen out specific sensitive risk groups and uncover more significant immune molecular mechanisms, which can improve prospects and support immunotherapy for PRAD.

In this study, we used a cohort of 191 cases after RP with long-term clinical follow-up from the Massachusetts General Hospital and a cohort reported by TCGA to construct and to validate a novel IRG-based signature for predicting the prognosis of PRAD patients. We undertook a comprehensive bioinformatics analysis to assess the correlation between IRG-based signatures and clinical features and explored the potential mechanisms that could induce malignant transformation. In addition, the TIMER database and CIBERSORT algorithm were used to clarify the infiltration of immune cells in PRAD to determine the individual immunophenotype of PRAD patients (Figure S1). Our novel risk signature could contribute to further elucidate the immune-related pathogenesis of carcinogenesis and progression in PRAD patients, and provide new targets for subsequent personalized immunotherapy.

2 | MATERIALS AND METHODS

2.1 | Data collection

Tissue samples for the discovery cohort (n = 210) were obtained from patients undergoing RP between September 1993 and December 1995 at the Massachusetts General Hospital. The clinical data, including demographic information, Gleason score, surgical margin status, pT stage, time to BCF, metastasis status, and time to

metastasis, were collected from patient medical records. Biochemical failure was defined as detectable serum PSA after RP. The detection threshold was based on the generally accepted standard at the time. From 1993 to 2001, the detection limit dropped from 0.5 to 0.1 ng/mL. The majority of patients on record (87%) underwent a second confirmatory PSA examination. Metastasis events were determined by metastatic lesions observed on imaging. Exclusion criteria were as follows: adjuvant radiotherapy or neoadjuvant therapy before BCF or lymph node metastasis present during RP. Patients with missing follow-up clinical information and unavailability of tumor tissue were excluded. The median follow-up time was 70.32 months (range, 1-138 months), from 1994 to 2005. Based on relevant scientific reports and reviews, 1536 genes related to the initiation and progression of PRAD were selected for expression profile analysis.[12] To identify a prognostic genetic signature for prostate cancer patients, we first undertook gene expression profiling on formalin-fixed and paraffin-embedded RP tissue samples from the discovery cohort using a customized 1536-gene DASL assay (Illumina). For the DASL bead microarray study, total RNA was isolated and purified using the High Pure RNA Isolation Kit (Roche), and cDNA was generated using the single-use cDNA Synthesis Kit (Illumina). cDNA sequences for the 1536 genes were generated, and a custom DASL assay panel was synthesized by Illumina. The DASL assay was carried out at the Mayo Clinic College of Medicine Genotyping Shared Resource. The DASL array dataset is available online (https://www.ncbi.nlm.nih.gov/projects/geo; accession no. GSE44353). Of the 210 patients, 19 were excluded due to quality control issues (insufficient RNA quantity or quality) in DASL. This study was approved by the Institutional Review Boards of Massachusetts General Hospital. In the validation cohorts, RNA-seq data and clinical information of 494 PRAD cases were downloaded from the TCGA database (http://tcga-data.nci.nih.gov). The data format was HTSeq-FPKM. Data preprocessing was carried out before executing the analyses. The R package "SVA" was used to correct for plate batch effects. The discovery cohort was used to construct a prognostic scoring system model, and the validation cohort (TCGA database) was used to validate the efficacy of this model.

2.2 | Weighted gene coexpression network analysis

Weighted gene coexpression network analysis was undertaken with the database from the discovery cohort using the WGCNA R package.[13] The group samples were hierarchically clustered according to gene expression levels. Outlier samples were excluded to ensure the relative stability of the gene network. Scale-free gene coexpression networks were constructed with a min-Module Size of 30 and merge-Cut Height of 0.25. The network met the scale-free condition while $R^2 > .8$. The soft-threshold power was set to six according to the mean degree of connectivity and scale-free fit index. The TOM was computed using the adjacency matrix, which was subsequently clustered according to the dissimilarity of gene connectivity. Different gene modules were defined using the dynamic tree-cutting algorithm, and the correlations between principal components of different gene modules and clinical features (age, Gleason score, surgical margin status, pT stage, BCF status, and metastasis status) were calculated. The BCF-related modules were defined based on *P* values ($P < .05$), which were selected for further analysis.

2.3 | Extraction of IRGs

The list of IRGs was downloaded from the InnateDB database[14] (https://www.innatedb.com/). Overlapping 152 IRGs from BCF-related modules and the InnateDB database were selected for further analyses.

2.4 | Gene Ontology and KEGG analyses

To determine the function of IRGs in the BCF-related modules, GO and KEGG analyses were carried out. The KEGG pathway and GO terms with $P < .05$ were considered significant pathway enrichments.

2.5 | Construction and validation of an IRG-based signature from BCF-related modules

The association between BCF and the expression of 152 IRGs was assessed using univariate Cox regression in PRAD patients. The IRGs in the selected module with *P* values less than .05 were identified as BCF-related IRGs. Among them, 50 IRGS were correlated with recurrence at $P < .05$. Random survival forest analysis was used to determine the variable importance factors distinguishing gene expression associated with patient survival using the randomForestSRC package in R. Multivariate stepwise Cox regression analysis was used to establish the IRG-based signature in PRAD. Each patient's risk score in both cohorts was calculated using the formula:

$$risk\ \text{score} = \Sigma \text{expgenei}^{*}\ \beta i.$$

where expgene, *i*, and βi represent the gene expression level, the number of signature genes, and the coefficient index, respectively. In total, 191 samples from the discovery cohort were used as the internal set, and 494 samples from the validation cohort were used as the external set. Based on the risk score (median cut-off value), all patients were further divided into low-risk and high-risk groups in both cohorts. Disease-free survival was calculated using the Kaplan-Meier method. The associations between the IRG-based signature and clinical parameters were evaluated with a χ^2 test. Univariate Cox regression analysis was used to identify the parameters related to the BCF of PRAD. In order to further explore whether the prognostic impact of the IRG-based signature was independent of other clinical parameters, multivariate Cox regression and stratified analyses were carried out. In order to determine the predictive accuracy of the IRG-based signature, the AUC was applied to validate the performance of the classifier.

2.6 | Construction and validation of nomogram

In the discovery and validation cohort, Harrell's C-index was used to assess the accuracy of the prognostic model based on signature-based risk scores and other clinical prognostic factors.[15] The establishment of the nomogram was integrated with signature-based risk scores and other clinical prognostic factors to predict the patients' prognosis using the "rms" package in R.[16] Calibration curves were used to evaluate the discrimination and accuracy of the nomogram.

2.7 | Tumor Immune Estimation Resource database analysis

The TIMER database[17,18] was utilized with TCGA RNA-seq data to assess TIICs in PRAD patients. Tumor-infiltrating immune cells included $CD8^+$ T cells, neutrophils, B cells, $CD4^+$ T cells, macrophages, and dendritic cells.

2.8 | Human leukocyte antigen-related gene analysis

RNA sequencing data of HLA-related genes of 494 PRAD cases were downloaded from the TCGA database (http://tcga-data.nci.nih.gov). The database was used to evaluate the expression of HLA-related genes between different risk groups. Variables between groups were compared using the Wilcoxon test.

2.9 | Calculation of relative abundance of immune cell subtypes

The abundance of 22 subtypes of TIICs in PRAD cases from the TCGA database was calculated using the CIBERSORT algorithm. The CIBERSORT algorithm is a deconvolution algorithm that can infer 22 subtypes of TIICs and harnesses the characteristics of specific gene expression to calculate the relative score of each immune cell subtype.[19,20] Data with a *P* value less than .05 after CIBERSORT were selected for subsequent analysis to increase the accuracy of the deconvolution algorithm. Data analysis was undertaken using the CIBERSORT package in R software. The Wilcox test was utilized to compare the relative abundance of the TIICs between high-risk and low-risk groups.

2.10 | Statistical analyses

All statistical analyses were carried out using R software. Comparison of Kaplan-Meier survival curves between low-risk and high-risk groups was undertaken with log-rank tests using the "survminer" package. Random forest analysis was carried out using the "randomForestSRC" package. Univariate and multivariate Cox regression analyses were used to ascertain independent prognostic factors. The associations of clinicopathologic factors between low-risk and high-risk groups were assessed using χ^2 tests. Variables between groups were compared using the Wilcoxon test. *P* values less than .05 were considered statistically significant.

3 | RESULTS

3.1 | Identification of BCF-related modules by construction of WGCNA

Weighted gene coexpression network analysis was undertaken on 191 cases after RP from the discovery cohort (Figure 1A). Weighted gene coexpression network analysis is a systematic biological method that uses gene expression data to establish a scale-free network. The scale-free R^2 (R^2 = .84) in WGCNA was determined to build the network to ensure that the coexpressed network was consistent with the scale-free network. The expression matrix was transformed into an adjacency matrix, and the adjacency matrix was subsequently transformed into a TOM. Eight gene modules were identified. The gray module represents the genes that were not assigned to any of the modules (Figure 1B). To investigate the relationships between modules and clinical traits, a module-trait analysis was carried out to select the key modules that were significantly associated with BCF (Figure 1C). The turquoise module, which contained 422 genes, showed the strongest correlation with BCF. The genes in the turquoise module were sorted for further analysis.

3.2 | Identification of IRGs from the BCF-related module

In total, 152 overlapping IRGs were extracted from the turquoise module and the InnateDB database. The GO analysis revealed that these IRGs were predominantly concentrated in biological processes such as the cell cycle, cell adhesion, and immune responses. The KEGG analysis revealed that the most significant enrichment pathways included focal adhesion, extracellular matrix-receptor interaction, and a p53 signaling pathway (Figure S2). The protein-protein interaction network of 152 IRGs was constructed using STRING, and presented with Cytoscape software[21] (Figure 1D).

3.3 | Construction and validation of prognostic IRG-based signature

Univariate Cox regression was carried out on 152 IRGs, and 50 IRGs were identified as BCF-related IRGs in PRAD ($P < .05$ as the criterion). Random forest and multivariate stepwise Cox regression analyses were used to identify the best gene-set for the prediction of BCF. In total, five IRGs were selected as the most significant gene-set associated with BCF, namely *BUB1B*, *COL4A6*, *FLRT2*, *NDN*, and *NID1*. The full gene names and functions of the five IRGs are shown

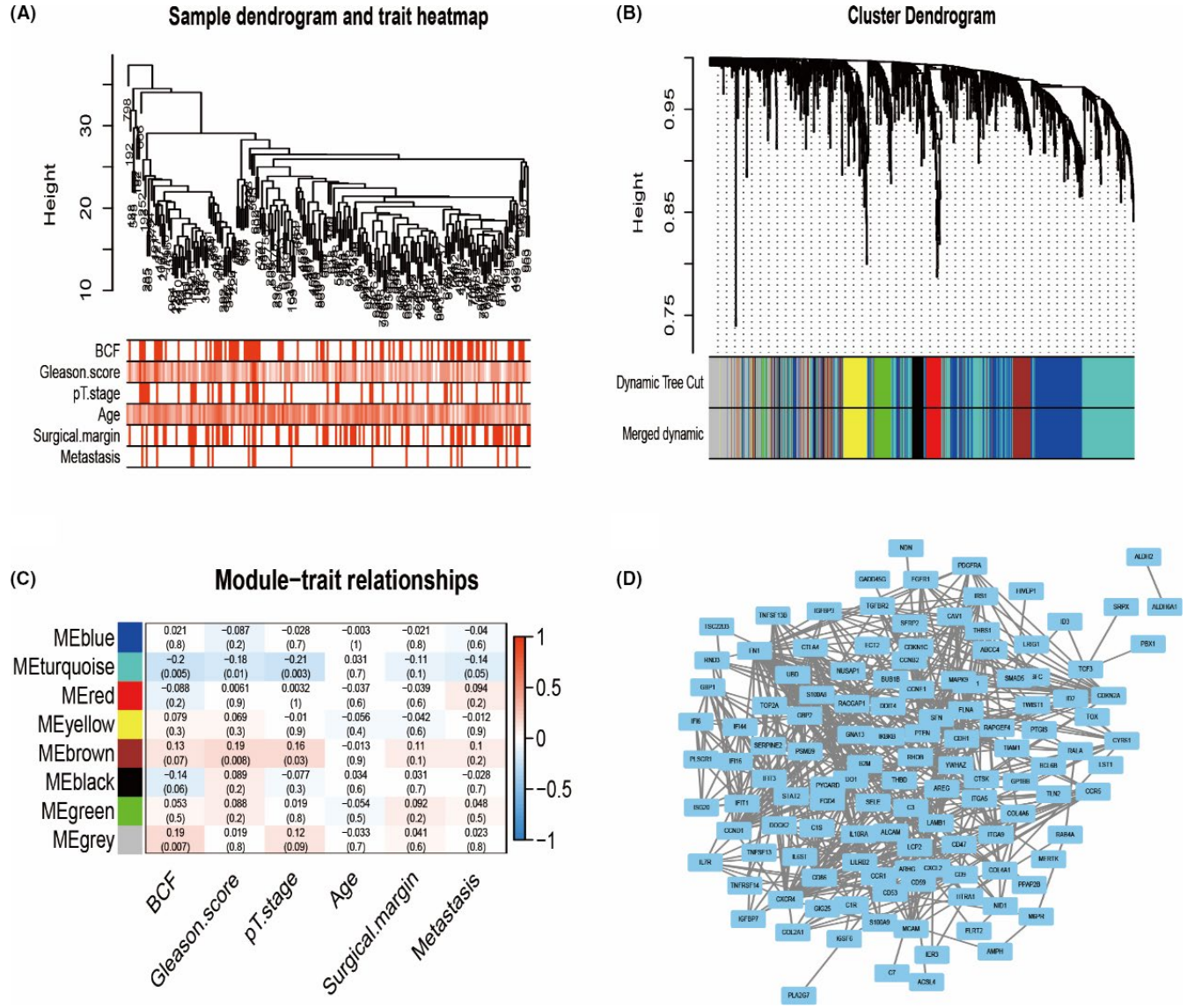

FIGURE 1 Network construction of weighted coexpressed genes and associations with clinical traits in prostatic adenocarcinoma. A, Hierarchical clustering tree of the discovery cohort; dendrogram tips are labeled with the discovery cohort unique name. In the hierarchical dendrogram, lower branches correspond to higher coexpression. Among them: biochemical failure (BCF) (no, 0; yes, 1), Gleason score (5-10), pT stage (≤pT2, 1; >pT2, 2), age (45-78 years), surgical margin (no, 0; yes, 1), metastasis (no, 0; yes, 1), The level of color increases as the value grows. B, The branches of the cluster dendrogram correspond to eight different gene modules based on topological overlaps. Each piece of the leaves on the cluster dendrogram represents a gene. C, Relationships between modules and clinical traits. Shading behind the number represents the strength of the correlation between the gene modules and clinical traits, which varies from blue to red. Each column corresponds to a clinical trait. D, Visualization of the coexpression network of the immune genes in the turquoise module. The more edges, the more significant the gene is. Based on weight, not all genes are represented

in Table S1. Before building a risk score model, the "SVA" software package was applied to normalize and remove batch effects from the discovery cohort and validation set. Each patient's risk score in both cohorts was calculated using the formula:

$$\text{Risk score} = (0.51871 \times \exp_{BUB1B}) + (0.54655 \times \exp_{NDN}) + (0.41306 \times \exp_{NID1}) + (-0.35744 \times \exp_{COL4A6}) + (-1.37699 \times \exp_{FLRT2})$$

After the risk scores were calculated, PRAD patients in both cohorts were separated into low-risk and high-risk groups based on the median cut-off point of the risk score. The risk score distribution of patients, BCF status in different risk groups, and expression profiles of five IRGs in both cohorts are shown in Figure 2A. Patients in the high-risk group had a higher BCF rate than patients in the low-risk group (Figure 2A). Kaplan-Meier analysis revealed that the high-risk group had a poorer DFS rate relative to that of the low-risk group ($P < .01$; Figure 2B). Time-dependent ROC curve analyses revealed that the IRG-based signature had strong predictive ability in both cohorts. In the discovery cohort, the values of 1-year, 3-year, and 5-year AUCs were 0.777, 0.730, and 0.726, respectively. In the validation cohort, the values of 1-year, 3-year, and 5-year AUCs were 0.716, 0.690, and 0.633, respectively (Figure 2C).

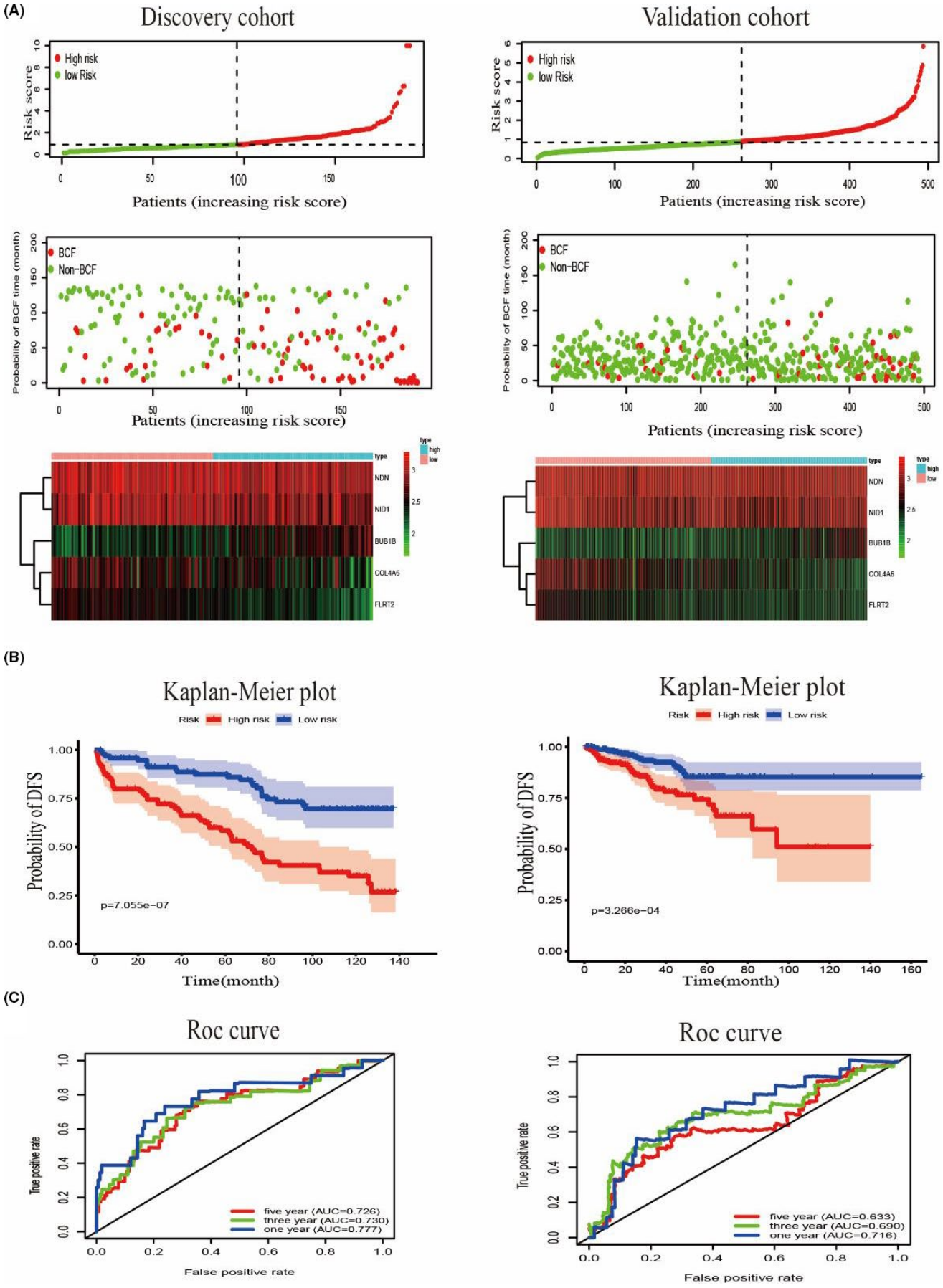
(A)
Discovery cohort
Validation cohort
High risk
low Risk
Risk score
Patients (increasing risk score)
BCF
Non-BCF
Probability of BCF time (month)
type
high
low
NDN
NID1
BUB1B
COL4A6
FLRT2
(B)
Kaplan-Meier plot
Risk
High risk
Low risk
Probability of DFS
p=7.055e-07
p=3.266e-04
Time(month)
(C)
Roc curve
True positive rate
False positive rate
five year (AUC=0.726)
three year (AUC=0.730)
one year (AUC=0.777)
five year (AUC=0.633)
three year (AUC=0.690)
one year (AUC=0.716)

FIGURE 2 Validation of the immune-related gene (IRG)-based signature in prostatic adenocarcinoma. A, Risk score distribution of IRG-based signature; patients' biochemical failure (BCF) status in different risk groups and heatmap of five IRG expression profiles. B, Kaplan-Meier curve plots to assess patient prognosis in different cohorts. C, Time-dependent receiver operating characteristic (ROC) curves to evaluate the accuracy of the IRG-based signature in different cohorts

3.4 | Associations between the IRG-based signature and clinical characteristics

To evaluate the predictive power of the IRG-based signature, univariate and multivariate Cox regression analyses were undertaken in both cohorts (Figure 3). Univariate Cox regression analysis revealed that the Gleason score, pT stage, and risk score were strongly correlated with BCF in discovery and validation cohorts ($P < .05$). In multivariate Cox regression analysis, the Gleason score and risk score retained their prognostic value in the discovery cohort ($P < .05$). However, in the validation cohort, risk score was not evidently associated with BCF ($P = .378$). The associations between signature-based risk score and patients' clinical characteristics are shown in Table 1. The Gleason score ($P < .01$), metastasis status ($P < .05$), and pT stage ($P < .01$) were significantly associated with risk score in both the discovery and validation cohorts. Patients with high Gleason score, advanced stage, and metastasis tended to have higher risk scores.

A stratified analysis was carried out to evaluate the prognostic value of the IRG-based signature in different subgroups in PRAD (Figure S3). Stratified analysis suggested that the IRG-based signature was a powerful tool to predict BCF in elderly patients (age >60 years), patients with a low (<8) or high Gleason score (≥8), early (≤pT2) or advanced stage (>pT2) patients, and patients with negative or positive surgical margins in the discovery cohort ($P < .05$). However, in the validation cohort, the IRG-based signature only showed prognostic efficacy in young patients (age ≤60 years), patients with a low Gleason score (<8), advanced stage (>pT2) patients, and patients with a negative surgical margin ($P < .05$). Collectively, the IRG-based signature demonstrated satisfactory prognostic values for BCF for specific clinical factors in PRAD.

The C-index was computed to assess the prognostic value of selected parameters (age, Gleason score, pT stage, surgical margin, and risk score). Compared with signature-based risk scores or certain clinicopathologic parameters, the prognostic model integrating risk score with clinical parameters had the best C-index in both cohorts (discovery cohort: 0.749, 95% CI, 0.688-0.809; validation cohort: 0.743, 95% CI, 0.674-0.812) (Table S2).

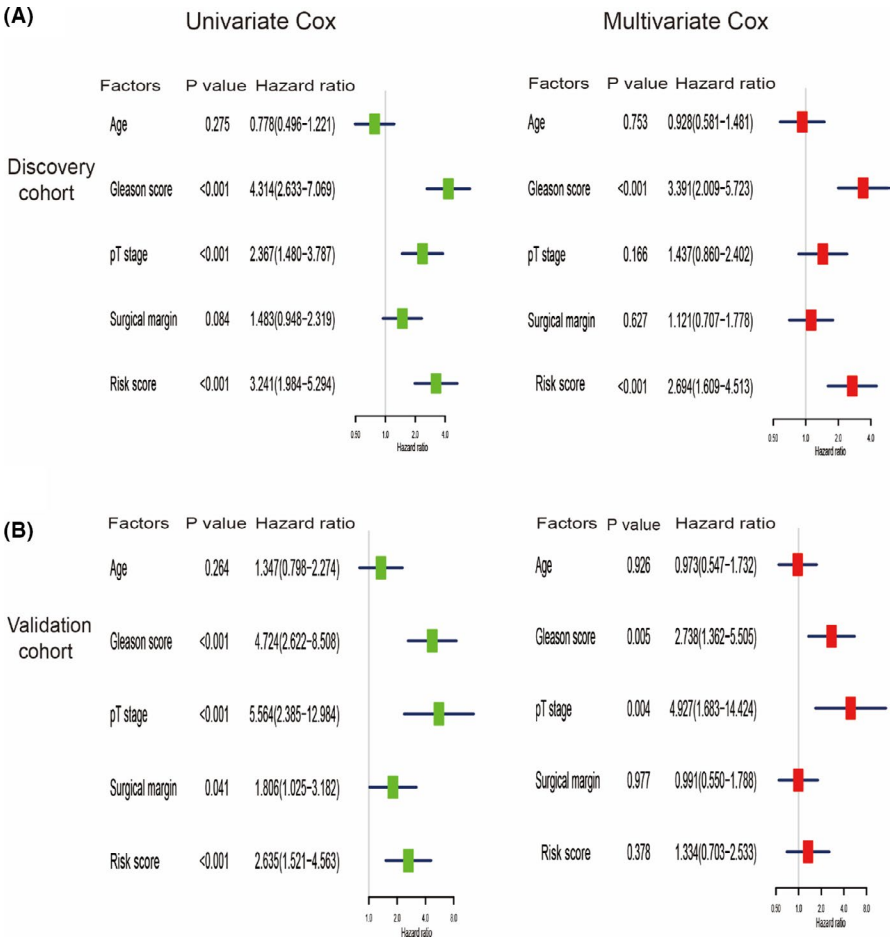

FIGURE 3 Univariate and multivariate Cox regression analyses in (A) discovery and (B) validation cohorts of patients with prostatic adenocarcinoma. In the univariate and multivariate Cox regression analyses, risk score, age, Gleason score, pT stage, and metastasis status are evaluated as continuous variables. $P < .05$ is considered statistically significant in all analyses

TABLE 1 Associations between signature-based risk score and clinical characteristics of patients with prostatic adenocarcinoma

Variable	Risk score (discovery cohort)			Risk score (validation cohort)		
	Low (%)	High (%)	*P* value	Low (%)	High (%)	*P* value
Age, years			.030			.145
<60	27 (39.7)	41 (60.3)		114 (57.0)	86 (43.0)	
≥60	69 (56.1)	54 (43.9)		148 (50.3)	146 (49.7)	
Gleason score			.008			<.000
<8	88 (54.3)	74 (45.7)		201 (69.3)	89 (30.7)	
≥8	8 (27.6)	21 (72.4)		61 (29.9)	143 (70.1)	
pT stage			.001			<.000
=T2	82 (56.9)	62 (43.1)		130 (69.9)	56 (30.1)	
>T2	14 (29.8)	33 (70.2)		127 (42.2)	174 (57.8)	
Metastasis			.017			<.000
No	91 (53.2)	80 (46.8)		240 (58.0)	174 (42.0)	
Yes	5 (25.0)	15 (75.0)		22 (27.5)	58 (72.5)	

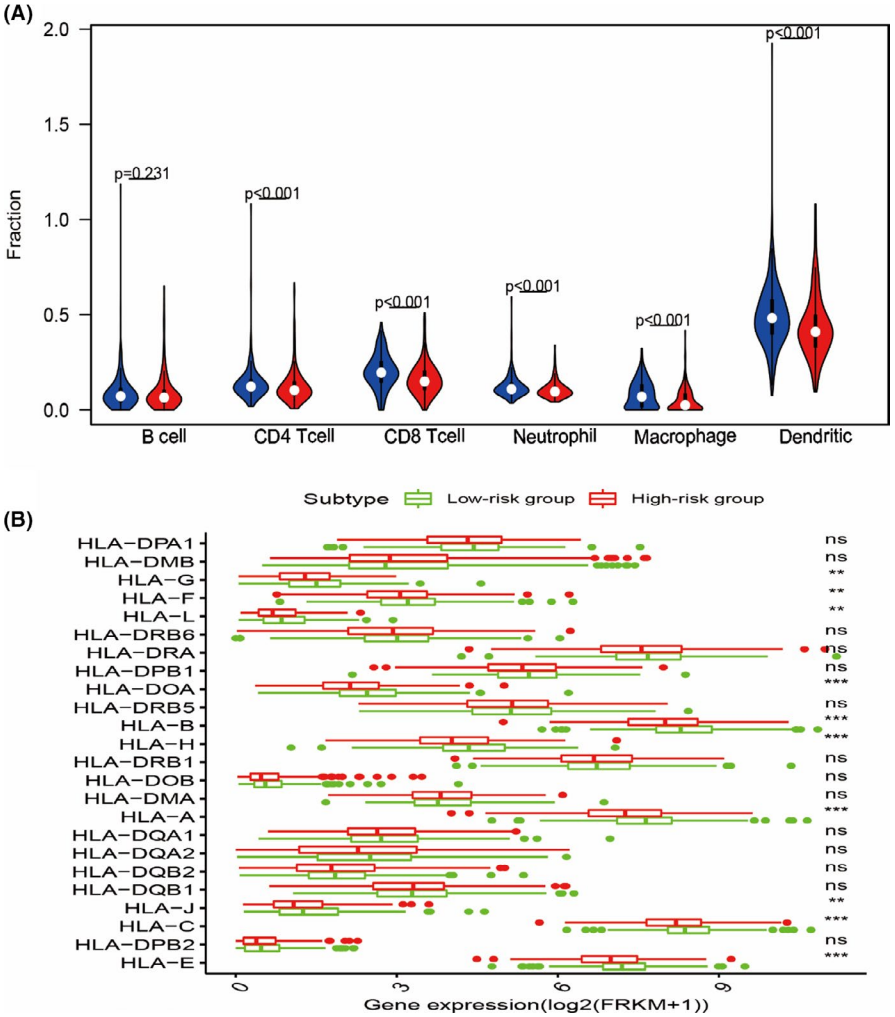

FIGURE 4 Infiltration of immune cells and expression of human leukocyte antigen (HLA)-related genes in low-risk and high-risk groups of prostatic adenocarcinoma (PRAD) samples. A, Violin plot shows the differential abundance of six infiltrative immune cells by TIMER database between high-risk (red) and low-risk (blue) groups of PRAD samples. B, Differential abundance of HLA-related genes between high- and low-risk groups in PRAD. Group differences were assessed by the Wilcoxon test. ns, not significant. $^{**}P < 0.01$; $^{***}P < 0.001$

To provide a quantitative method to predict the probability of BCF in PRAD after RP for clinicians, we constructed a nomogram integrating the signature-based risk score with classical prognostic parameters. Calibration plots revealed that the nomogram performed well for predicting 1-year, 3-year, and 5-year BCF (Figure S4).

3.5 | Correlation between signature-based risk score and TIIC

To explore whether the signature-based risk score could precisely assess the immune status of the tumor microenvironment, a TIMER

database analysis was carried out to evaluate the correlation between signature-based risk score and infiltration of immune cells. Results indicated that the infiltration rates of five types of immune cells between the high-risk and low-risk groups were significantly different, including neutrophils, $CD8^+$ T cells, macrophages, $CD4^+$ T cells, and dendritic cells ($P < .05$) (Figure 4A). The relationship between the signature-based risk score and expression of HLA-related genes was also assessed. The expression of main HLA I-related genes (including HLA-A, HLA-B, and HLA-C) between high-risk and low-risk groups was significantly different in PRAD. The risk score was negatively correlated with the expression of HLA I-related genes ($P < .05$, Figure 4B).

In order to further explore the specific immunophenotype of PRAD, the mRNA expression of PRAD patients in the TCGA database was used to infer the infiltration of 22 subtypes of immune cells using the CIBERSORT algorithm, and the differences between high-risk and low-risk groups were investigated. In the TCGA database, 123 PRAD cases were enrolled with CIBERSORT ($P < .05$). CD4 naive T cells were excluded after estimation, as they were almost absent in the prostate tissue. As shown in Figure 5A, macrophages M0 and M2, $CD8^+$ T cells, $CD4^+$ T cells, and naive B cells accounted for the majority of infiltrating immune cells in PRAD. The proportion of immune cell subtypes was substantially different between the high-risk and low-risk groups (Figure 5C). Among 123 PRAD patients, the infiltration levels of $CD4^+$ memory-activated T cells and regulatory T cells were higher in the high-risk group than in the low-risk group ($P < .05$). Compared with the low-risk group, the infiltration levels of plasma cells, monocytes, activated mast cells, and neutrophils were significantly decreased in the high-risk group ($P < .05$). The infiltration level of $CD8^+$ T cells tended to be higher in the high-risk group than in the low-risk group; however, it did not reach significance ($P = .075$).

4 | DISCUSSION

The prostate is an organ with multiple tumor-related antigens as potential therapeutic targets. As such, PRAD is regarded as an ideal tumor model for immunotherapy. In addition, PRAD is typically a torpid disease that provides abundant time for the appearance of an antitumor immune response.[22] However, HLA I may be downregulated in PRAD, rendering antigen presentation ineffective for immune escape by the growth in the number of regulatory T cells, generation of immunosuppressive cytokines, or induction of T-cell apoptosis by false expressions.[23] Previous clinical trials have reported that patients with advanced prostate cancer show overwhelming resistance to immune checkpoint inhibitors such as programmed cell death protein 1, cytotoxic T-lymphocyte antigen 4, and lymphocyte activation gene-3, which could be attributed to the rare expression of immune checkpoint markers on prostate tumor cells. Nevertheless, recent clinical tests have confirmed that immunotherapy plays a vital role in PRAD. Therefore, novel IRGs hold potential for risk stratification and prognosis. Immune-related genes have been verified as independent predictors in several malignant tumors,[24-27] indicating that the IRG-based signatures could expound local immune responses and immune status.

In this study, a BCF-related module was identified in a cohort of RP patients with long-term follow-up (discovery cohort) using WGCNA. In total, 152 IRGs were identified from this specific module. Five IRGs (*BUB1B*, *NDN*, *NID1*, *COL4A6*, and *FLRT2*) were verified as components of a risk signature to divide PRAD patients into high-risk and low-risk groups. The IRG-based signature showed good performance in the risk stratification of PRAD patients in the internal (discovery) and external (validation) cohorts. Moreover, a univariate Cox regression analysis revealed that Gleason score, pT stage, and risk score showed prognostic value regarding BCF in PRAD. Furthermore, in the multivariate Cox regression analysis, the Gleason score and risk score maintained their prognostic value in the discovery cohort. Nevertheless, the risk score did not show the desired result in the validation cohort. Compared with the aforementioned clinicopathologic predictors, the risk score still showed independent prognostic value based on the univariate and multivariate Cox regression analyses. Furthermore, we also observed that the IRG-based signature showed satisfactory accuracy regarding the 1-, 3-, and 5-year DFS of PRAD patients in both cohorts.

Our results suggest that the IRG-based signature could be used as an appropriate index in the stratification of PRAD. Patients in the high-risk group showed a higher incidence of BCF. The expression of the five IRGs was significantly different between low-risk and high-risk groups. Among these IRGs, *BUB1B*, *NDN*, and *NID1* were identified as risk-associated genes. Conversely, *COL4A6* and *FLRT2* were identified as protective genes. Association analyses between the signature-based risk score and clinicopathologic factors revealed that patients with a high-risk score had a stronger correlation with high Gleason score, advanced stage, high incidence of metastasis, and poor prognosis. This reflected the contribution of an immunosuppressive tumor microenvironment to tumor progression and recurrence in the high-risk group. Furthermore, stratified analysis revealed that the IRG-based signature had better stratification efficacy for low Gleason score groups (<8), advanced stage groups (>pT2), and negative surgical margin groups in both cohorts. Harrell's C-index revealed that the prognostic model integrating risk score with clinical parameters had the best C-index in both cohorts. This suggested that the signature-based risk score combined with clinical parameters could enhance the predictive accuracy for BCF in PRAD.

This study identified five IRGs as a novel risk signature for PRAD. These genes have been reported to participate in immune response regulation. *BUB1B* is a mitotic checkpoint that contributes to the fidelity of chromosome segregation during the proliferation of primary T cells.[28] Precise regulation of the cell cycle is crucial for maintaining T cell homeostasis and preventing lymphoproliferative diseases. High *BUB1B* expression has been reported in several malignancies, and is correlated with poor outcomes, including PRAD.[29] *NDN* is a well characterized member of the MAGE protein family. During the process of tumor transformation, genes of this family may be reactivated, expressed, and turned into antigenic targets for immune

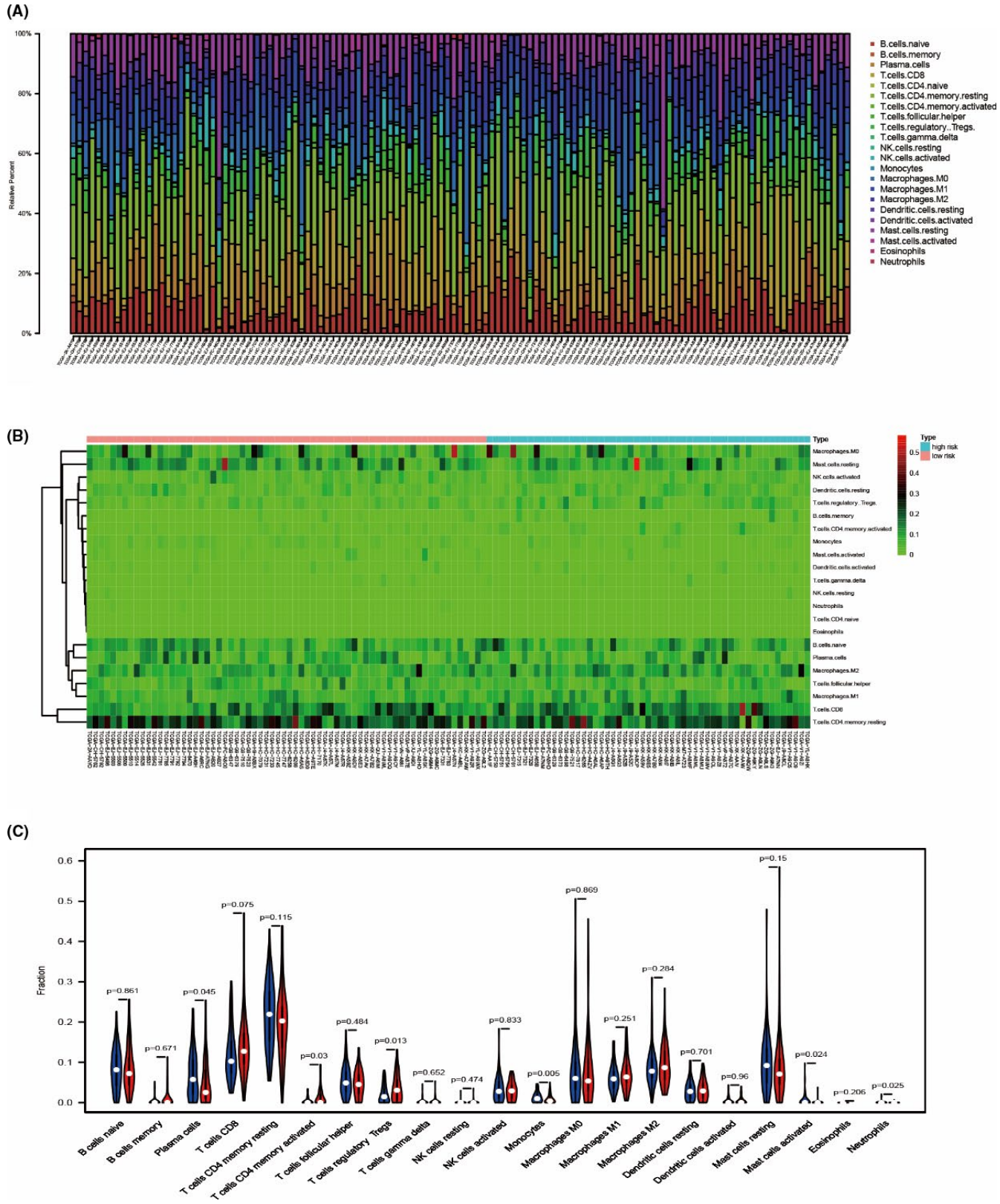

FIGURE 5 Correlation between the risk score and 22 subtypes of tumor-infiltrating immune cells (TIICs) in prostatic adenocarcinoma (PRAD) samples. A, Relative proportions of 22 subtypes of TIICs in high- and low-risk groups. B, Bar plot shows the percentages of 22 subtypes of TIICs calculated by CIBERSORT algorithm between high- and low-risk groups in PRAD patients from The Cancer Genome Atlas database (high risk, 55 samples; low risk, 68 samples). C, Violin plot shows the different proportions of 21 subtypes of TIICs between low- and high-risk groups in PRAD samples. Horizontal and vertical axes represent 21 subtypes of TIICs and relative percentages, respectively. Blue and red represent low- and high-risk groups, respectively, in PRAD samples. Data were assessed by the Wilcoxon test. NK, natural killer

system recognition and attack.[30] *NDN* may be involved in the process of immune destruction against several early tumor cells.[31] *NDN* hypermethylation is associated with better survival in head and neck squamous cell carcinoma patients.[32] *NID1* is a mesenchymal-related gene strongly linked to poor prognosis in ovarian cancer patients. *NID1* facilitates the invasion, migration, and chemoresistance of ovarian cancer by epithelial-mesenchymal transition through the activation of ERK/MAPK signaling.[33] *COL4A6* encodes one of the six subunits of type IV collagen. The downregulation of *COL4A6* may promote the progression and invasion of PRAD by activating the p-FAK/MMP-9 signaling pathway.[34] *FLRT2* is a typical member of the FLRT family, and encodes cell adhesion molecules. The diverse methylation of the *FLRT2* transcription start site could contribute to modulating the expression and the interaction of *FLRT2* and *FGFR2*, thereby regulating PRAD progression.[35] These five IRGs have shown prognostic value in various cancers. However, the mechanistic underpinnings of their involvement in immune modulation remain to be elucidated. Therefore, further exploration is necessary to expound the correlations between risk stratification and infiltration of immune cells in PRAD.

In this study, the infiltration of immune cells in PRAD patients was analyzed with the TIMER database. Patients in the low-risk group showed higher infiltration rates of immune cells when compared to patients in the high-risk group, including $CD8^+$ T cells, $CD4^+$ T cells, macrophages, dendritic cells, and neutrophils. The expression of HLA I-related genes (including HLA-A, HLA-B, and HLA-C) was also significantly increased in low-risk PRAD patients. The BCF of PRAD could be closely related to the infiltration level of immune cells, expression of immune components, and immune status. A higher immune status predicts a lower risk of BCF in PRAD. Previous studies have verified that immune cells differentiate into diverse subsets with inverse functions.[36,37] The infiltration of 22 subtypes of immune cells was also analyzed using the CIBERSORT algorithm with the TCGA database. In 123 PRAD patients, risk score was positively related to activated $CD4^+$ memory T cells and regulatory T cells, and was negatively related to plasma cells, monocytes, activated mast cells, and neutrophils. These results indicate that PRAD differs from other tumors and could have a specific immunophenotype and individual immune prognostic factors. Consequently, further research is necessary to determine possible molecular mechanisms linking the IRGs in the novel risk signature and tumor immune status of PRAD.

Limitations of this research include its retrospective nature and the need to explore the molecular mechanisms by which IRGs in the special module affect the occurrence, progression, and prognosis in PRAD. As a result of limitations associated with the experimental conditions, the specific molecular mechanisms were not explored further. Nevertheless, in addition to improving the present prognostic efficacy, we expect that this novel IRG-based signature will facilitate clinical research with the purpose of developing novel therapeutic strategies for PRAD.

In summary, we have identified a BCF-related module with a long-term follow-up database (discovery cohort) using WGCNA. An IRG-based signature was identified from the specific module, which could function as an index for assessing tumor immune status and risk stratification in PRAD.

ACKNOWLEDGMENTS

This work was supported by grants from National Natural Science Foundation of China (81874099, 82072813), Natural Science Foundation of Guangdong Province (2020A1515010473), Science and Technology Project of Bureau of Health in Guangzhou Municipality (20171A011239), Guangzhou Municipal Science and Technology Project (201803040001), and The Science Foundation of Guangzhou First People's Hospital (Q2019020, Q2019007).

DISCLOSURE

The authors declare that they have no competing interests.

ETHICAL APPROVAL

This study was approved by the Hospital Ethics Committee of Massachusetts General Hospital. All of the patients gave informed consents before undergoing surgery. Written informed consent was obtained from all patients before the study. The principles of the 1964 Helsinki Declaration were followed in this study. In order to protect patients' privacy, we used Case No. to replace patients' names.

ORCID

Hai-Bo Zhao https://orcid.org/0000-0001-5168-5757
Wei-De Zhong https://orcid.org/0000-0002-0430-2845

REFERENCES

1. World Health Organization (2020) Latest global cancer data: cancer burden rises to 19.3 million new cases and 10.0 million cancer deaths in 2020 questions and answers (Q&A). https://www.iarc.fr/faq/latest-global-cancer-data-2020-qa/. Accessed December 15, 2020.
2. Diaz M, Peabody JO, Kapoor V, et al. Oncologic outcomes at 10 years following robotic radical prostatectomy. *Eur Urol*. 2015;67:1168-1176.
3. Antwi SO, Steck SE, Zhang H, et al. Plasma carotenoids and tocopherols in relation to prostate-specific antigen (PSA) levels among men with biochemical recurrence of prostate cancer. *Cancer Epidemiol*. 2015;39:752-762.
4. 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;71:630-642.
5. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol*. 2010;17(6):1471-1474.
6. Liu Y, Jing R, Xu J, et al. Comparative analysis of oncogenes identified by microarray and RNA-sequencing as biomarkers for clinical prognosis. *Biomark Med*. 2015;9:1067-1078.
7. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med*. 2013;19:1423-1437.
8. Pietras K, Ostman A. Hallmarks of cancer: interactions with the tumor stroma. *Exp Cell Res*. 2010;316:1324-1331.
9. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144:646-674.
10. Venturini NJ, Drake CG. Immunotherapy for prostate cancer. *Cold Spring Harb Perspect Med*. 2019;9:a030627.

11. Shenderov E, Antonarakis ES. Reimagining vaccines for prostate cancer: back to the future. *Clin Cancer Res*. 2020;26:5056-5058.
12. Wu CL, Schroeder BE, Ma XJ, et al. Development and validation of a 32-gene prognostic index for prostate cancer progression. *Proc Natl Acad Sci USA*. 2013;110:6121-6126.
13. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
14. Bhattacharya S, Andorf S, Gomes L, et al. ImmPort: disseminating data to the public for the future of immunology. *Immunol Res*. 2014;58:234-239.
15. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361-387.
16. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol*. 2015;16:e173-e180.
17. Li T, Fan J, Wang B, et al. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res*. 2017;77:e108-e110.
18. Li T, Fu J, Zeng Z, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res*. 2020;48:W509-W514.
19. Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*. 2019;37:773-782.
20. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12:453-457.
21. Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017;45:D362-D368.
22. Drake CG. Prostate cancer as a model for tumour immunotherapy. *Nat Rev Immunol*. 2010;10:580-593.
23. Drake CG, Jaffee E, Pardoll DM. Mechanisms of immune evasion by tumors. *Adv Immunol*. 2006;90:51-81.
24. Ascierto ML, Kmieciak M, Idowu MO, et al. A signature of immune function genes associated with recurrence-free survival in breast cancer patients. *Breast Cancer Res Treat*. 2012;131:871-880.
25. Kim K, Jeon S, Kim TM, Jung CK. Immune gene signature delineates a subclass of papillary thyroid cancer with unfavorable clinical outcomes. *Cancers (Basel)*. 2018;10(12):494.
26. Shen S, Wang G, Zhang R, et al. Development and validation of an immune gene-set based prognostic signature in ovarian cancer. *EBioMedicine*. 2019;40:318-326.
27. Yang W, Lai Z, Li Y, et al. Immune signature profiling identified prognostic factors for gastric cancer. *Chin J Cancer Res*. 2019;31:463-470.
28. Baek KH, Shin HJ, Yoo JK, et al. p53 Deficiency and defective mitotic checkpoint in proliferating T lymphocytes increase chromosomal instability through aberrant exit from mitotic arrest. *J Leukoc Biol*. 2003;73:850-861.
29. Fu X, Chen G, Cai ZD, et al. Overexpression of BUB1B contributes to progression of prostate cancer and predicts poor outcome in patients with prostate cancer. *Onco Targets Ther*. 2016;9:2211-2220.
30. Zhao Z, Lu F, Zhu F, Yang H, Chai Y, Chen S. Cloning and biological comparison of Restin, novel member of Mage superfamily. *Sci China C Life Sci*. 2002;45:412-420.
31. Xiao J, Chen HS. Biological functions of melanoma-associated antigens. *World J Gastroenterol*. 2004;10:1849-1853.
32. Virani S, Bellile E, Bradford CR, et al. NDN and CD1A are novel prognostic methylation markers in patients with head and neck squamous carcinomas. *BMC Cancer*. 2015;15:825.
33. Zhou Y, Zhu Y, Fan X, et al. NID1, a new regulator of EMT required for metastasis and chemoresistance of ovarian cancer cells. *Oncotarget*. 2017;8:33110-33121.
34. Ma JB, Bai JY, Zhang HB, Gu L, He D, Guo P. Downregulation of collagen COL4A6 is associated with prostate cancer progression and metastasis. *Genet Test Mol Biomarkers*. 2020;24:399-408.
35. Wu Y, Davison J, Qu X, et al. Methylation profiling identified novel differentially methylated markers including OPCML and FLRT2 in prostate cancer. *Epigenetics*. 2016;11:247-258.
36. Long J, Wang A, Bai Y, et al. Development and validation of a TP53-associated immune prognostic model for hepatocellular carcinoma. *EBioMedicine*. 2019;42:363-374.
37. Pinto MP, Balmaceda C, Bravo ML, et al. Patient inflammatory status and CD4+/CD8+ intraepithelial tumor lymphocyte infiltration are predictors of outcomes in high-grade serous ovarian cancer. *Gynecol Oncol*. 2018;151:10-17.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Zhao H-B, Zeng Y-R, Han Z-D, et al. Novel immune-related signature for risk stratification and prognosis in prostatic adenocarcinoma. *Cancer Sci*. 2021;112:4365–4376. https://doi.org/10.1111/cas.15062