Identification and validation of a prognostic 5-protein signature for biochemical recurrence following radical prostatectomy in prostate cancer

Xiangkun Wu
Guangzhou Medical University
Wenjie Li
Guangzhou Medical University
Daojun Lv
Guangzhou Medical University
Yongda Liu (13719007083@163.com)
Guangzhou Medical University
Di Gu (sveong@163.com)
Guangzhou Medical University https://orcid.org/0000-0002-2411-8164

Primary research

Keywords: biochemical recurrence, prostate cancer, prognosis, signature, nomogram

Posted Date: May 20th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-29488/v1

Abstract

Background : Biochemical recurrence (BCR) is considered as an indicator for prostate cancer (PCa)-specific recurrence and mortality. However, lack of effective prediction model to assess the prognosis of patients for optimization of treatment. The aim of this work was to construct a protein-based nomogram that could predict BCR for PCa.

Materials and methods: Univariate Cox regression analysis was conducted to identify candidate proteins from the Cancer Genome Atlas (TCGA) database. LASSO Cox regression was further conducted to pick out the most significant prognostic proteins and formulate the proteins signature for predicting BCR. Additionally, a nomogram was constructed by multivariate Cox proportional hazards regression.

Results: We established a 5-protein-based signature which was well used to identify PCa patients into high- and low-risk groups. Kaplan-Meier analysis demonstrated patients with higher BCR generally had significantly worse survival than those with lower BCR ($p<0.0001$). Time-dependent receiver operating characteristic curve expounded that ours signature had excellent prognostic efficiency for 1-, 3- and 5-year BCR (area under curve in training set: 0.691, 0.797, 0.808 and 0.74, 0.739, 0.82 in the test set). Univariable and multivariate Cox regression analysis showed that this 5-protein signature was an independent of several clinical signatures including age, Gleason score, T stage, N status, PSA and residual tumor. Moreover, a nomogram was constructed and calibration plots confirmed the its predictive value in 3-, 5- and 10-year BCR overall survival.

Conclusion: Our study identified a 5-protein-based signature and constructed a prognostic nomogram that reliably predicts BCR in prostate cancer. The findings might be of paramount importance in tumor prognosis and medical decision-making.

Introduction

Prostate cancer (PCa) is the second leading cause of tumor death among American male, accounting for 20% newly-diagnosed cancer with 31,620 deaths in 2019 [1]. Although radical prostatectomy (RP) is considered to be an effective therapy for PCa patients, recent work reveals that approximately 20%-40% patients suffered from biochemical recurrence (BCR) after RP [2]. BCR is defined as recurrent prostate specific antigen (PSA) more than 0.2 μg/L and an indicator for distant metastasis or PCa-specific mortality [2, 3]. 32%-45% among BCR patients with post-RP would die from PCa within 15 years [4]. Only those with high risk (with 8-10 Gleason score or a PSA doubling time of <12 months) would benefit from salvage treatment [5]. Thus, early identification of PCa patients with high BCR risk have great importance for management strategies.

Known to relate to the survival of PCa patients, Gleason score, PSA, and tumor stage are currently recommended to facilitate the prediction of BCR [6]. However, the prediction based on clinicopathological features has a limited accurateness [7]. Patients with the same clinicopathological features might undergo infinitely different clinical endpoints [8]. Thus, accumulating nuanced BCR risk stratifications

have been developed by using additional molecular biomarkers. Strand et al.[9] find several DNA methylation markers may be related to PCa initiation and progression. Further evidence demonstrates that miRNAs and lncRNAs might be crucial regulators of biologic developments in PCa progression, having potential as novel biomarkers for BCR [10].

It is well acknowledged that in malignant prostate disease multiple, although biochemical processes from DNA to protein are influenced by many complicated biological factors, proteins directly reveal functions of genes. Moreover, the correlation between levels of protein and mRNA expression is not always observed [11]. Thus, aggressiveness of tumors may be better correlated with the quantification of protein [12]. However, little work to date addresses the potential function of protein-based signatures to predict BCR in PCa. In the present study, in order to predict the BCR of PCa patients, we firstly proposed to 1) construct a multi-protein-based signature and nomogram with combination of clinicopathological variables, thus aiding to prognosis BCR for PCa patients, 2) validate the predictive ability throughout time-dependent receiver operating characteristic (ROC) curve, calibration plots, concordance index (C-index), and decision curve analysis (DCA), 3) and perform GO (Gene Ontology) pathway enrichment analyses and Gene Set Enrichment Analysis (GSEA) to investigate the potential biological functions.

Materials And Methods

Collection of Protein and Clinical Data

The preset endpoint of our work was BCR after RP. The data acquisition and application were performed based on the publication policies and guidelines for the protection of human subjects.

Reverse-phase protein array (RPPA) protein data for prostate adenocarcinoma was searched and derived from the Cancer Proteome Atlas (TCPA) [13]. PCa samples for all trials were from RP specimens. ‘Impute’ package in R was applied to pre-process the protein profile data as well as fill lacking values. Before use all selected expression datasets were standardized.

The Cancer Genome Atlas (TCGA) was used to download clinical data and genes sequencing information (Fragments Per Kilobase Of Exon Per Million Fragments Mapped, FPKM) from PCa samples. PCa Patients irrespectively of age, vital status, tumor type, race and ethnicity were included from this study. After analyzing the download data, patients without data integrality were eliminated: 1) with follow-up periods less than 30 days, and 2) deficient clinical information of age, recurrence, TNM stage. Finally, a total of 341 patients (recurrence-free 282 and 59 with BCR) were selected for further analysis and randomly assigned to the test set (n=172) and training set (n=169), respectively.

Identification and validation of the multitype-protein-based prognostic signature

Univariate Cox regression analysis was applied to identify the prognostic protein signature in the training set. To construct prognostic protein signature, only proteins with $|\log_2$ fold-change (FC)$| > 2$ and $P < 0.05$ were marked as significantly expressed proteins and deemed as candidate proteins for further model

construction. Log-rank tests and Kaplan–Meier (KM) survival curve analysis were subsequently used to investigate the prognostic capability of identified proteins after univariate Cox regression analysis. Using 'glmnet' package in R, least absolute shrinkage and selection operator (LASSO)-penalized cox regression analysis was conducted to further narrow proteins for predicting BCR in the training set. Regression coefficients and coefficients of other unrelated variables were set as zero by LASSO in terms of the regulation weight λ [14]. The best λ was gotten base on minimum cross-validation mistake using 10-fold cross-validation. In terms of the optimal λ value, a list of prognostic proteins with related coefficients was screened from the levels of protein profiling and BCR data.

In order to investigate the prognostic accuracy of final proteins, hierarchical clustering analysis was conducted to classify the data in terms of the similar expression patterns throughout 'heatmap' package in R in the training dataset. The cut-off risk score was calculated by median risk score using Survival in R because it is easy to for clinical application. Patients in each dataset were allocated into either low- or high-risk group using mean risk score based on proteins signature. In order to further prove the discriminatory power of the protein-signature, we conducted the KM analysis with log-rank test, ROC curve and C-index.

Univariate and multivariate Cox regression analysis of BCR in all datasets were conducted to explore whether this protein-base-signature was independent of age, PSA, Gleason scores and tumor stage in each dataset. KM curves were applied to investigate the diagnostic value of multi-protein prognostic signature in different clinical features, including age, Gleason score, T stage, N status, PSA and residual tumor. Additionally, the area under curve (AUC) were calculated by time-dependent ROC curve to compare the prognostic performance between the protein-based signature and above independent clinical variables in the training, test and entire sets.

Construction and validation of predictive nomogram

Using the package of 'rms' in R, all independent prognostic proteins and relevant clinical data were included in the construction of a prognostic nomogram throughout a multivariate Cox regression analysis in the entire set. The stepwise method was applied to pick out the most ideal model, and a risk score was assigned to the expression level of each prognostic protein and coefficients weighted throughout the penalized Cox model.

In order to investigate the predictive accuracy among the nomogram, the signature and other clinicopathological factors including age, Gleason score, T stage, N status, PSA and residual tumor, the AUC of ROC was performed. Calibration plots were conducted to further assess the predictive performance of the nomogram. Prognostic results and outcomes of the nomogram were shown in the calibrate curve, in which the 45° line represented the best prediction. Bootstraps for 200 resamples were applied to these activities. C-index was subsequently performed to explore the predicted probabilities of the nomogram. Finally, a decision curve analysis (DCA) was used so as to investigate the clinical net benefit of diverse probability thresholds for possible clinical outcomes and reliability of the nomogram [15].

Functional enrichment analysis and identification of proteins

We first calculated the Pearson correlation coefficient between levels of proteins from the signature and expression of corresponding genes. With screened proteins after univariate Cox regression analysis, we performed GO pathway enrichment analysis. A *p*-value of < 0.05 was considered as the cut-off criteria. GSEA was conducted to investigate potential biological functions throughout GSEA software on JAVA platform for proteins used for the construction of the signature[16]. We obtained the annotated gene set c2.cp.kegg.v7.0.symbols.gmt by using the Molecular Signatures Database (MSigDB), which was latter selected as the reference set to investigate whether genes are distributed randomly or enriched in high/low-expression groups. FDR<0.05 and p<0.05 were preset as the cut-off criteria to identify an enriched group. To obtain co-expression proteins, Sankey diagram was performed by using 'ggplot2' package in R.

Statistical analysis

Patients with PCa were randomly assigned into high- and low-risk groups by the 'caret' package of R (set. Seed =1,000, p = 0.50). A signature was constructed by univariate Cox regression analysis throughout 'Survival' package in R. To evaluate the prognostic power of multi-protein-based signature, time-dependent receiver operating characteristic (ROC) analysis and concordance index (C-index) were conducted based on the 'survival ROC' R package. BCR, from the day of RP to the time of recurrence or the date censored, was calculated in terms of KM model, and the log-rank test was used to evaluate the statistical differences between high/low-risk groups. Multivariate Cox regression model was conducted for the formulation of nomogram using 'rms' package. Hazard ratios (HR) and 95% confidence intervals were obtained to identify proteins related to BCR. Calibration plots were used to investigate whether actual outcomes approximate discriminatory power for nomogram. All statistical calculations were conducted with R 3.5.0 software.

Result

Collection of samples and patients

After exclusion of samples without adequate clinical information, totally 341 patients were included in the present study from TCGA. The average age was 61.5 years and median follow-up was 2.68 years. Moreover, 30 of 341 patients had low Gleason scores (≤6) PCa and 140 patients had high Gleason scores (≥8) PCa. All patients were undergone RP and pathologically diagnosed with BRC, with assigned to training set (n=169) and testing set (n=172), respectively.

Prognostic proteins were screened by univariate Cox regression analysis to filtered out proteins correlated with BCR. In terms of the cut-off standard of P value <0.05 and $|\log_2 FC| > 2$, totally 21 proteins significantly related to BCR were identified finally. As clearly shown in **Supplementary Figure S1**, KM analyses expounds the accuracy of the 21 selected proteins and discriminative power for further analysis (p<0.05).

Development and validation of proteins signature

LASSO Cox regression analysis was conducted to construct a prognostic model in the training set, which picked out 5 proteins (alpha-Catenin, BRD4, DJ1, SMAD1 and YB1) after initial filtration of univariate Cox regression identified 21 proteins **(Figure 1)**. An equation to calculate the risk score for their BCR risk was derived according to selected levels of five proteins weighted by the regression coefficients, as following: risk score = (-2.771× levels of alpha-Catenin) + (1.577× levels of BRD4) + (-2.239× levels of DJ1) + (2.152× levels of SMAD1) + (2.428× levels of YB1). Among these five prognostic proteins, three (BRD4, SMAD1, YB1) demonstrated positive coefficients, suggesting high expression levels were associated with high-risk BCR. Two (alpha-Catenin and DJ1) in the Cox regression analysis showed negative coefficient, indicating that their high expression levels were related to better BCR.

To investigate the predictive performance of the signature, patients were assigned into the high- and low-risk groups based on median risk score as cut-off value of each protein. PCa patient cohorts with a risk score of 1.804 or lower were divided into the low-risk group, otherwise the others belonged to the high-risk group **(Figure 2A)**. For the BCR, the higher risk score meant worse prognosis. Thereby, the higher levels of proteins with positive weighting coefficient suggested higher risk scores. The distribution of survival status of the PCa patients and the levels of proteins were analyzed. The results demonstrated that compared with low-risk groups, death of patients was significantly more in high-risk groups **(Figure 2B)**. The levels of proteins with positive coefficients were higher in high-risk groups **(Figure 2C)**. We also discovered that patients with high BCR risk inclined to express high-risk proteins, whereas samples with low BCR tended to express protective proteins.

Kaplan-Meier analysis demonstrated patients with higher BCR generally had significantly worse survival than those with lower BCR ($p<0.0001$) **(Figure 3)**. In the training set, AUCs of the 5-protein-based signature at 1-, 3- and 5-year survival times were 0.691, 0.797, 0.808 and 0.74, 0.739, 0.82 for the test set severally, suggesting that the prognostic signature had a great specificity and sensitivity **(Figure 3)**. The C-index of the signature was 0.679 (95%CI: 0.599 to 0.759) in the training set, 0.704 (95%CI: 0.613 to 0.794) in the test set and 0.693 (95%CI: 0.634 to 0.752) in the entire set.

Throughout the univariate and multivariate Cox proportional hazards regression analyses, the 5-protein predictive signature was confirmed to be independent of with other clinicopathological factors, including age, Gleason grades, T stage, N status, PSA and residual tumors in predicting the BCR-free survival **(Table 1)**. The KM survival analysis also indicated the discriminative capability of the signature in different clinical prognostic features **(Supplementary Figure S2).** To further evaluate the predictive accuracy between the signature and the other clinicopathological factors, we calculated the AUC of ROC which showed that the 5-protein-based signature had significantly better prognostic performance than any other clinical factors in the training, test and entire sets **(Figure 4).**

Identification and validation of the nomogram

The nomogram was conducted in the entire set by multivariate Cox regression analysis of 5 proteins with preset clinicopathological covariables, including age, Gleason grades, T stage, N status, PSA and residual tumors. The result demonstrated great prognostic performance in BCR of PCa patients **(Figure 5A)**. Calibration plots confirmed the predictive value of the prognostic nomogram in 3-, 5-and 10-year BCR overall survival (OS) **(Figure 5B)**, indicating the good agreement with the actual outcome. The C-index of the nomogram was 0.777 (95%CI: 0.699 to 0.855) in the training set, 0.771 (95%CI: 0.691 to 0.851) in the test set and 0.764 (95%CI: 0.701 to 0.827) in the entire set. Finally, net benefit curves outlined the nomogram was better than the signature and other clinicopathological factors (**Figure 5C**).

Validation and functional Characteristics of the 5 proteins

There were positive correlations between 5 proteins and corresponding genes by calculated the Pearson correlation coefficient (**Supplementary figure S3**). Outcomes of GO enrichment analysis indicated that these protein-related genes are enriched in immune- or cell differentiation-related GO terms (**Figure 6A**), suggesting that the effect of prognostic proteins on cancer might be related to the tumor microenvironment. Additionally, GSEA in TCGA database was conducted to ascertain the five proteins related biological signaling pathway between high- and low-risk groups (**Figure 6B**). According to significant protein sets (FDR<0.05 and p<0.05), five pathways were screened: 1) base excision repair, 2) DNA replication, 3) nucleotide excision repair, 4) pyrimidine metabolism, and 5) spliceosome. Finally, Sankey diagram revealed the association with co-expression proteins and the 5-protein signature, which may interact with each other by certain molecular mechanism (**Figure 6C**).

Discussion

The majority of patients with early BCR will develop clinical recurrence and require timely intervention [2]. Up to date, some novel biomarkers were identified for the BCR of PCa [6, 9, 10]. However, most of these studies were only concerned with one or a few genes, and little work was carried out for clinical predictive performance of proteins. Their clinical applicability was also restricted due to high cost [14]. Due to the heterogeneity of prostate cancers [17], a protein-base-panel of biomarkers may be more useful in prognosis malignant prostate disease than a single gene. To the best of our knowledge, it is the first time to construct a prognosis protein-based model for patients with PCa after RP.

In the present study, we established and validated the 5-protein-based signature (alpha-Catenin, BRD4, DJ1, SMAD1 and YB1) to prognosis BCR of patient with PCa. Initial concerns on the poor value of proteins is not supported in the present study [18], because significant correlations between gene expressions and levels of protein were observed. The results of KM curves and C-index revealed that 5-proteins-based signature may be of importance in categorizing patients into high- and low-risk BCR groups as an effective prognostic indicator. Shao et al [6]. established a 5-lncRNAs-based signature to predict PCa BCR with AUC=0.72. However, our 5-protein-based signature had the more clinical utility, outperforming the known model with AUC=0.809 in the entire cohort, 0.808 in the test cohort and 0.820 in the training cohort. Additionally, the current TNM staging system is closely associated with the prognosis

of PCa. Consistently, in the present study the univariable and multivariable Cox regression analysis also showed that tumor TN stage was a significant prognostic factor for PCa. Of note, our 5-protein signature turned out independent of tumor stage throughout KM analysis, suggesting the ability to discriminate PCa patients with high BCR risk from the stratified groups.

Furthermore, Greke et al. identified a four-protein cancer nomogram of PTEN, SPP1, SMAD4, and CCND1 that were related to lethal outcomes among PCa patients based on clinical features [19]. However, their nomogram failed to provide long-term and independent predictive information beyond clinical factors. In contrast, our 5-protein–based nomogram had greater discriminatory ability than other clinicopathological features. Moreover, calibration plots and DCA showed great predictive performance in 3-, 5-and 10-year BCR overall survival. Overestimating the risk of indolent PCa patients who are more appropriate for active surveillance implies a higher public health burden because of overtreatment [6]. Thereby, our nomogram with strong discriminative power could provide new insights on appropriate treatments and better clinical management for PCa patient cohorts.

Given considerable value to discriminate BCR, the potential predictive performance of the individual proteins included in our final five-protein-signature needed to be revealed. Our work revealed that three proteins (BRD4, YB1 and SMAD1) were risk factors for PCa patients. It is coincident with the study of Tan et al. [20] that report that expression levels of BRD4 are notably increased in malignant prostate specimens and related to clinical stage and metastasis. Further analysis demonstrates that BRD4 even mediates migration and invasion of castration-resistance prostate cancer (CRPC) through direct transcriptional regulation [21]. YB-1 has also been revealed to correlate with cancerous transformation and poor outcomes in CRPC. Around 66% of patients with high YB-1 expression were reported to relapse within 5 years of postoperative chemotherapy [22]. Potential molecular mechanism may be explained that upregulated YB-1 contributed to reduced intracellular androgen accumulation, thus weaning PCa off androgen dependency and upregulating tumor survival [23]. In accordance with the present study, SMAD1 is reported to be significantly elevated in patients with high-risk PCa [24]. The same conclusion is also drawn by increasing evidence which show the downregulation of SMAD1 contributed to the PCa proliferation, migration and invasion [25]. However, there is limited work regarding the association between SMAD1 and PCa, so that the underlying mechanism of these findings require further investigation.

In the present study, two (alpha-Catenin and DJ1) were protective factors for PCa progression. This finding reflects the study of Aaltomma et al.[26] who expounds that changes in the alpha-catenin-regulated cell-cell adhesion mechanism seems be present in almost half of all prostate tumors. Furthermore, alpha-catenin is a promising prognostic marker for PCa specific survival, lack of which may indicate PSA failure as well as shortened survival [26, 27]. It is well-demonstrated that lack of alpha-catenin expression leads to reduced cell-cell adhesion and loss of the epithelial phenotype, which could be reversed after repletion of alpha-catenin [28]. Although Xu et al. [29] indicates that PSA with DJ-1/PARK7 has novel potential to identify population without cancer from those with PCa, significance of DJ-

1/ PARK7 in PCa progression is not well understood. More research is needed to explore more about DJ-1 in PCa.

Our signature is derived from BCR-related proteins. Hence, our signature is closely related to BCR and will be appropriate for prognostic evaluation. Furthermore, above pathway analysis confirms once again that our 5-protein signature is closely associated with cancer metastasis and BCR. Interestingly, the Sankey diagram indicated the possibility of interaction between five proteins in our signature and co-expression ones. This finding is consistent with the study of Augustin et al.[30] who suggest that prostate tumor grade is associated with expressions of p53 and alpha-catenin. Mutations of the p53 gene and depletion of alpha-catenin are observed in pats of PCa [30]. In addition, Sankey diagram in the present study broadly supported the work of Gulino-Debrac et al.[31] who indicates that alpha-catenin and VEGFR2 seems to be related to the mechanotransduction mechanism for adherent junction strengthening endothelial cell-cell contacts. Though most co-expression proteins have not yet functionally annotated in PCa, undoubtedly, future studies should reveal potential molecular connections in the PCa progression.

Although the protein-base model showed an accurate survival prognosis, there are several limitations of our work. Firstly, only patients with complete BCR information were included in the present study, which may cause a selection bias. The size of samples and patients in our work was also limited. More large-scale trials are needed to investigate the prognostic value of this 5-proteins signature. Secondly, we only included data set from TCGA platform, which does not encompass all potential proteins. The potential molecular mechanisms of several proteins in our signature are not completely clear. Further larger researches are needed to explore the exact molecular mechanisms of these prognostic proteins.

Conclusion

To predict biochemical recurrence following radical prostatectomy in prostate cancer, we firstly identified and validated an innovative prognostic proteomic signature, which could identify PCa patients into high- and low-risk groups. Moreover, this 5-protein-based signature is independent of clinical factors and outperforms other known prognostic signatures, indicating the discriminative power and contribution to personalized management of PCa patients.

Abbreviations

BCR: biochemical recurrence; PCa: prostate cancer; RP: radical prostatectomy; CRPC: castration-resistance prostate cancer; ROC: receiver operating characteristic; C-index: concordance index; DCA: decision curve analysis; GO: Gene Ontology; GSEA: Gene Set Enrichment Analysis; RPPA: Reverse-phase protein array; TCPA: Cancer Proteome Atlas; TCGA: Cancer Genome Atlas; FPKM: Fragments Per Kilobase Of Exon Per Million Fragments Mapped; KM: Kaplan–Meier; LASSO: least absolute shrinkage and selection operator; AUC: area under curve; DCA: decision curve analysis; MSigDB: Molecular Signatures Database; HR: hazard ratios.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

All data generated or analyzed during this study are included in the main text and supplementary material.

Competing interests

The authors have no possible conflict of interest.

Funding

This work was partially supported by grants from the following: National Natural Science Foundation of China (No.81600542); Guangdong Basic and Applied Basic Research Foundation (Grant NO. 2019A1515110033); Distinguished Young Talents in Higher Education Foundation of Guangdong Province (Grant NO. 2019KQNCX115); China Postdoctoral Science Foundation (Grant NO. 2019M662865); Achievement cultivation and clinical transformation application cultivation projects of the First Affiliated Hospital of Guangzhou Medical University(Grant NO. ZH201908).

Authors' contributions

WJL, DJL and YDL reviewed relevant literatures and drafted the manuscript. XKW and DG conducted all statistical analyses. All authors read and approved the final manuscript.

Acknowledgements

All authors appreciate The First Affiliated Hospital of Guangzhou Medical University.

Ethical Statement

The authors are responsible for all aspects of the work and ensure that the accuracy and integrity of the work are appropriately investigated and resolved.

Data availability statement

The datasets generated or analyzed in our work study are available in the TCGA (https://portal.gdc.cancer.gov/) and TCPA (https://www.tcpaportal.org/).

References

1. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2019. CA Cancer J Clin 2019, 69(1):7-34.
2. Van den Broeck T, van den Bergh RCN, Arfi N, Gross T, Moris L, Briers E, Cumberbatch M, De Santis M, Tilki D, Fanti S 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.
3. Cornford P, Bellmunt J, Bolla M, Briers E, De Santis M, Gross T, Henry AM, Joniau S, Lam TB, Mason MD et al: EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part II: Treatment of Relapsing, Metastatic, and Castration-Resistant Prostate Cancer. Eur Urol 2017, 71(4):630-642.
4. Brockman JA, Alanee S, Vickers AJ, Scardino PT, Wood DP, Kibel AS, Lin DW, Bianco FJ, Jr., Rabah DM, Klein EA et al: Nomogram Predicting Prostate Cancer-specific Mortality for Men with Biochemical Recurrence After Radical Prostatectomy. Eur Urol 2015, 67(6):1160-1167.
5. Van den Broeck T, van den Bergh RCN, Briers E, Cornford P, Cumberbatch M, Tilki D, De Santis M, Fanti S, Fossati N, Gillessen S et al: Biochemical Recurrence in Prostate Cancer: The European Association of Urology Prostate Cancer Guidelines Panel Recommendations. Eur Urol Focus 2020, 6(2):231-234.
6. Shao N, Tang H, Qu Y, Wan F, Ye D: Development and validation of lncRNAs-based nomogram for prediction of biochemical recurrence in prostate cancer by bioinformatics analysis. J Cancer 2019, 10(13):2927-2934.
7. Remmers S, Verbeek JFM, Nieboer D, van der Kwast T, Roobol MJ: Predicting biochemical recurrence and prostate cancer-specific mortality after radical prostatectomy: comparison of six prediction models in a cohort of patients with screening- and clinically detected prostate cancer. BJU Int 2019.
8. Goodman M, Ward KC, Osunkoya AO, Datta MW, Luthringer D, Young AN, Marks K, Cohen V, Kennedy JC, Haber MJ et al: Frequency and determinants of disagreement and error in gleason scores: a population-based study of prostate cancer. Prostate 2012, 72(13):1389-1398.
9. Strand SH, Orntoft TF, Sorensen KD: Prognostic DNA methylation markers for prostate cancer. Int J Mol Sci 2014, 15(9):16544-16576.
10. Karatas OF, Guzel E, Suer I, Ekici ID, Caskurlu T, Creighton CJ, Ittmann M, Ozen M: miR-1 and miR-133b are differentially expressed in patients with recurrent prostate cancer. PLoS One 2014, 9(6):e98675.
11. Guo Y, Xiao P, Lei S, Deng F, Xiao GG, Liu Y, Chen X, Li L, Wu S, Chen Y et al: How is mRNA expression predictive for protein expression? A correlation study on human circulating monocytes. Acta Biochim Biophys Sin (Shanghai) 2008, 40(5):426-436.
12. Jia Z, Zhu J, Zhuo Y, Li R, Qu H, Wang S, Wang M, Lu J, Chater JM, Ma R et al: Offsetting Expression Profiles of Prognostic Markers in Prostate Tumor vs. Its Microenvironment. Front Oncol 2019, 9:539.
13. Li J, Lu Y, Akbani R, Ju Z, Roebuck PL, Liu W, Yang JY, Broom BM, Verhaak RG, Kane DW et al: TCPA: a resource for cancer functional proteomics data. Nat Methods 2013, 10(11):1046-1047.

14. Tang J, Ren J, Cui Q, Zhang D, Kong D, Liao X, Lu M, Gong Y, Wu G: A prognostic 10-lncRNA expression signature for predicting the risk of tumour recurrence in breast cancer patients. J Cell Mol Med 2019, 23(10):6775-6784.
15. Vickers AJ, Cronin AM, Elkin EB, Gonen M: Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak 2008, 8:53.
16. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005, 102(43):15545-15550.
17. Shtivelman E, Beer TM, Evans CP: Molecular pathways and targets in prostate cancer. Oncotarget 2014, 5(17):7217-7259.
18. de Sousa Abreu R, Penalva LO, Marcotte EM, Vogel C: Global signatures of protein and mRNA expression levels. Mol Biosyst 2009, 5(12):1512-1526.
19. Gerke TA, Martin NE, Ding Z, Nuttall EJ, Stack EC, Giovannucci E, Lis RT, Stampfer MJ, Kantoff PW, Parmigiani G et al: Evaluating a 4-marker signature of aggressive prostate cancer using time-dependent AUC. Prostate 2015, 75(16):1926-1933.
20. Tan Y, Wang L, Du Y, Liu X, Chen Z, Weng X, Guo J, Chen H, Wang M, Wang X: Inhibition of BRD4 suppresses tumor growth in prostate cancer via the enhancement of FOXO1 expression. Int J Oncol 2018, 53(6):2503-2517.
21. Shafran JS, Andrieu GP, Gyorffy B, Denis GV: BRD4 Regulates Metastatic Potential of Castration-Resistant Prostate Cancer through AHNAK. Mol Cancer Res 2019, 17(8):1627-1638.
22. Janz M, Harbeck N, Dettmar P, Berger U, Schmidt A, Jurchott K, Schmitt M, Royer HD: Y-box factor YB-1 predicts drug resistance and patient outcome in breast cancer independent of clinically relevant tumor biologic factors HER2, uPA and PAI-1. Int J Cancer 2002, 97(3):278-282.
23. Gimenez-Bonafe P, Fedoruk MN, Whitmore TG, Akbari M, Ralph JL, Ettinger S, Gleave ME, Nelson CC: YB-1 is upregulated during prostate cancer tumor progression and increases P-glycoprotein activity. Prostate 2004, 59(3):337-349.
24. Wahdan-Alaswad RS, Bane KL, Song K, Shola DT, Garcia JA, Danielpour D: Inhibition of mTORC1 kinase activates Smads 1 and 5 but not Smad8 in human prostate cancer cells, mediating cytostatic response to rapamycin. Mol Cancer Res 2012, 10(6):821-833.
25. Chen QG, Zhou W, Han T, Du SQ, Li ZH, Zhang Z, Shan GY, Kong CZ: MiR-345 suppresses proliferation, migration and invasion by targeting Smad1 in human prostate cancer. J Cancer Res Clin Oncol 2016, 142(1):213-224.
26. Aaltomaa S, Karja V, Lipponen P, Isotalo T, Kankkunen JP, Talja M, Mokka R: Reduced alpha- and beta-catenin expression predicts shortened survival in local prostate cancer. Anticancer Res 2005, 25(6C):4707-4712.
27. van Oort IM, Tomita K, van Bokhoven A, Bussemakers MJ, Kiemeney LA, Karthaus HF, Witjes JA, Schalken JA: The prognostic value of E-cadherin and the cadherin-associated molecules alpha-,

beta-, gamma-catenin and p120ctn in prostate cancer specific survival: a long-term follow-up study. Prostate 2007, 67(13):1432-1438.

28. Vasioukhin V, Bauer C, Degenstein L, Wise B, Fuchs E: Hyperproliferation and defects in epithelial polarity upon conditional ablation of alpha-catenin in skin. Cell 2001, 104(4):605-617.
29. Xu L, Lee JR, Hao S, Ling XB, Brooks JD, Wang SX, Gambhir SS: Improved detection of prostate cancer using a magneto-nanosensor assay for serum circulating autoantibodies. PLoS One 2019, 14(8):e0221051.
30. Augustin H, Hammerer PG, Graefen M, Palisaar J, Daghofer F, Huland H, Erbersdobler A: Characterisation of biomolecular profiles in primary high-grade prostate cancer treated by radical prostatectomy. J Cancer Res Clin Oncol 2003, 129(11):662-668.
31. Gulino-Debrac D: Mechanotransduction at the basis of endothelial barrier function. Tissue Barriers 2013, 1(2):e24180.

Tables

Table 1. Univariate and multivariate Cox proportional hazards regression analyses

Variable	Univariate analysis			Multivariate analysis		
	HR	95%CI of HR	P value	HR	95%CI of HR	P value
Training dataset (n=169)						
Age						
<65					1.00	
>65				1.01	0.95 to 1.07	0.87
Gleason scores						
<=6		1.00			1.00	
7	0.52	0.06 to 4.84	0.56	0.96	0.11 to 8.18	0.97
8,9,10	1.81	0.20 to 16.21	0.60	6.70	0.90 to 49.62	0.06
T stage						
T1&2		1.00			1.00	
T3&4	2.71	0.76 to 9.7	0.13	4.64	1.41 to 15.30	0.01 *
N status						
N0		1.00			1.00	
N1	1.28	0.55 to 2.97	0.56	2.68	1.20 to 6.03	0.02 *
PSA						
<10 ng/mL					1.00	
10 to 20 ng/mL				1.52	0.65 to 3.56	0.33
>20 ng/mL				1.28	0.50 to 3.23	0.60
Residual tumor						
No		1.00			1.00	
Yes	1.16	0.54 to 2.50	0.71	2.14	1.04 to 4.42	0.04 *
Risk score						
low		1.00			1.00	
high	3.65	1.27 to 10.50	0.02 *	6.33	2.42 to 16.58	0.0002 ***
Test dataset (n=172)						
Age						
<65					1.00	
>65				1.06	1.00 to 1.12	0.04 *
Gleason scores						
<=6		1.00			1.00	
7	4.27		1.00			
8,9,10	5.13		1.00			
T stage						
T1&2		1.00			1.00	
T3&4	2.71	0.69 to 10.73	0.15	5.44	1.58 to 18.67	0.007 **
N status						
N0		1.00			1.00	
N1	1.11	0.46 to 2.67	0.82	2.24	1.03 to 4.85	0.04 *
PSA						
<10 ng/mL					1.00	
10 to 20 ng/mL				1.09	0.43 to 2.76	0.86
>20 ng/mL				1.23	0.46 to 3.31	0.69
Residual tumor						

No		1.00			1.00	
Yes	1.28	0.57 to 2.87	0.55	2.33	1.10 to 4.91	0.03 *
Risk score						
low		1.00			1.00	
high	4.53	1.89 to 10.87	0.0007 ***	5.70	2.42 to 13.43	<0.0001 ***
Entire dataset (n=341)						
Age						
<65					1.00	
>65				1.03	0.99 to 1.08	0.11
Gleason scores						
<=6		1.00				
7	1.76	0.23 to 13.56	0.59	2.76	0.37 to 20.82	0.33
8,9,10	3.40	0.44 to 26.47	0.24	10.57	1.45 to 76.81	0.02 *
T stage						
T1&2		1.00			1.00	
T3&4	2.52	1.01 to 6.31	0.048 *	4.89	2.07 to 11.53	0.0003 ***
N status						
N0		1.00			1.00	
N1	1.11	0.61 to 2.00	0.74	2.36	1.35 to 4.11	0.003 **
PSA						
<10 ng/mL					1.00	
10 to 20 ng/mL				1.32	0.70 to 2.45	0.39
>20 ng/mL				1.26	0.64 to 2.49	0.50
Residual tumor						
No		1.00			1.00	
Yes	1.22	0.70 to 2.11	0.49	2.16	1.29 to 3.62	0.004 **
Risk score						
low		1.00			1.00	
high	4.22	2.18 to 8.14	<0.0001 ***	5.98	3.17 to 11.31	<0.0001 ***

Supplemental Figure Legends

Supplementary figure S1: Kaplan–Meier survival curve analysis of the 21 prognostic proteins in the signature.

Supplementary figure S2: Kaplan–Meier survival analysis of the five proteins risk score in diverse subgroups. (A) age <65/>65 (B) Gleason score 6/7, Gleason score 8/9/10, (C) N0 status, N1 status, (D) T1-T2 stage, T3-T4 stage, (E) PSA 0-10, PSA>10, (F) R0 non-residual-cancer, R1 residual-cancer.

Supplementary figure S3: Pearson correlation coefficient between levels of 5 proteins and their corresponding genes. (A) BRD4-BRD4, (B) CTNNA1-ALPHACATENIN, (C) PARK7-DJ1, (D) SMAD1-SMAD1, and (E) YB1-YBX1.

Figures

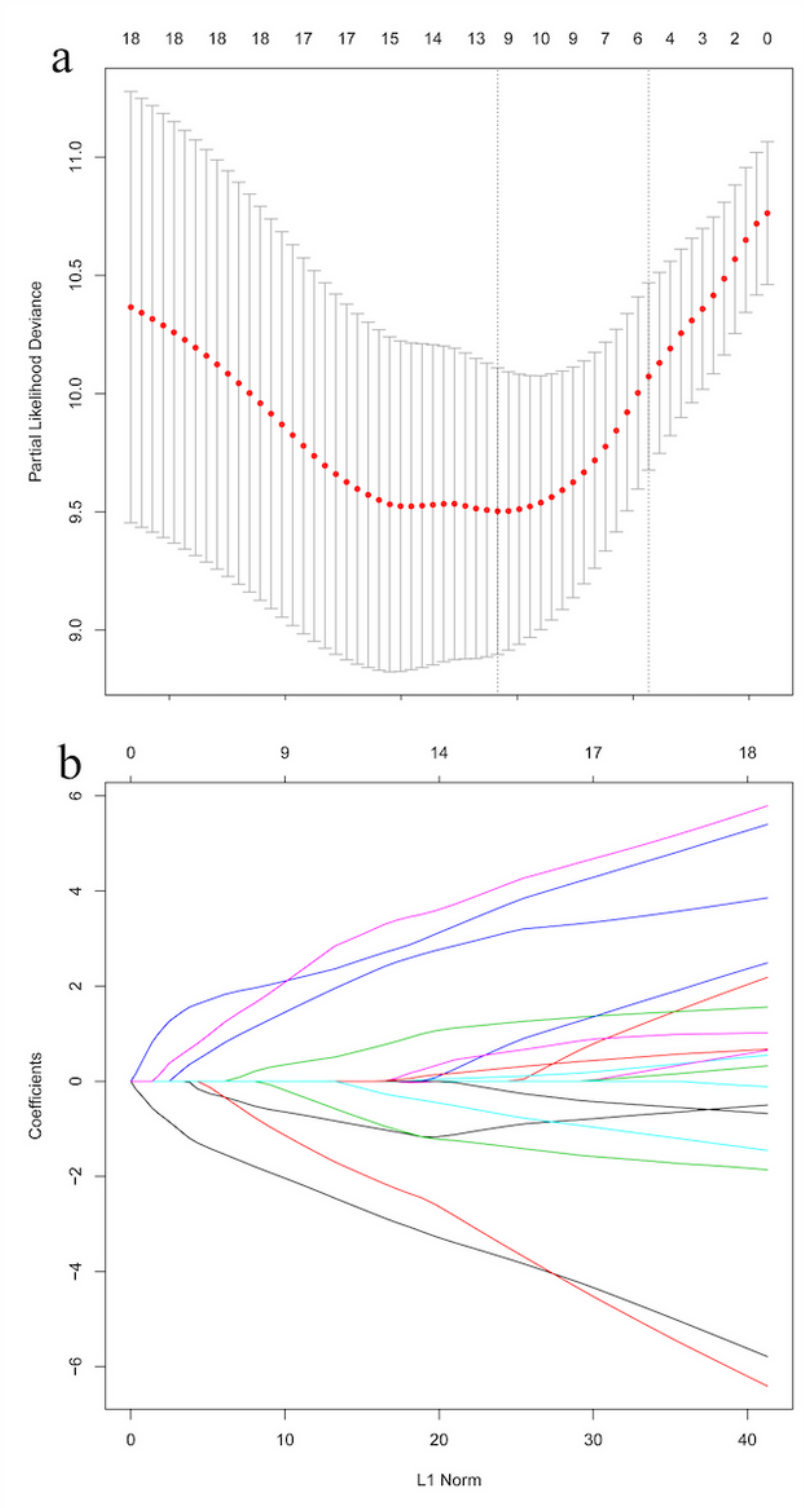

Figure 1

LASSO Cox regression analysis. (A) Ten-time cross-validation for tuning parameter selection in the lasso model. (B) LASSO coefficient profiles of the 21 predictive proteins associated with BCR. A vertical line is drawn at the value selected by 10-fold cross-validation.

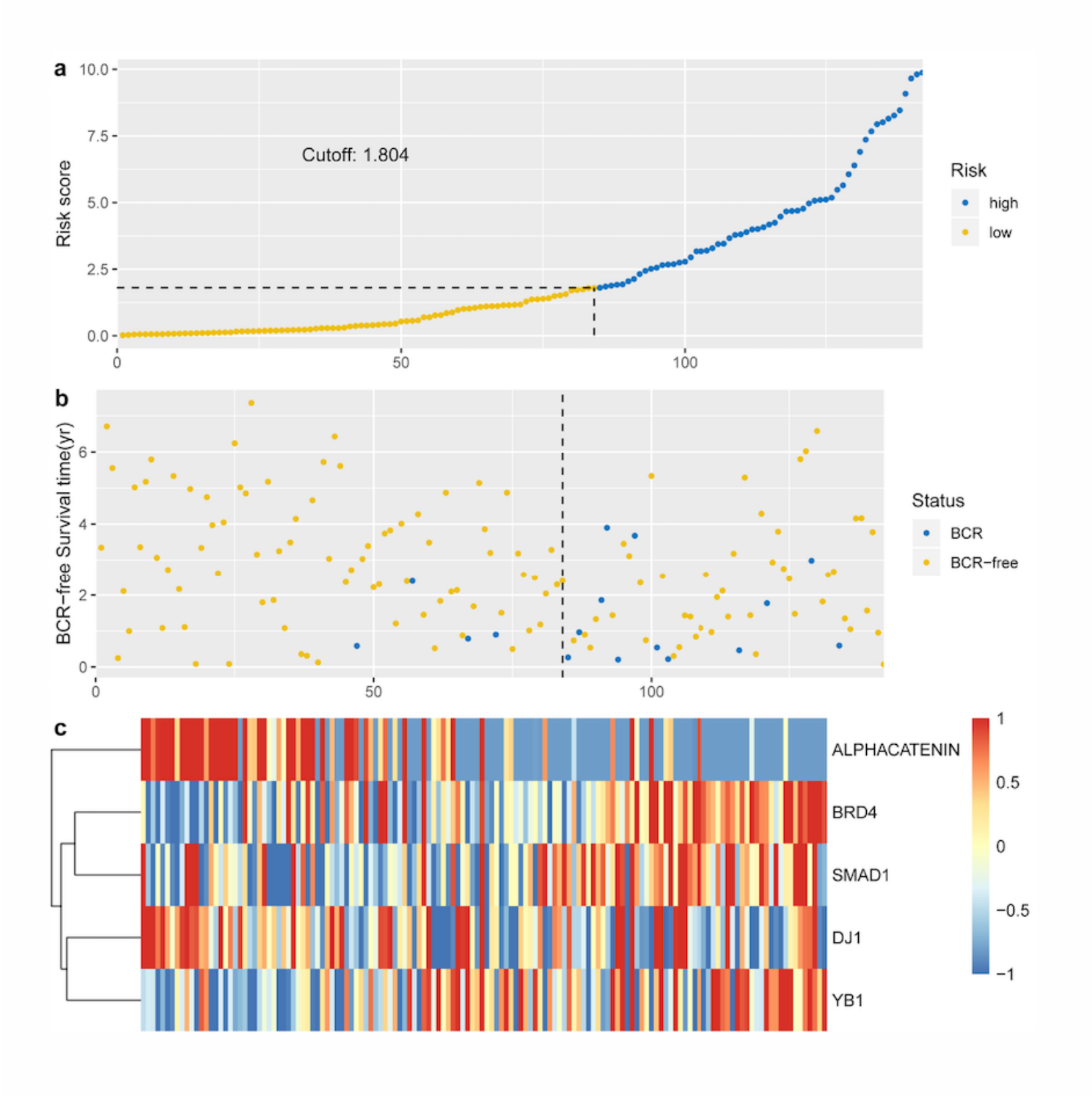

Figure 2

Identification of the integrated prognostic classifier in the training set. (a) The distribution of risk score. The median risk score cut-off is 1.804. (b) Each point in the scatterplot represents the survival status of patients. (c) Heat map showed differentially expressed proteins between BCR-free patients and patients with BCR.

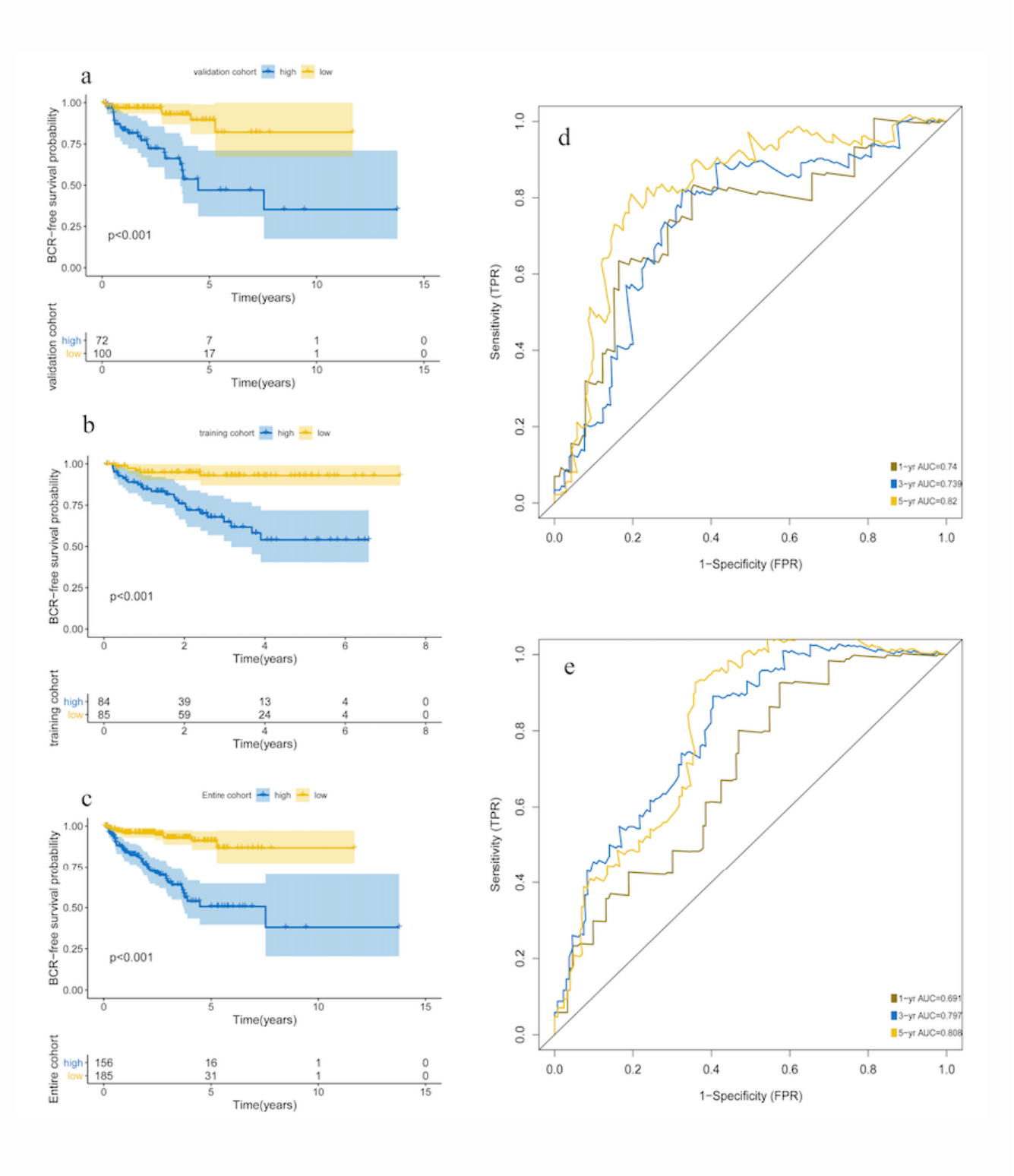

Figure 3

Kaplan-Meier survival analysis and ROC curves according to the prognostic signature. Kaplan–Meier curves for (A) the training group (N=172); (B) the test group (N=169); (C) the entire group (N=341). ROC curves at 1, 3, 5 years for (D) training group, (E) test group.

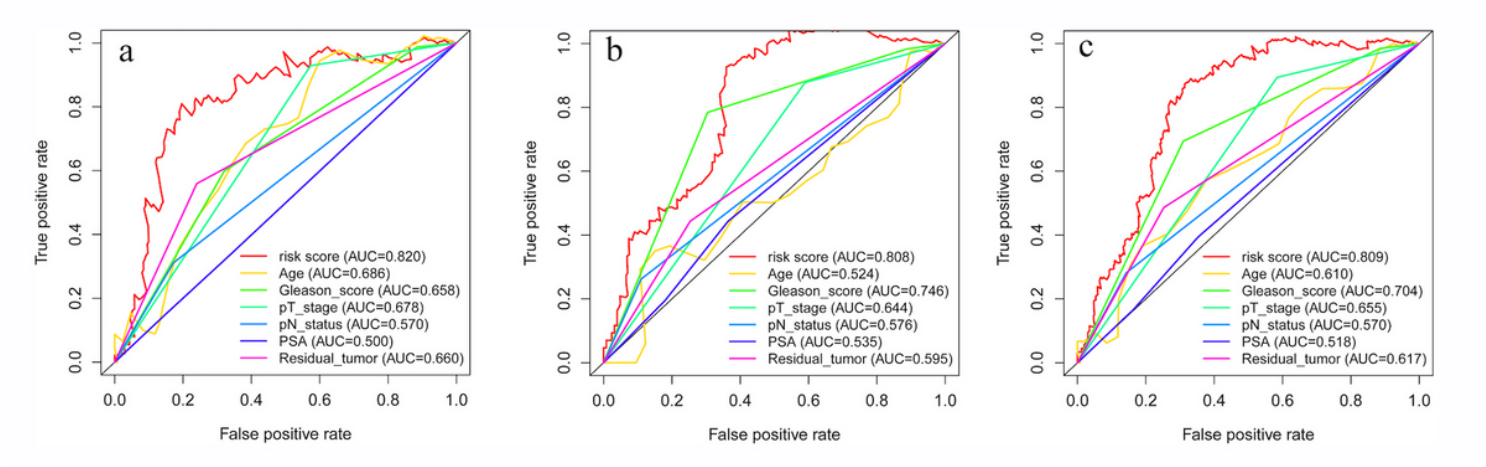

Figure 4

ROC curves compare the prognostic power between the prognostic signature and clinicopathological features. (A) Training set. (B) Test set. (C) Entire set. P values indicate the area under curve (AUC) at 5 years for multi-protein-based signature verse the AUC at 5 years for other features.

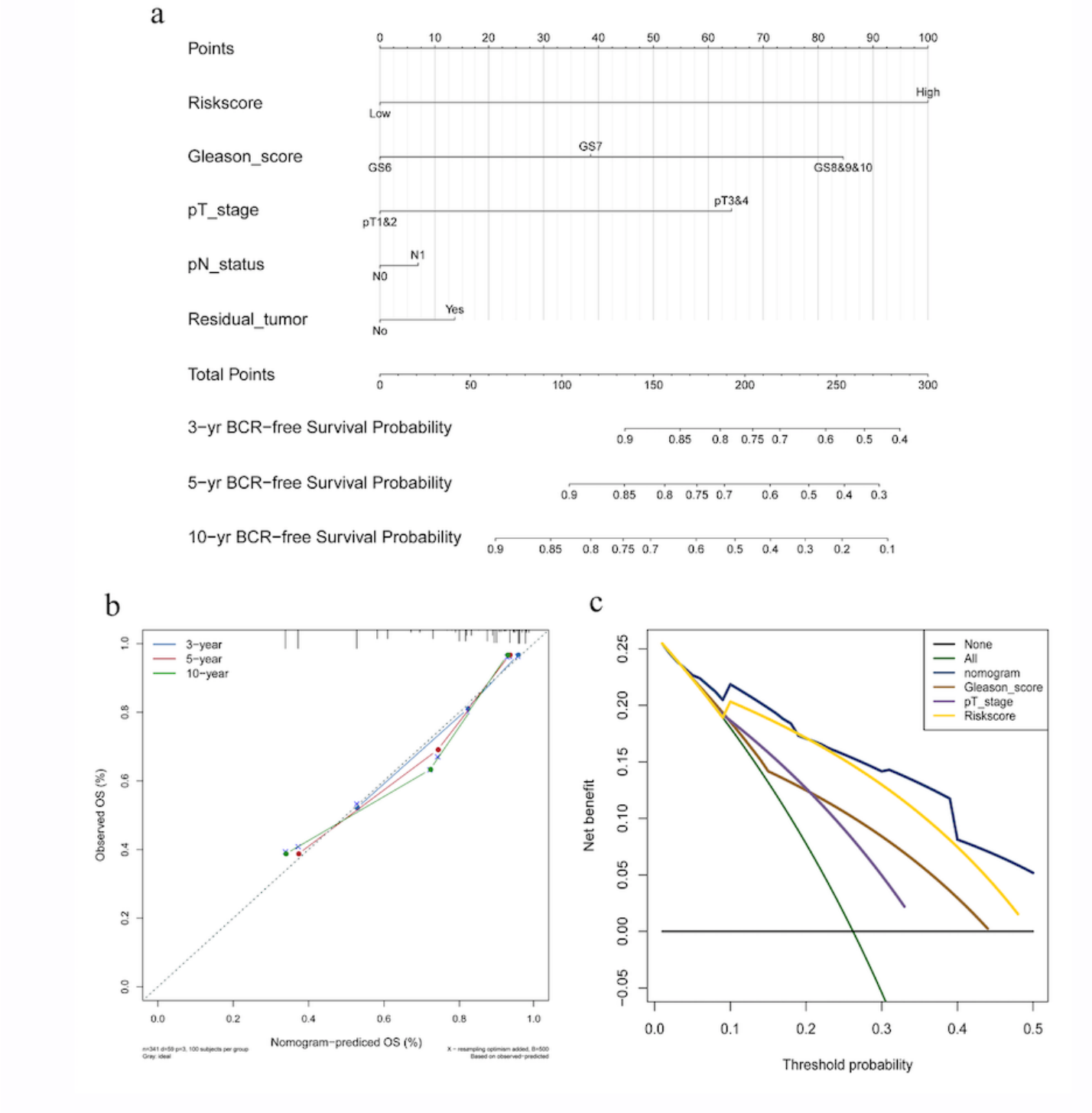

Figure 5

PCa survival nomogram, calibration curve and Net benefit curves. (A) Survival nomogram. each variable axis represents the value of individual patient, and a line is plotted upward to decide the number of points received for each variable value. The Total Points axis represents the total of these numbers, and a line is plotted downward to the survival axes to deicide the probability of 3-, 5- and 10-year BCR-free survival. (B) The calibration curve for predicting PCa patient survival at 3, 5 and 10 years in the entire cohort. (C) Net

benefit curves for the nomogram, signature and other clinical features, along with their confidence intervals. P values indicate the area under curve (AUC) at 5 years for nomogram verse the AUC at 5 years for other features.

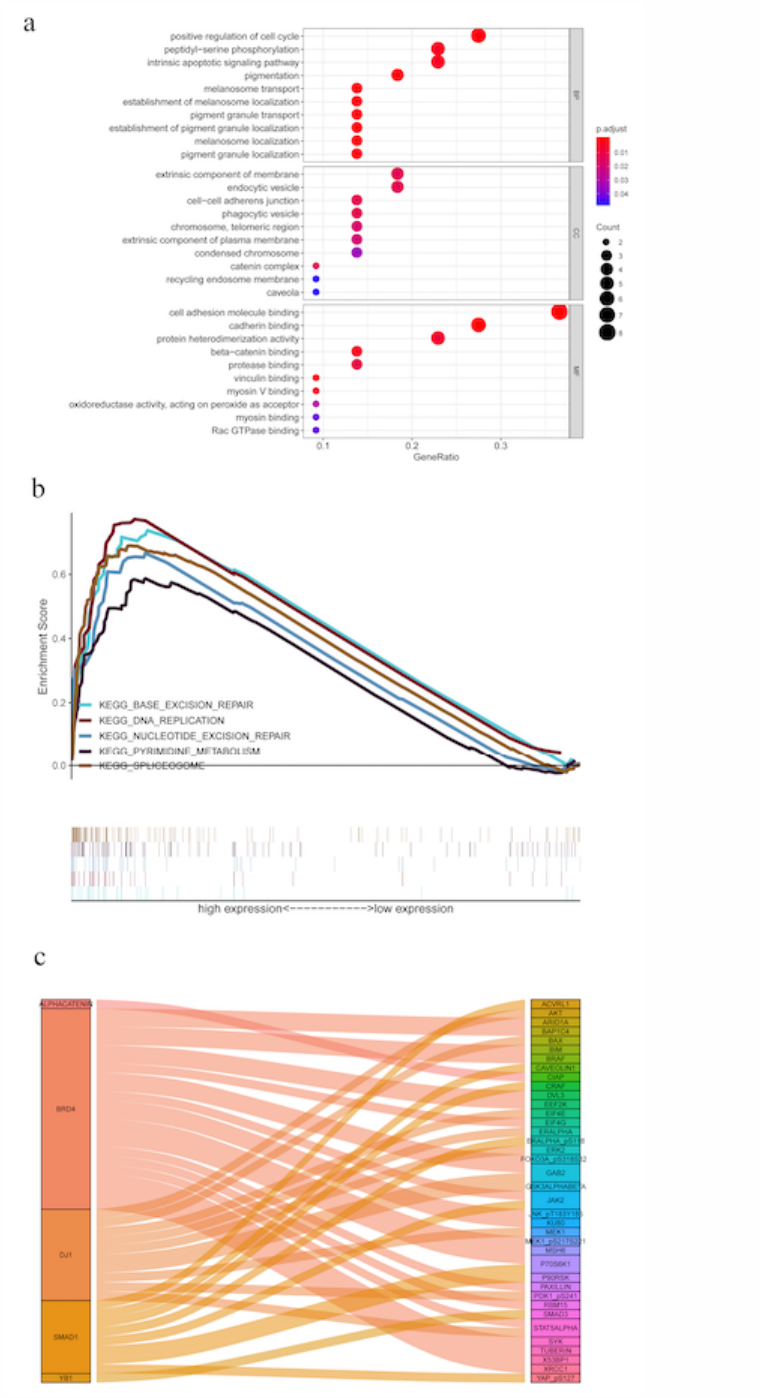

Figure 6

Functional Characteristics of 5 proteins. (A) GO enrichment analysis, (B) Gene set enrichment analysis, (C) Sankey diagram of the signature.

Supplementary Files

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

- SupplementaryfigureS1.tif
- SupplementaryfigureS3.tif
- SupplementaryfigureS2.tif