ORIGINAL RESEARCH

Cancer Medicine _Open Access_ WILEY

A four-gene signature associated with clinical features can better predict prognosis in prostate cancer

Penghui Yuan[1] | Le Ling[1] | Qing Fan[3] | Xintao Gao[1] | Taotao Sun[1] | Jianping Miao[2] | Xianglin Yuan[3] | Jihong Liu[1] | Bo Liu[3] (iD)

[1]Department of Urology Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]Department of Geriatrics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[3]Department of Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Correspondence
Bo Liu, Department of Oncology, Tongji Hospital, Huazhong University of Science and Technology, 1095, Jiefang Road, Wuhan, PR China.
Email: boliu888@hotmail.com

Funding information
National Natural Science Foundation of China, Grant/Award Number: 81902619

Abstract

Prostate cancer (PCa) is one of the most deadly urinary tumors in men globally, and the 5-year over survival is poor due to metastasis of tumor. It is significant to explore potential biomarkers for early diagnosis and personalized therapy of PCa. In the present study, we performed an integrated analysis based on multiple microarrays in the Gene Expression Omnibus (GEO) dataset and obtained differentially expressed genes (DEGs) between 510 PCa and 259 benign issues. The weighted correlation network analysis indicated that prognostic profile was the most relevant to DEGs. Then, univariate and multivariate COX regression analyses were conducted and four prognostic genes were obtained to establish a four-gene prognostic model. And the predictive effect and expression profiles of the four genes were well validated in another GEO dataset, The Cancer Genome Atlas and the Human Protein Atlas datasets. Furthermore, combination of four-gene model and clinical features was analyzed systematically to guide the prognosis of patients with PCa to a largest extent. In summary, our findings indicate that four genes had important prognostic significance in PCa and combination of four-gene model and clinical features could achieve a better prediction to guide the prognosis of patients with PCa.

KEYWORDS

clinical features, four-gene signature, prognosis, prostate cancer

1 | INTRODUCTION

Prostate cancer (PCa) is the second most commonly diagnosed and the fifth leading death of cancer in men worldwide.[1] It is observed that the incidence of PCa has been growing globally, particularly in Asia, Northern and Western Europe.[2] The metastasis of PCa is the dominant factor of PCa-related death, resulting in the 5-year mortality over 70%.[3] As a significant sign of prognosis, biochemical recurrence (BCR), defined as

the return of prostate-specific antigen (PSA), occurs in 27%-53% of patients after radical prostatectomy and radiotherapy,[4] and tends to happen months or years ahead of other clinical symptoms of PCa recurrence [5] and increases the risk of developing distant metastases, PCa-specific, and overall mortality.[6] Although extensive research has been conducted on the mechanisms of carcinogenesis, the etiology of PCa still remains unclear. Therefore, it is significant to explore biomarkers for early diagnosis, prognosis, and personalized therapy of PCa.

Various molecular aberrations and genetic mutations exist in the pathogenesis of PCa.[7,8] In the meantime, cumulative studies have been conducted to identify novel molecular biomarkers and refine diagnosis and prognosis of PCa recently,[5] nevertheless, not all molecular alternations influence the tumor outcome, and a single gene aberration does not necessarily have a good predictive value due to complex carcinogenesis and individual difference.[9] It will be more valuable to explore multiple molecules combined with clinical features for better prediction of diagnosis and prognosis in PCa. Nowadays, analysis of microarray and high-throughput sequencing technologies has advanced diagnosis and prognosis of various diseases.[10] It provides an efficient tool to decipher critical molecular alternations especially in cancer. A comprehensive understanding of molecular patterns of PCa as well as clinical traits would contribute to identifying the prognostic risk of patients and achieving personalized therapy by a systematic model.

In this study, we performed an integrated analysis in multiple microarrays and identified hub genes affecting the prognosis. Then, a four-gene prognostic model was constructed and patients of PCa were stratified based on risk score. The predictive value and gene profiles were validated successfully in other independent PCa datasets. Finally, combination of four-gene model and clinical features was analyzed systematically to guide the prognosis of patients with PCa to a largest extent. In summary, the study is aimed to add novel knowledge of PCa development and prognosis by analyzing the genetic changes and clinical traits comprehensively.

2 | MATERIALS AND METHODS

2.1 | Data acquisition

Transcriptome data of PCa was obtained from the Gene Expression Omnibus (GEO) datasets (https://www.ncbi.nlm.nih.gov/geo) (GSE21034,[11] GSE55945,[12] GSE46602,[13] GSE62872,[14] and GSE29079[15]) on April 3, 2020. Each eligible dataset incorporated at least twenty samples of tumor and benign issue. After screening, 510 tumor samples and 259 benign samples were acquired for further analysis finally. This study were approved by our Institutional Research Ethics Committee.

2.2 | Integrated analysis of microarray datasets

Raw transcriptome data in each dataset was normalized by "limma" package [16] in R software (Version 3.6.2). After averaging the expression values of the genes corresponding to the multi-microarray probes and base-2 logarithm (log2) transformation, log2 fold-change (log_2FC) values of differentially expressed genes (DEGs) expressions were calculated between tumor and normal issues by "limma" package. Later, the integrated analysis was conducted across the five microarray datasets by robust rank aggregation algorithm with R package "RobustRankAggregation" based on a prioritized gene list.[17] DEGs with $|log_2FC| > 1$ and adjust $P < .01$ were regarded as statistical significance for further analysis.

2.3 | Functional enrichment analysis

To identify the biological functions related to DEGs, the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of DEGs were conducted by R package "clusterProfiler" [18] with the thresholds of adjusted $P < .05$ and $q < 0.05$. The GO and KEGG clusters were visualized by the R package "GOplot".[19]

2.4 | Evaluation of gene modules and correlation with clinical factors

To establish the gene interaction modules and assess the relationship between DEGs and clinical factors as a whole, the weighted correlation network analysis (WGCNA) was conducted for DEGs by R package "WGCNA".[20] Clinical data were extracted from GSE70769 dataset,[21] which including clinical stage, T stage, Gleason score, prognostic conditions as well as gene expression profiles. The value of soft threshold (power) was set to obtain the optimal scale free topology fit model index (scale free R^2) and mean connectivity in the meantime. Based on a topological overlap measure to determine the degree of dissimilarity among genes, cluster dendrogram of genes was obtained. After clustering modules and genes, clinical factors in GSE70769 were involved in correlation analysis with the module eigengenes. $P < .05$ was regarded as statistical significance.

2.5 | Construction and assessment of prognostic model based on DEGs

Differentially expressed genes were applied for prognostic analysis with clinical information in the GSE70769 dataset. First, univariate Cox regression method was employed to obtain candidate genes related to prognosis with $P < .05$ between patient BCR free survival and gene expression levels. Next, the LASSO Cox regression analysis by R package "glmnet"[22] was applied to further screen the candidate genes with more prognostic value based on penalty parameter tuning performed via 10-fold cross-validation.[23] Finally, the selected genes were involved in the

multivariate Cox regression model by a stepwise method. After these, a risk score tool was acquired based on the sum of gene expression level (Exp_i) and regression coefficient (β_i) based on the multivariate Cox regression model (risk score = $\Sigma Exp_i \times \beta_i$). The patients with clinical information were classified into different risk groups according the medium value of risk score. Then, Kaplan-Meier (KM) survival curves by R package "survival," receiver operational characteristic (ROC) curves by R package "timeROC"[24] and C-index were processed to evaluate the prognostic value of the multigene model. Finally, the model was validated in another GEO dataset (GSE116918) including 248 patients and The Cancer Genome Atlas (TCGA) dataset including 499 patients (https://portal.gdc.cancer.gov/).[25] And the prognostic value of the multigene model was also evaluated compared with T stage in GSE70769.

2.6 | Verification of the expression profiles of prognostic genes

To validate the value of prognostic genes above, the expression profiles of these genes were analyzed in TCGA prostate adenocarcinoma cohort[26] with the Mann-Whitney *U* test. $P < .05$ was regarded as statistical significance. Furthermore, to determine the clinical relevance of these genes, immunohistochemical data were downloaded from the Human Protein Atlas (HPA) (https://www.proteinatlas.org/)[27] to compare the levels of protein encoded by these genes. The results of immunohistochemistry in tumor and normal issues were shown by the same antibody.

2.7 | Combination of prognostic genes and clinical data for prognostic evaluation

To assess the association between the prognostic genes and other clinical features (including risk score based on prognostic genes, Gleason score, preoperative level of PSA, clinical stage, extra capsular extension, and positive surgical margins) and construct a more systematic prognostic model, univariate and multivariate Cox regression analyses were processed with BCR free survival as the dependent variable and prognostic genes as well as clinical features as the independent variables. Hazard value and *P* value (<.05) were obtained for assessment of prognosis. Based on the results, a new model combining prognostic genes and significant clinical features was constructed. Similarly, the KM survival curves, ROC curves, and C-index were shown. Furthermore, a nomogram was established to show prognostic profiles and calibration curves were drawn to verify the accuracy of the nomogram by R package "rms".[28]

3 | RESULTS

3.1 | Identification of DEGs based on integrated analysis

The flow diagram in this study is illustrated in Figure 1. The information of five GEO datasets involved in this study is shown in Table S1. There were 510 PCa samples and 259 benign samples for further analysis. After normalization and integrated analysis in each dataset (Figure S1), a total of 270 DEGs involving 148 downregulated and 122 upregulated genes were identified by robust rank aggregation algorithm. The top 20 DEGs in both of them are shown in Figure 2. AMACR, ACSM1, ERG, DNAH5, and CRISP3 were top five genes in upregulated sets and NEFH, SLC14A1, CD177, KRT5, and MME were top five genes in downregulated sets.

3.2 | Identification of biological functions related to DEGs

GO and KEGG enrichment analyses were conducted to reveal the biological roles of DEGs. The GO profiles of DEGs are shown in Figure 3A. It revealed that these DEGs had a close relationship with metabolism and peptidase activities (Figure 3B and Table S2). The top five GO terms were enzyme inhibitor activity, peptidase regulator and inhibitor activities, modified amino acid binding, and extracellular matrix binding. Similarly, KEGG profiles are shown in Figure 3C. These DGEs mainly participated in Wnt signaling pathway, glutathione metabolism, metabolism of drug, and xenobiotics by cytochrome P450. Also, PCa pathway was involved significantly (Figure 3D and Table S3).

3.3 | Weighted correlation network analysis

To evaluate the clinical profiles of DEGs, WGCNA analysis was conducted to determine gene modules and correlation with clinical traits. Based on the optimal soft threshold, seven gene modules were identified and cluster dendrogram was obtained (Figure S2). The relationships between modules and clinical features including BCR status, BCR free survival, Gleason score, PSA level, clinical stage, and pathological T stage are shown in Figure 4. Of all the modules, three modules correlated strongly with BCR status ($P < .05$) and BCR free survival ($P < .01$), respectively. One module had a negative correlation with Gleason score ($P < .05$), while there was no significant correlation between gene modules and the other clinical factors. Based on the results, clinical prognosis was considered as the

FIGURE 1 The flow diagram of the study. GEO, Gene Expression Omnibus; GO, Gene Ontology; HPA, Human Protein Atlas; KEGG, Kyoto Encyclopedia of Genes and Genomes; PCa, prostate cancer; TCGA, The Cancer Genome Atlas; WCGNA, weighted correlation network analysis
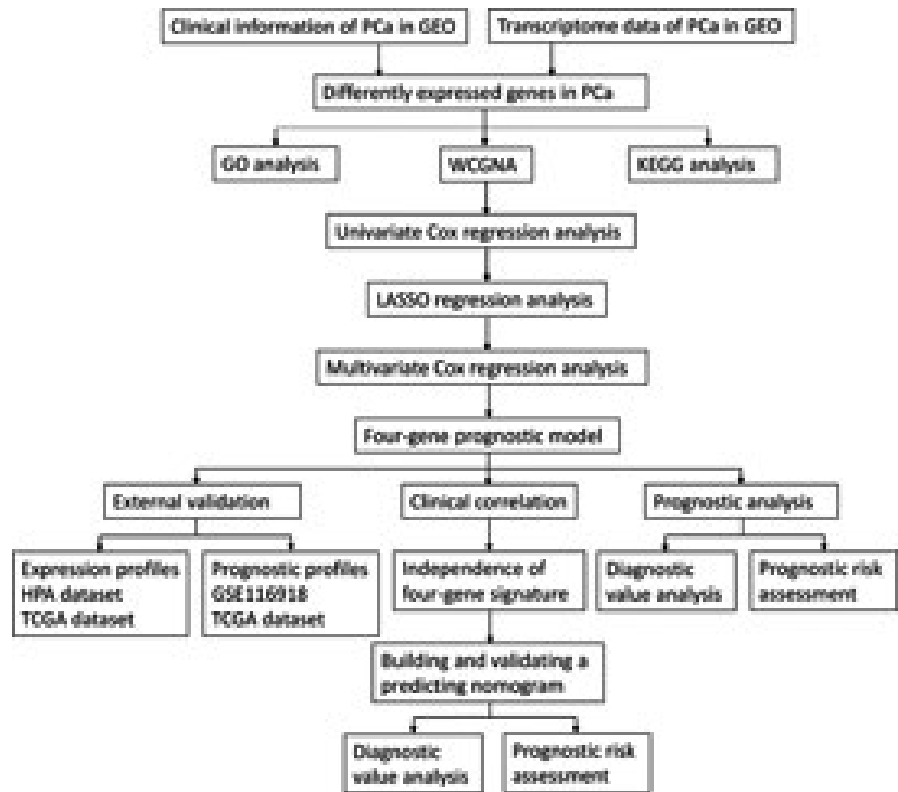
main association with DEGs. Prognosis analysis deserved to be conducted subsequently.
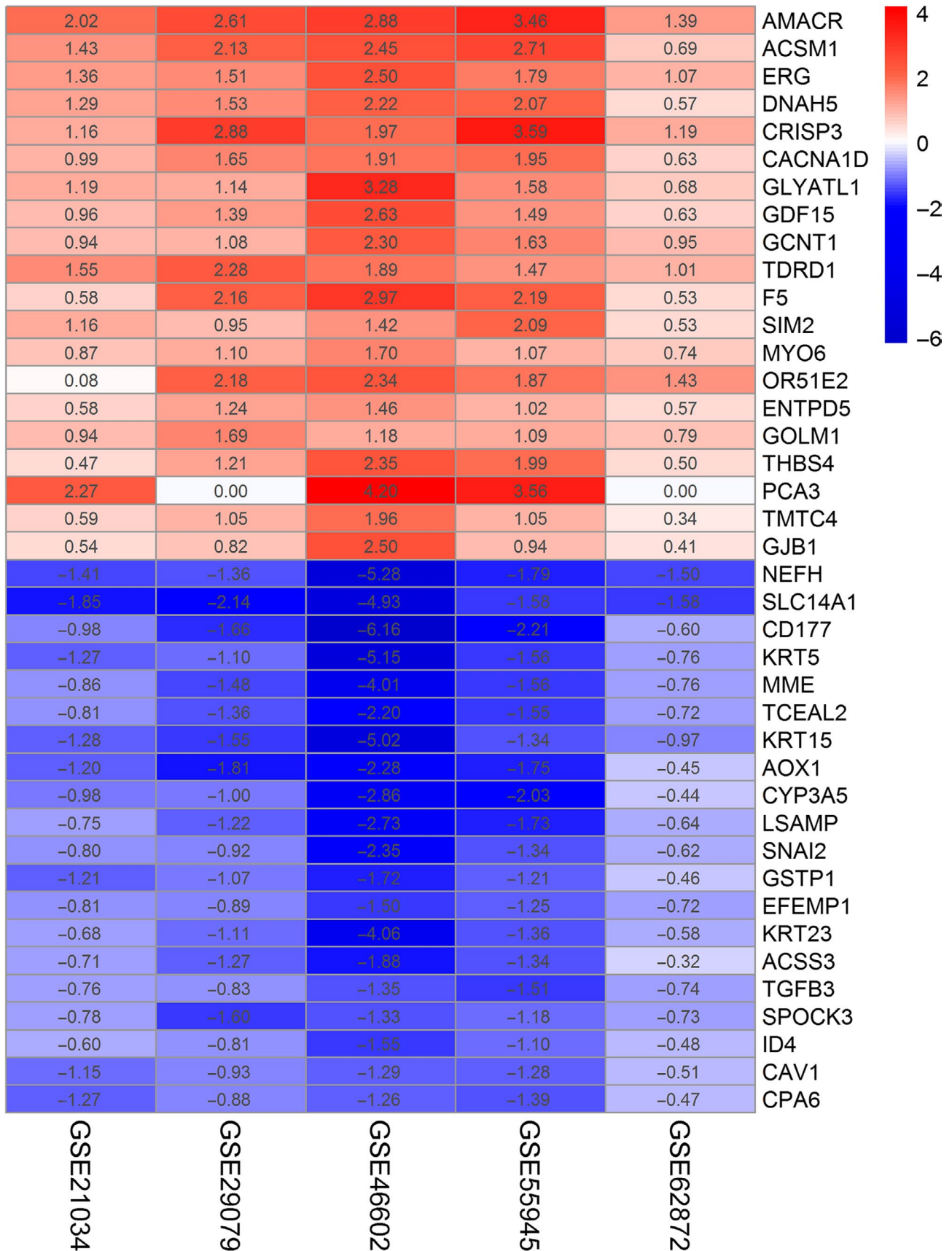
3.4 | Prognostic analysis of DEGs

To determine the prognostic value of DEGs, a univariate Cox regression was conducted to investigate the relationship between DEGs and BCR free survival in GSE 70 769 (Table S4). Then, 73 candidate DEGs were involved in LASSO Cox regression analysis and four effective DEGs were selected (Figure S3). Finally, ANO4, EZH2, PARM1, and SRD5A2 were regarded as prognostic genes and four-gene prognostic model was constructed in stepwise multivariate Cox regression analysis. Risk scores for patients were calculated based on the integrated combination of gene expression level and corresponding regression coefficient. That is, risk score $= (-1.25558 \times$ expression of ANO4) $+ (1.337196 \times$ expression of EZH2) $+ (-0.36793 \times$ expression of PARM1) $+ (-0.73096 \times$ expression of SRD5A2). Based on medium value of risk scores, patients in GSE70769 were divided into low- and high-risk groups. Prognostic curves between BCR free survival and risk level were drawn (Figure 5A). It showed that low-risk group had a better BCR free survival than that in high-risk group ($P < .0001$). In the meantime, the AUCs of gene model corresponding to 1-, 3-, and 5-year BCR free survival calculated based on ROC curves were 0.83, 0.799, and 0.81 (Figure 5B). It suggested good specificity and sensitivity of the four-gene prognostic model.

## 3.5 | External validation of four-gene model in another GEO and TCGA datasets

To assess the external predictive effect of four-gene model, clinical data in GSE116918 including 248 patients and TCGA dataset including 499 patients was analyzed. After calculating risk score for patients in GSE116918 and TCGA, patients were divided into low- and high-risk groups using the four-gene prognostic model. Consistent with prognostic information in GSE70769, BCR free survival was significantly better in the low-risk group than that in high-risk group in GSE116918 ($P = .00499$) and TCGA ($P = .00013$), respectively. These findings indicated the predictive effect of four-gene model was well validated (Figure 5C-F). In addition, we compared the predictive effect of four-gene model with T stage in GSE70769. The patients were classified into different groups according the specific T stage. Patients with advanced T stage (T3 or higher) showed worse prognosis ($P < .0001$) compared with lower T stage. The AUCs of 0.816, 0.747, and 0.722 at 1, 3, and 5 years were similar with four-gene model (Figure S4). Also, the C-index for four-gene model was 0.701(0.638, 0.764), which was comparable with T stage (0.708(0.634, 0.782)).

3.6 | Validation of the expression of prognostic genes

Besides external validation of prognostic prediction in GEO dataset, the expression patterns of prognostic genes were also

2.02	2.61	2.88	3.46	1.39	AMACR
1.43	2.13	2.45	2.71	0.69	ACSM1
1.36	1.51	2.50	1.79	1.07	ERG
1.29	1.53	2.22	2.07	0.57	DNAH5
1.16	2.88	1.97	3.59	1.19	CRISP3
0.99	1.65	1.91	1.95	0.63	CACNA1D
1.19	1.14	3.28	1.58	0.68	GLYATL1
0.96	1.39	2.63	1.49	0.63	GDF15
0.94	1.08	2.30	1.63	0.95	GCNT1
1.55	2.28	1.89	1.47	1.01	TDRD1
0.58	2.16	2.97	2.19	0.53	F5
1.16	0.95	1.42	2.09	0.53	SIM2
0.87	1.10	1.70	1.07	0.74	MYO6
0.08	2.18	2.34	1.87	1.43	OR51E2
0.58	1.24	1.46	1.02	0.57	ENTPD5
0.94	1.69	1.18	1.09	0.79	GOLM1
0.47	1.21	2.35	1.99	0.50	THBS4
2.27	0.00	4.20	3.56	0.00	PCA3
0.59	1.05	1.96	1.05	0.34	TMTC4
0.54	0.82	2.50	0.94	0.41	GJB1
−1.41	−1.36	−5.28	−1.79	−1.50	NEFH
−1.85	−2.14	−4.93	−1.58	−1.58	SLC14A1
−0.98	−1.66	−6.16	−2.21	−0.60	CD177
−1.27	−1.10	−5.15	−1.56	−0.76	KRT5
−0.86	−1.48	−4.01	−1.56	−0.76	MME
−0.81	−1.36	−2.20	−1.55	−0.72	TCEAL2
−1.28	−1.55	−5.02	−1.34	−0.97	KRT15
−1.20	−1.81	−2.28	−1.75	−0.45	AOX1
−0.98	−1.00	−2.86	−2.03	−0.44	CYP3A5
−0.75	−1.22	−2.73	−1.73	−0.64	LSAMP
−0.80	−0.92	−2.35	−1.34	−0.62	SNAI2
−1.21	−1.07	−1.72	−1.21	−0.46	GSTP1
−0.81	−0.89	−1.50	−1.25	−0.72	EFEMP1
−0.68	−1.11	−4.06	−1.36	−0.58	KRT23
−0.71	−1.27	−1.88	−1.34	−0.32	ACSS3
−0.76	−0.83	−1.35	−1.51	−0.74	TGFB3
−0.78	−1.60	−1.33	−1.18	−0.73	SPOCK3
−0.60	−0.81	−1.65	−1.10	−0.48	ID4
−1.15	−0.93	−1.29	−1.28	−0.51	CAV1
−1.27	−0.88	−1.26	−1.39	−0.47	CPA6
GSE21034	GSE29079	GSE46602	GSE55945	GSE62872	

FIGURE 2 A heat map of the top 20 significantly upregulated and downregulated DEGs based on integrated analysis. The expression profile of genes in each dataset is shown by $\log_2$FC value in each column. The color layout from blue to red represents the expression level of DEGs from downregulation to upregulation. DEGs, differentially expressed genes; $\log_2$FC value, base-2 logarithm transformation fold-change value
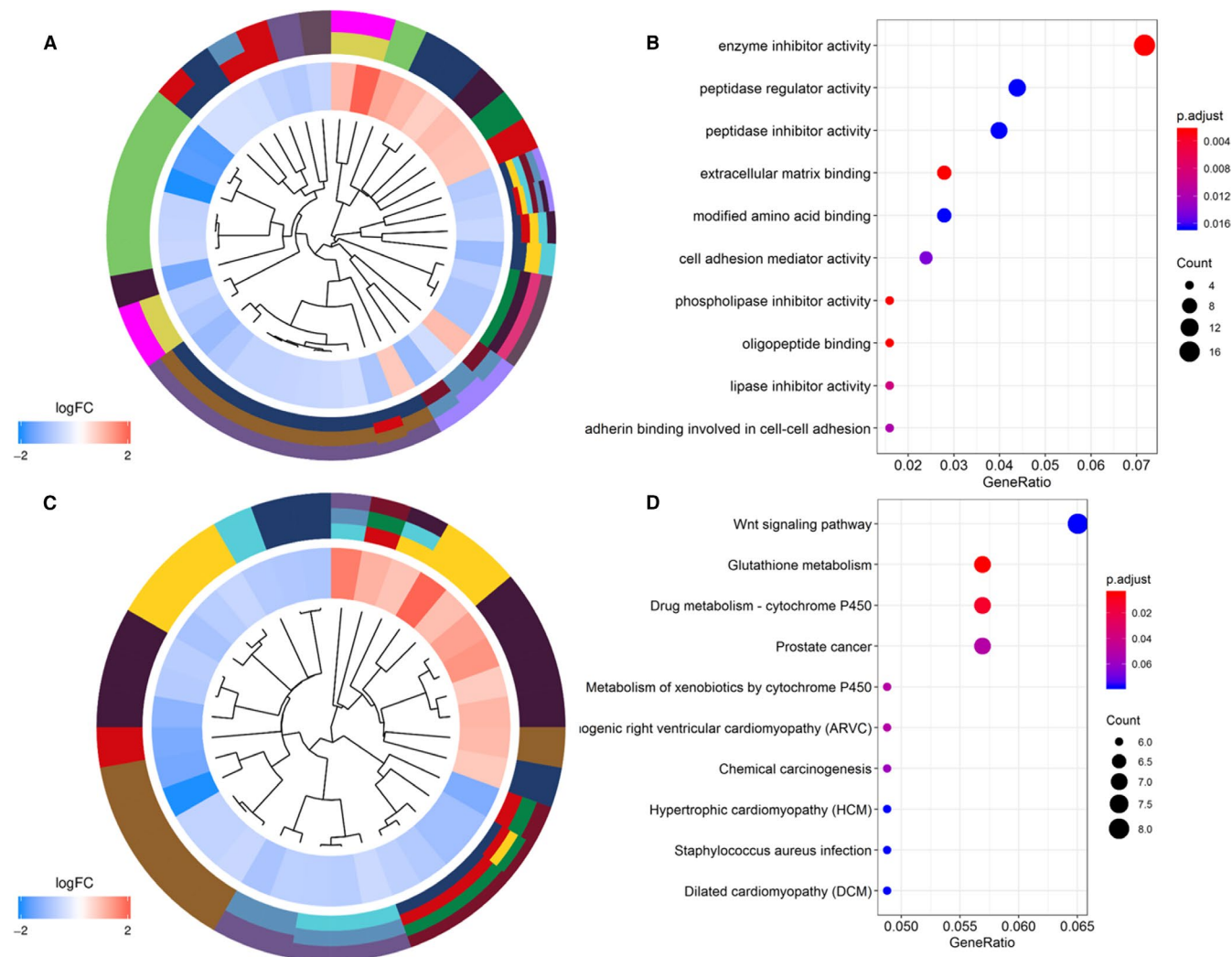
FIGURE 3 Functional enrichment analysis of DEGs. A, GO cluster. B, GO analysis of representative DEGs. The *y*-axis shows enriched GO terms. C, KEGG cluster. D, KEGG analysis of representative DEGs. The *y*-axis shows enriched KEGG pathways. For (A) and (C), the innermost part shows the hierarchical clustering of the DEGs. The middle part represents the expression profiles of DEGs, in which the color layout from blue to red indicates the expression level of DEGs from downregulation to upregulation. And the outermost part represents the GO terms (A) and KEGG pathways (C) associated with DEGs. DEGs, differentially expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes

assessed in TCGA and HPA datasets. As shown in TCGA (Figure 6A), the expression levels of ANO4 and PARM1 were significantly lower and EZH2 was significantly higher in PCa issue compared with normal issue ($P < .05$). What is more, in the HPA database, EZH2 was strongly positive in PCa issue and negative in normal tissue, while PARM1 was moderately positive in normal issue and negative in PCa tissue (Figure 6B). ANO4 and SRD5A2 were not found in the HPA database. All of the results were consistent with data in GEO datasets.

3.7 | Assessment of independent predictive value of gene model

Four-gene model and other clinical features including risk score, Gleason score, preoperative level of PSA, clinical stage,

extra capsular extension, and positive surgical margins from GSE 70769 were involved in univariate Cox regression analysis first. It showed that risk score, Gleason score, extra capsular extension, and positive surgical margins had a relationship with prognosis ($P < .05$) (Figure S5). Then, these significant factors were incorporated into the multivariate Cox regression analysis. The result showed that risk score, Gleason score, extra capsular extension, and positive surgical margins had independent prognostic value associated with BCR free survival (Figure S5).

3.8 | Construction and validation of a systematic prognostic model

To construct a most effective prognostic model in PCa, risk score based on four genes and clinical features in

Module–trait relationships

multivariate Cox regression model were combined. The KM curves showed consistent results with four-gene model (Figure 7A), but the AUCs of 0.879, 0.833, and 0.825 at 1, 3, and 5 years were higher (Figure 7B). The C-index for the combined model was 0.732(0.668, 0.797). These results indicated better predictive value of combined model compared with T stage. Furthermore, a nomogram including risk score, Gleason score, extra capsular extension, and positive surgical margins was developed and the probability of recurrence free survival at 1, 3, and 5 years could be calculated based on these parameters (Figure 8A). Finally, calibration plots were drawn to reflect the accuracy of nomogram. Doublication degree between the solid line and dotted line showed the nomogram had a good predictive value (Figure 8B). In addition, the combined model based on four genes and clinical features was also successfully validated in another PCa cohort

GSE70768 with 112 patients (Figure S6). The AUCs at 1, 3, and 5 years were 0.878, 0.768, and 0.778, respectively. The C-index for the combined model was 0.688 (0.594, 0.782).

4 | DISCUSSION

Prostate cancer is one of the most deadly urinary tumors in men globally.[1] Study reported that there was 29 430 PCa-related death cases in the United States in 2018.[29] Therefore, it is of significant importance to expound biomarkers of PCa to achieve better diagnostic and prognostic evaluation. Overall survival and recurrence free survival are common indictors to assess the prognosis in cancers. In PCa, the emergence of BCR after radical prostatectomy and radiotherapy is confirmed as a decisive risk factor for PCa-specific and

FIGURE 5 Prognosis analysis in training and validation datasets. Kaplan-Meier BCR free survival curves for patients in GSE70769 (A), GSE106918 (C), and TCGA datasets (E) based on high- and low-risk groups divided by risk score. ROC curve analyses of prognostic prediction for patients in GSE70769 (B), GSE106918 (D), and TCGA datasets (F). BCR, biochemical recurrence; ROC curve, receiver operational characteristic curve; TCGA, The Cancer Genome Atlas
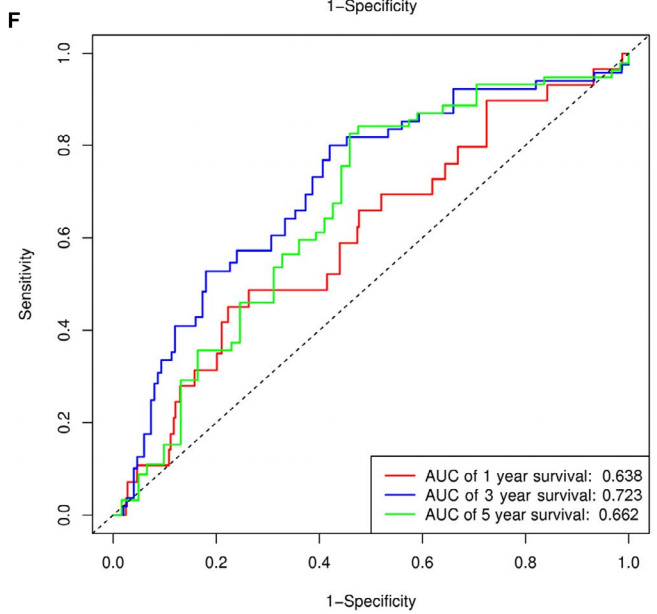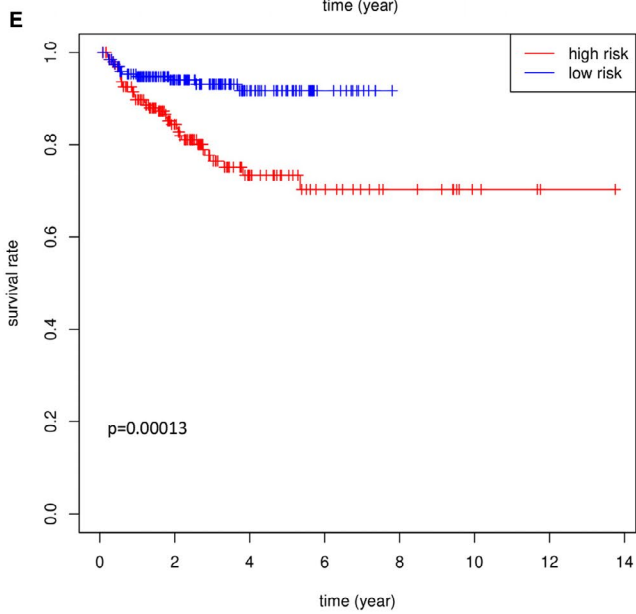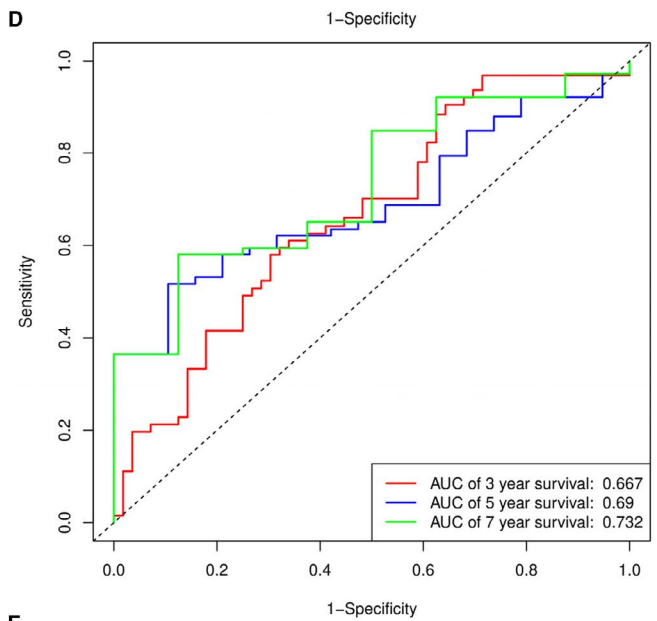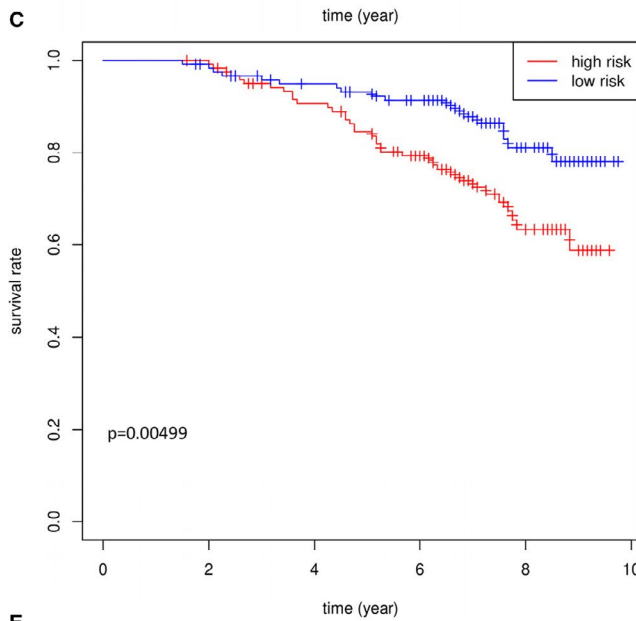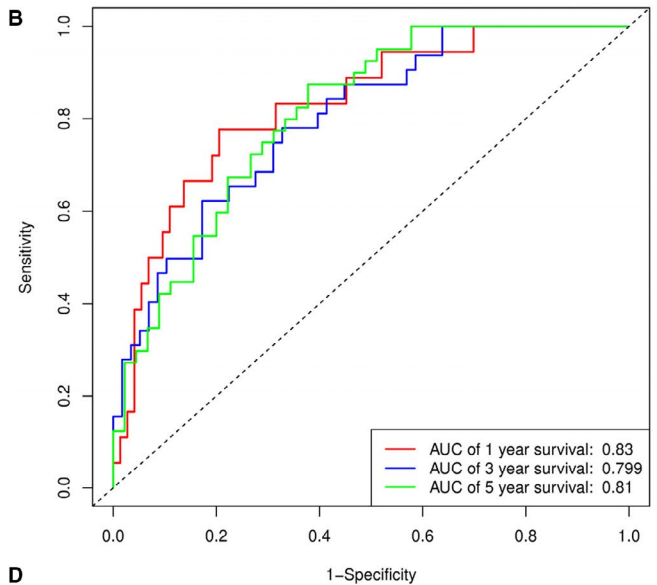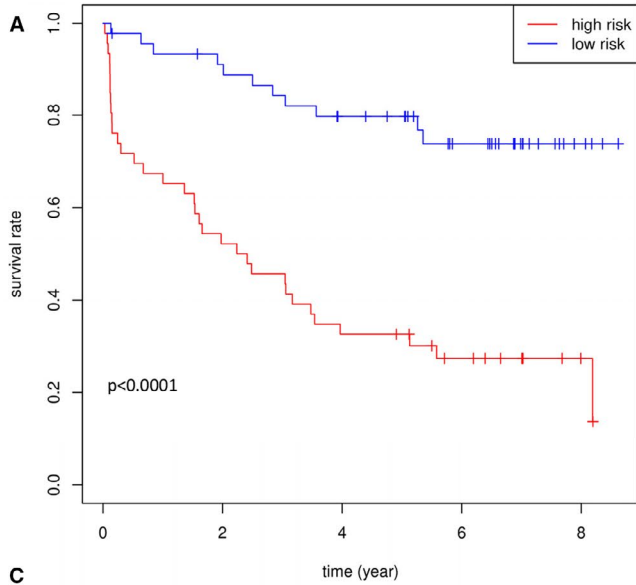
overall mortality.[6] Though some studies have researched gene expression profiles in PCa, systematic evaluation of molecular patterns as well as clinical information in PCa is limited.

To expound the potential biomarkers in PCa to a large extent, in the present study, we conducted an integrated analysis across multiple microarrays in GEO datasets. Significantly altered genes were identified based on robust rank aggregation.[30] Then, a total of 270 DEGs involving 148 downregulated and 122 upregulated genes were obtained. GO enrichment analysis showed DEGs were significantly associated with 23 GO terms.

Cell adhesion [31] and extracellular matrix [32] are well-known components during the progression and metastasis of PCa. Peptidase activities also have a tight relationship with PCa. The kallikrein (KLK)-related peptidase gene family plays an important role in the biological metabolism.[33] KLK3 (PSA) is the most commonly used clinical marker for PCa at present.[34] Besides, KLK2 and KLK4 are predominantly prostate specific and regulated by androgens.[35] In KEGG analysis, DEGs mainly participated in Wnt signaling pathway, glutathione, and cytochrome P450 metabolism. Studies related to PCa have illustrated the underlying importance of molecular alterations
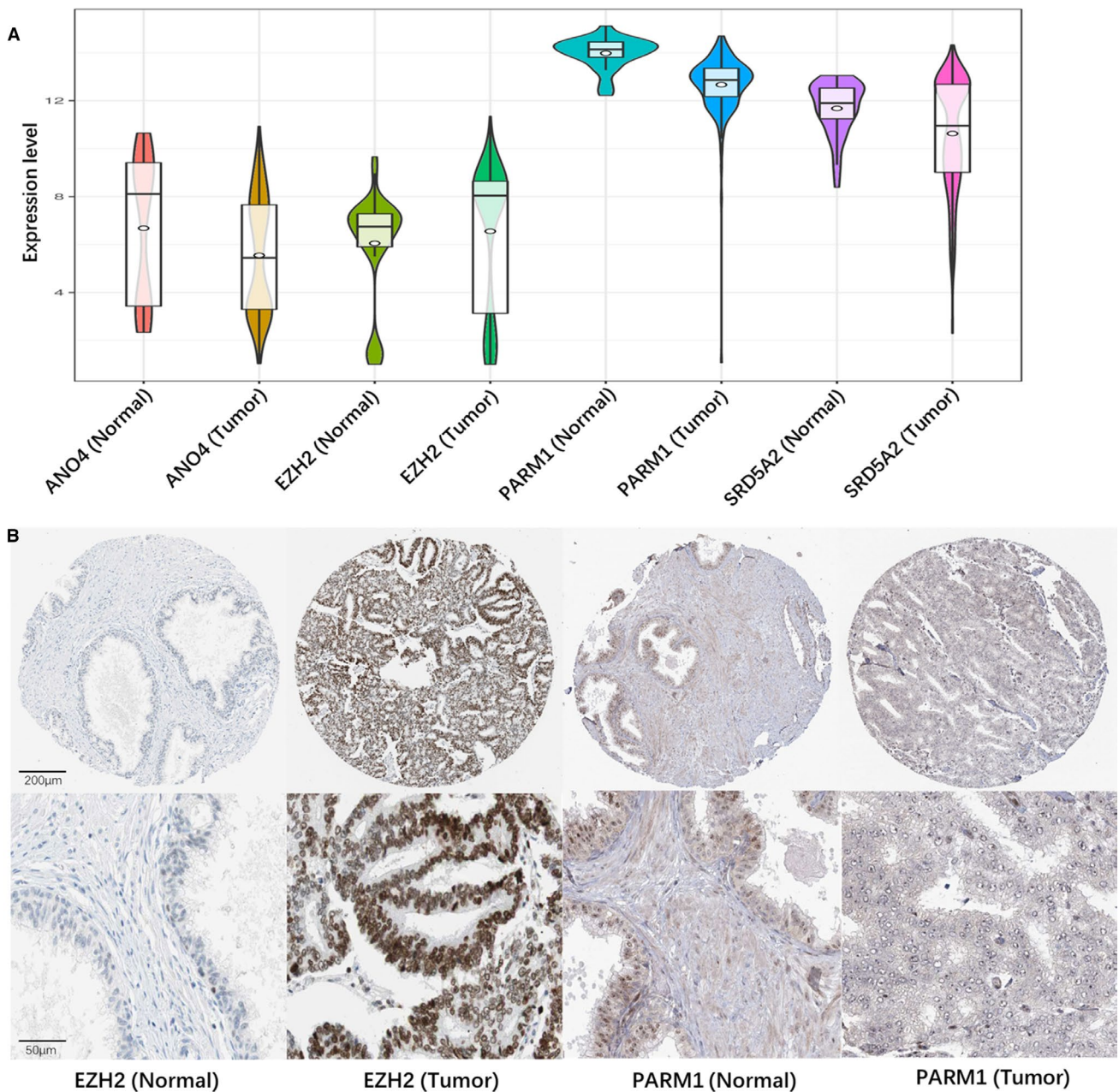
FIGURE 6 The expression profiles of prognostic genes in TCGA and HPA datasets. A, The expression levels after log2 transformation of the four genes in the TCGA prostate cancer RNA-seq dataset. B, The immunohistochemical results of EZH2 and PARM1 in HPA. TCGA, The Cancer Genome Atlas; HPA, Human Protein Atlas; log2 transformation, base-2 logarithm transformation; RNA-seq, RNA sequencing
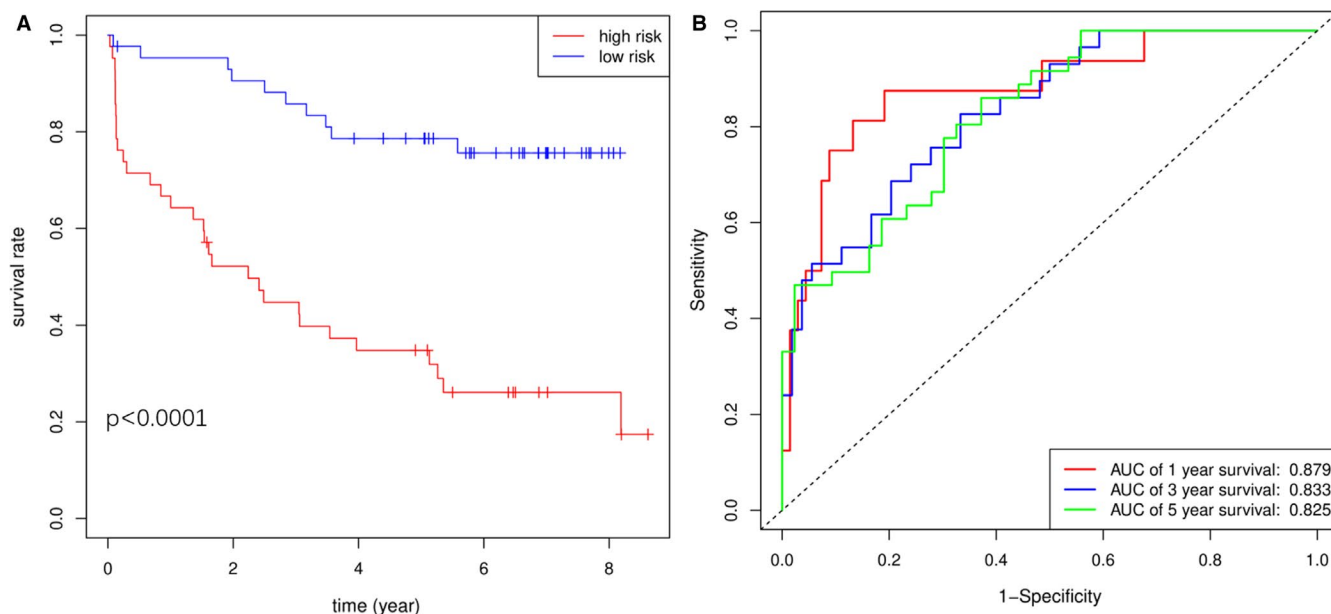
FIGURE 7 Prognosis analysis based on the combined model including four genes and clinical features. A, Kaplan-Meier BCR free survival curves for patients in high- and low-risk groups based on the combined model. B, ROC curve analyses of the combined model. BCR, biochemical recurrence; ROC curve, receiver operational characteristic curve

in Wnt signaling pathway.[36] In PCa microenvironment, Wnt-related proteins derived from the tumor promote resistance to therapy.[37] Besides, Wnt-β catenin facilitates self-renewing in PCa progenitors.[37] Glutathione plays a pivotal role in the development of cancer.[38] Glutathione metabolism is involved in the proliferation of PCa cells.[39] As for cytochrome P450 metabolism, Chang et al[40] reported that CYP1B1 inhibition suppressed tumorigenicity of PCa via caspase-1 activation. And CYP17A1 was reported to be involved in the biosynthesis of androgen in human.[41]

To reveal the clinical profiles of DEGs, a co-expression network by WCGNA was constructed between DEGs and clinical traits. Although only seven gene modules were obtained, three of them had a strong correlation with BCR status and BCR free time. These revealed the prognostic profile in DEGs. After univariate, LASSO and multivariate Cox regression analyses, ANO4, EZH2, PARM1, and SRD5A2 were screened out as prognostic genes ultimately. Then, a four-gene prognostic model was constructed and patients were grouped based on risk score. The AUCs of the ROC curve for predicting the 1, 3, and 5-year BCR free survival were 0.83, 0.799, and 0.81, respectively, indicating that the model had a good performance for prognostic prediction. Expression profiles of four genes in TCGA showed similar trend with GEO datasets. Furthermore, the predictive effect of the model was well validated in another PCa cohort and TCGA datasets.

Anoctamins, known as transmembrane 16 proteins, are a family of calcium-activated chloride channels and involved in many biological processes, including membrane excitability, ion homeostasis, and cell proliferation.[42] In cancer research,

Britschgi et al[43] found ANO1 promoted breast cancer progression by CAMK and EGFR pathways. Liu et al[44] reported that upregulation of ANO1 is involved in the pathological process of metastatic PCa and inhibition of ANO1 is a promising method in PCa therapy. ANO4 colocalizes with the endoplasmic reticulum Ca^{2+}-ATPase and reduces Ca^{2+} store release, probably acting as a leakage channel.[45] An analysis of single-nucleotide polymorphisms in anoctamin genes showed that ANO4 gene expression had a protective effect on the prognosis of PCa,[46] which is consistent with our results. However, the function of ANO4 in the pathological process of PCa is unknown, and molecular experiments need conducting to explore it further.

Enhancer of zeste homolog 2 (EZH2) is the enzymatic subunit of Polycomb repressive complex 2 related to transcriptional silencing.[47] EZH2 alterations have been associated with cancer progression.[48] EZH2 overexpression correlated with poor prognosis in tumors including prostate, breast, and bladder cancer.[49] In PCa, EZH2 is reported to methylate the androgen receptor (AR) and modulate AR recruitment.[50] After inhibiting EZH2, proliferation of PCa cells decreased and antitumor activity of AR antagonists in castration-resistant PCa increased.[49,51] Also, EZH2 inhibition was powerful to prevent the progression of neuroendocrine PCa.[52] Nowadays, EZH2 inhibitors are being evaluated in PCa patients such as CPI-1205.[48] It will be significant to improve the clinical outcome in PCa patients.

Prostate androgen-regulated mucin-like protein 1 (PARM1) is a member of the mucin family and is expressed at the surface of epithelial cells to promote cell survival.[53] PARM1 was reported to promote cardiomyogenic differentiation through the
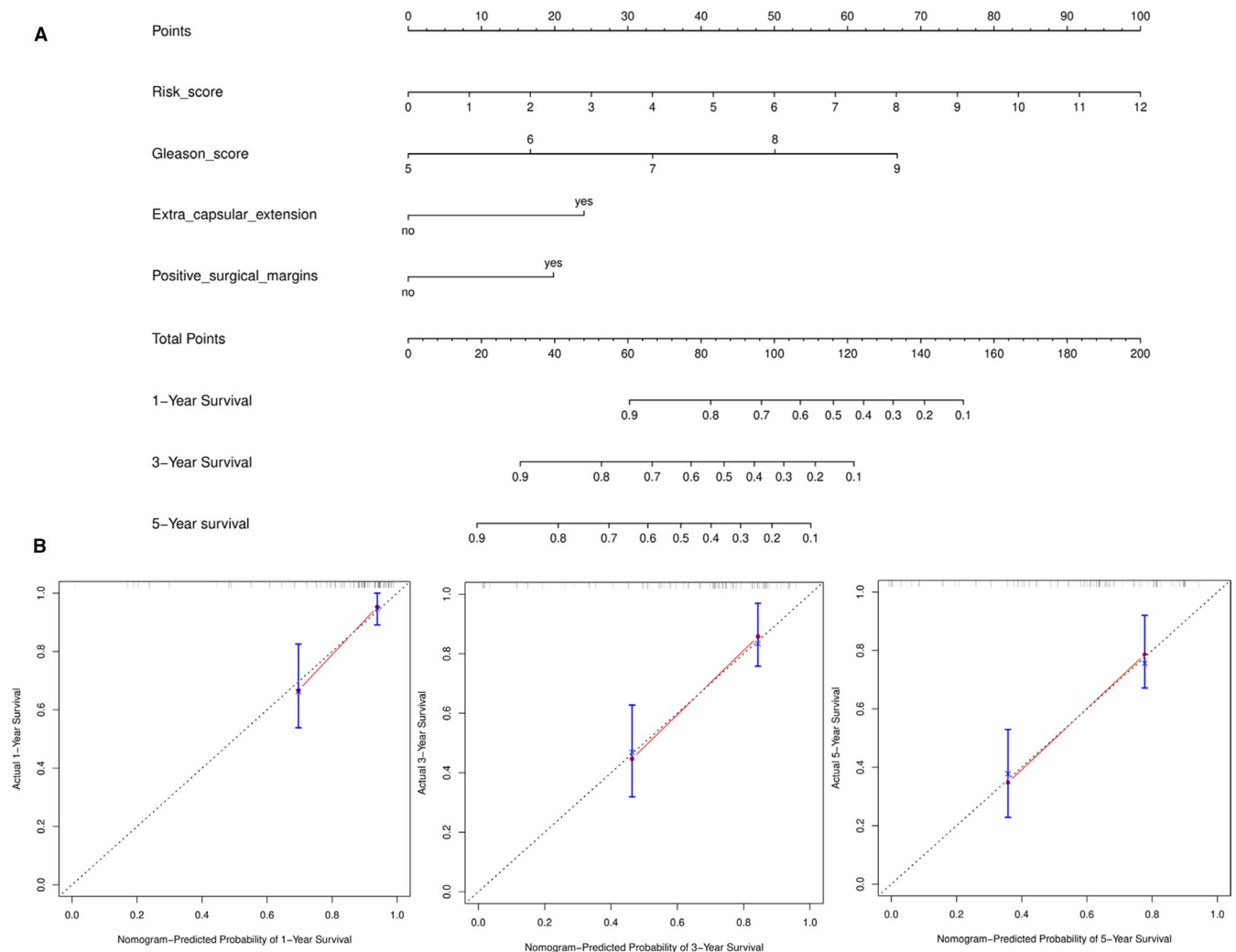
FIGURE 8 Nomogram evaluating prognosis based on the combined model. A, The 1-, 3-, and 5-y BCR free survival could be evaluated by adding up the points of four-gene risk score, Gleason score and events of extra capsular extension and positive surgical margins. B, The calibration curves for predicting 1-, 3-, and 5-y BCR free survival for patients with prostate cancer

Smad signaling pathway.[54] It is also involved in the cancer development. PARM1 could suppress the proliferation of colorectal cancer cells.[55] In PCa, Shola et al [56] found that PARM1 served as tumor suppressor to induce apoptosis of cancer cells through the Smad signaling pathway. In this study, downregulation of PARM1 showed a poor prognostic outcome in PCa patients and its expression was significantly lower in PCa issue compared with benign issue in TCGA and HPA datasets. However, it is also reported that PRAM1 enhanced cell growth in leukemia.[53] The specific function of PARM1 in PCa need to be expounded further.

Steroid 5 alpha-reductase 2 (SRD5A2) encodes a microsomal protein expressed at high levels in androgen-sensitive tissues like prostate. In our study, SRD5A2 was regarded as a protective gene in the prognosis of PCa. SRD5A2 is involved in converting testosterone to dihydrotestosterone in prostate cells.[57] Aggarwal et al[58] reported that SRD5A2 could reduce cell migration and invasion by indirectly regulating ERK/MAPK pathway in PCa cells. Also, the expression of SRD5A2 was often found to be downregulated as androgen dependency is lost in advanced stages of metastasis.[59] Furthermore, SRD5A2 polymorphism could be a promising biomarker for metastatic PCa patients treated with primary androgen-deprivation therapy.[60] Therefore, SRD5A2 may serve as a molecular target in the advanced PCa.

In a word, the four prognostic genes obtained in our study were well validated in theory and multiple datasets. It would be significant for the guidance of prognosis in PCa. Considering that traditional clinical parameters are also associated with prognosis, we combined four-gene model and clinical factors to analyze the prognosis of PCa to a largest extent. Finally, we obtained a combined prognostic model including four genes and Gleason score, extra capsular extension, and positive surgical margins. Expectedly, the AUCs of the ROC curve for prognosis were higher than four-gene model, which achieved a better predictive value. Based on the nomogram, prognostic probability for each patient could

be predicted according to the corresponding parameters. A personalized evaluation was achieved in a sense.

There are some limitations in this study. First, the clinical data for prognostic analysis.

is not large enough, so prospective studies should be conducted further. Second, the specific mechanisms of four genes in PCa are need to be explored in molecular experiments.

In conclusion, we performed an integrated analysis in multiple microarrays and constructed a four-gene prognostic model with ANO4, EZH2, PARM1, and SRD5A2 in PCa cohort. Then, the gene model was well validated in other datasets. Finally, combination of four-gene model and clinical features achieved a better prediction to guide the prognosis of patients with PCa.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

No potential conflicts of interest are to be disclosed by authors.

AUTHOR CONTRIBUTION

Conceptualization: PY and BL; Data curation and methodology: LL, XG, TS, JM, XY, and JL; Formal analysis: PY, LL, and QF; Funding acquisition: BL; Writing-original draft: PY and BL; Writing-review & editing: JM, XY, and JL; Supervision: JL and BL.

ETHICAL APPROVAL

This study were approved by our Institutional Research Ethics Committee (Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology).

DATA AVAILABILITY STATEMENT

The datasets involved in our study were extracted from TCGA (https://portal.gdc.cancer.gov/), GEO (https://www.ncbi.nlm.nih.gov/geo), and HPA (https://www.proteinatlas.org/). All the data we used are available in our study.

ORCID

Bo Liu https://orcid.org/0000-0001-5118-1890

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424.

2. Teoh JYC, Hirai HW, Ho JMW, Chan FCH, Tsoi KKF, Ng CF. Global incidence of prostate cancer in developing and developed countries with changing age structures. *PLoS One*. 2019;14(10):e0221775.

3. Norum J, Nieder C. Treatments for metastatic prostate cancer (mPC): a review of costing evidence. *Pharmacoeconomics*. 2017;35(12):1223-1236.

4. Artibani W, Porcaro AB, De Marco V, Cerruto MA, Siracusano S. Management of biochemical recurrence after primary curative treatment for prostate cancer: a review. *Urol Int*. 2018;100(3):251-262.

5. Kontos CK, Adamopoulos PG, Scorilas A. Prognostic and predictive biomarkers in prostate cancer. *Expert Rev Mol Diagn*. 2015;15(12):1567-1576.

6. Van den Broeck T, van den Bergh RCN, Arfi N, et al. Prognostic value of biochemical recurrence following treatment with curative intent for prostate cancer: a systematic review. *Eur Urol*. 2019;75(6):967-987.

7. Barbieri CE, Baca SC, Lawrence MS, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet*. 2012;44(6):685-689.

8. Robinson D, Van Allen EM, Wu YM, et al. Integrative clinical genomics of advanced prostate cancer. *Cell*. 2015;161(5):1215-1228.

9. Latonen L, Afyounian E, Jylha A, et al. Integrative proteomics in prostate cancer uncovers robustness against genomic and transcriptomic aberrations during disease progression. *Nat Commun*. 2018;9(1):1176.

10. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57-63.

11. Taylor BS, Schultz N, Hieronymus H, et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell*. 2010;18(1):11-22.

12. Arredouani MS, Lu B, Bhasin M, et al. Identification of the transcription factor single-minded homologue 2 as a potential biomarker and immunotherapy target in prostate cancer. *Clin Cancer Res*. 2009;15(18):5794-5802.

13. Mortensen MM, Hoyer S, Lynnerup AS, et al. Expression profiling of prostate cancer tissue delineates genes associated with recurrence after prostatectomy. *Sci Rep*. 2015;5:16018.

14. Penney KL, Sinnott JA, Tyekucheva S, et al. Association of prostate cancer risk variants with gene expression in normal and tumor tissue. *Cancer Epidemiol Biomarkers Prevent*. 2015;24(1):255-260.

15. Brase JC, Johannes M, Mannsperger H, et al. TMPRSS2-ERG-specific transcriptional modulation is associated with prostate cancer biomarkers and TGF-beta signaling. *BMC Cancer*. 2011;11:507.

16. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.

17. Liu X, Wu J, Zhang D, et al. Identification of potential key genes associated with the pathogenesis and prognosis of gastric cancer based on integrated bioinformatics analysis. *Front Genet*. 2018;9:265.

18. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-287.

19. Walter W, Sanchez-Cabo F, Ricote M. GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics (Oxford, England)*. 2015;31(17):2912-2914.

20. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.

21. Ross-Adams H, Lamb AD, Dunning MJ, et al. Integration of copy number and transcriptomics provides risk stratification in prostate

cancer: a discovery and validation cohort study. *EBioMedicine*. 2015;2(9):1133-1144.

22. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1-22.

23. Long J, Zhang L, Wan X, et al. A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma. *J Cell Mol Med*. 2018;22(12):5928-5938.

24. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. 2013;32(30):5381-5397.

25. Jain S, Lyons CA, Walker SM, et al. Validation of a Metastatic Assay using biopsies to improve risk stratification in patients with prostate cancer treated with radical radiation therapy. *Ann Oncol*. 2018;29(1):215-222.

26. Chen Z, Hu H. Identification of prognosis biomarkers of prostatic cancer in a cohort of 498 patients from TCGA. *Curr Probl Cancer*. 2019;43(6):100503.

27. Asplund A, Edqvist PH, Schwenk JM, Ponten F. Antibodies for profiling the human proteome—the Human Protein Atlas as a resource for cancer research. *Proteomics*. 2012;12(13): 2067-2077.

28. Jiang W, Guo Q, Wang C, Zhu Y. A nomogram based on 9-lncRNAs signature for improving prognostic prediction of clear cell renal cell carcinoma. *Cancer Cell Int*. 2019;19:208.

29. Cheng B, He Q, Cheng Y, et al. A three-gene classifier associated with microrna-mediated regulation predicts prostate cancer recurrence after radical prostatectomy. *Front Genet*. 2019;10:1402.

30. Zhang Y, Tao Y, Ji H, et al. Genome-wide identification of the essential protein-coding genes and long non-coding RNAs for human pan-cancer. *Bioinformatics (Oxford, England)*. 2019;35(21):4344-4349.

31. Kolonin MG, Sergeeva A, Staquicini DI, et al. Interaction between tumor cell surface receptor RAGE and proteinase 3 mediates prostate cancer metastasis to bone. *Can Res*. 2017;77(12):3144-3150.

32. Walker C, Mojares E, Del Rio HA. Role of extracellular matrix in development and cancer progression. *Int J Mol Sci*. 2018;19(10):1–31.

33. Gan L, Lee I, Smith R, et al. Sequencing and expression analysis of the serine protease gene cluster located in chromosome 19q13 region. *Gene*. 2000;257(1):119-130.

34. Lai J, An J, Nelson CC, Lehman ML, Batra J, Clements JA. Analysis of androgen and anti-androgen regulation of KLK-related peptidase 2, 3, and 4 alternative transcripts in prostate cancer. *Biol Chem*. 2014;395(9):1127-1132.

35. Lawrence MG, Stephens CR, Need EF, Lai J, Buchanan G, Clements JA. Long terminal repeats act as androgen-responsive enhancers for the PSA-kallikrein locus. *Endocrinology*. 2012;153(7):3199-3210.

36. Murillo-Garzon V, Kypta R. WNT signalling in prostate cancer. *Nat Rev Urol*. 2017;14(11):683-696.

37. Sun Y, Zhu D, Chen F, et al. SFRP2 augments WNT16B signaling to promote therapeutic resistance in the damaged tumor microenvironment. *Oncogene*. 2016;35(33):4321-4334.

38. Bansal A, Simon MC. Glutathione metabolism in cancer progression and treatment resistance. *J Cell Biol*. 2018;217(7):2291-2298.

39. Yu YP, Yu G, Tseng G, et al. Glutathione peroxidase 3, deleted or methylated in prostate cancer, suppresses prostate cancer growth and metastasis. *Can Res*. 2007;67(17):8043-8050.

40. Chang I, Mitsui Y, Kim SK, et al. Cytochrome P450 1B1 inhibition suppresses tumorigenicity of prostate cancer via caspase-1 activation. *Oncotarget*. 2017;8(24):39087-39100.

41. Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr Rev*. 2011;32(1):81-151.

42. Hartzell C, Putzier I, Arreola J. Calcium-activated chloride channels. *Annu Rev Physiol*. 2005;67:719-758.

43. Britschgi A, Bill A, Brinkhaus H, et al. Calcium-activated chloride channel ANO1 promotes breast cancer progression by activating EGFR and CAMK signaling. *Proc Natl Acad Sci USA*. 2013;110(11):E1026-E1034.

44. Liu W, Lu M, Liu B, Huang Y, Wang K. Inhibition of Ca(2+)-activated Cl(−) channel ANO1/TMEM16A expression suppresses tumor growth and invasiveness in human prostate carcinoma. *Cancer Lett*. 2012;326(1):41-51.

45. Cabrita I, Benedetto R, Fonseca A, et al. Differential effects of anoctamins on intracellular calcium signals. *FASEB J*. 2017;31(5):2123-2134.

46. Yu CC, Chen LC, Huang CY, et al. Genetic association analysis identifies a role for ANO5 in prostate cancer progression. *Cancer Med*. 2020;9(7):2372-2378.

47. Kim KH, Roberts CW. Targeting EZH2 in cancer. *Nat Med*. 2016;22(2):128-134.

48. Saad F, Shore N, Zhang T, Sharma S, Cho HK, Jacobs IA. Emerging therapeutic targets for patients with advanced prostate cancer. *Cancer Treat Rev*. 2019;76:1-9.

49. Varambally S, Dhanasekaran SM, Zhou M, et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature*. 2002;419(6907):624-629.

50. Kim E, Kim M, Woo DH, et al. Phosphorylation of EZH2 activates STAT3 signaling via STAT3 methylation and promotes tumorigenicity of glioblastoma stem-like cells. *Cancer Cell*. 2013;23(6):839-852.

51. Comet I, Riising EM, Leblanc B, Helin K. Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. *Nat Rev Cancer*. 2016;16(12):803-810.

52. Dardenne E, Beltran H, Benelli M, et al. N-Myc induces an EZH2-mediated transcriptional program driving neuroendocrine prostate cancer. *Cancer Cell*. 2016;30(4):563-577.

53. Charfi C, Levros LC Jr, Edouard E, Rassart E. Characterization and identification of PARM-1 as a new potential oncogene. *Mol Cancer*. 2013;12:84.

54. Nakanishi N, Takahashi T, Ogata T, et al. PARM-1 promotes cardiomyogenic differentiation through regulating the BMP/Smad signaling pathway. *Biochem Biophys Res Comm*. 2012;428(4):500-505.

55. Beck SE, Jung BH, Del Rosario E, Gomez J, Carethers JM. BMP-induced growth suppression in colon cancer cells is mediated by p21WAF1 stabilization and modulated by RAS/ERK. *Cell Signal*. 2007;19(7):1465-1472.

56. Shola DT, Wang H, Wahdan-Alaswad R, Danielpour D. Hic-5 controls BMP4 responses in prostate cancer cells through interacting with Smads 1, 5 and 8. *Oncogene*. 2012;31(19):2480-2490.

57. Russell DW, Wilson JD. Steroid 5 alpha-reductase: two genes/two enzymes. *Annu Rev Biochem*. 1994;63:25-61.

58. Aggarwal S, Singh M, Kumar A, Mukhopadhyay T. SRD5A2 gene expression inhibits cell migration and invasion in prostate cancer cell line via F-actin reorganization. *Mol Cell Biochem.* 2015;408(1-2):15-23.

59. Habib FK, Ross M, Bayne CW, Bollina P, Grigor K, Chapman K. The loss of 5alpha-reductase type I and type II mRNA expression in metastatic prostate cancer to bone and lymph node metastasis. *Clin Cancer Res.* 2003;9(5):1815-1819.

60. Shiota M, Fujimoto N, Yokomizo A, et al. The prognostic impact of serum testosterone during androgen-deprivation therapy in patients with metastatic prostate cancer and the SRD5A2 polymorphism. *Prostate Cancer Prostatic Dis.* 2016;19(2):191-196.

SUPPORTING INFORMATION

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