

PRIMARY RESEARCH

Open Access



# A Gleason score-related outcome model for human prostate cancer: a comprehensive study based on weighted gene co-expression network analysis

Yongzhi Wang and Zhonghua Yang\*

## Abstract

**Background:** Prostate cancer (PCa) is the second leading cause of cancer death in men in 2018. Thus, the evaluation of prognosis is crucial for clinical treatment decision of human PCa patients. We aim to establishing an effective and reliable model to predict the outcome of PCa patients.

**Methods:** We first identified differentially expressed genes between prostate cancer and normal prostate in TCGA-PRAD and then performed WGCNA to initially identify the candidate Gleason score related genes. Then, the candidate genes were applied to construct a LASSO Cox regression analysis model. Numerous independent validation cohorts, time-dependent receiver operating characteristic (ROC), univariate cox regression analysis, nomogram were used to test the effectiveness, accuracy and clinical utility of the prognostic model. Furthermore, functional analysis and immune cells infiltration were performed.

**Results:** Gleason score-related differentially expressed candidates were identified and used to build up the outcome model in TCGA-PRAD cohort and was validated in MSKCC cohort. We found the 3-gene outcome model (CDC45, ESPL1 and RAD54L) had good performance in predicting recurrence free survival, metastasis free survival and overall survival of PCa patients. Time-dependent ROC and nomogram indicated an ideal predictive accuracy and clinical utility of the outcome model. Moreover, outcome model was enriched in 28 pathways by GSVA and GSEA. In addition, the risk score was positively correlated with memory B cells, native CD4 T cells, activated CD4 memory T cells and eosinophil, and negatively correlated with plasma cells, resting CD4 memory T cells, resting mast cells and neutrophil.

**Conclusions:** In summary, our outcome model proves to be an effective prognostic model for predicting the risk of prognosis in PCa.

**Keywords:** Prostate cancer, WGCNA, LASSO, Prognosis, TCGA, GEO

## Background

Prostate cancer (PCa) is the second leading cause of cancer death in men in 2018 [1]. Due to metastasis, the 5-year relative survival rate of distant PCa is only 30% [2].

Despite decades of efforts in research, the standard treatment options and guidelines for PCa patients diagnosed with metastatic progression have remained unchanged [3, 4]. Clinically, the Gleason scoring system has been widely used for assessment of prognosis of PCa based on histology [5]. However, in patients with metastatic PCa, metastatic biopsies are rarely performed [6]. Although

\*Correspondence: yangzhonghua@whu.edu.cn  
Department of Urology, Zhongnan Hospital of Wuhan University,  
Wuhan 430071, China



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

prostate specific antigen (PSA) screening contributes to decrease PCa metastases and mortality [7], over-diagnosis and over-treatment become a controversial issue [8, 9].

With the development of precision medicine (PM), individualized treatment based on cancer genomic data has been achievable [10]. In view of the multiple treatment options and inconsistent outcomes of PCa, reliable biomarkers might help to optimize clinical decisions [11, 12]. For instance, as for diagnosis, the combination of digital-rectal examination and PSA value provides the risk stratification in most patients but could not give more details for the following steps. Therefore, prostate health indexes (PHI), four-kallikrein panel (4K), and even combination of PHI and 4K have been introduced [13–15]. The prognosis of PCa is greatly variable because of its various features [16]. However, there is little research about the biomarkers of PCa prognosis. PTEN, as a suppressor of prostate cancer, might be a prognostic biomarker for PCa [17]. Several miRNA, such as miR-145, has been demonstrated to suppress the androgen receptor in PCa cells and correlate to PCa prognosis [18]. However, there is still a huge distance between clinical application and these biomarkers because of lacking enough validation. Therefore, establishing an effective prognostic model for PCa has significant clinical implications.

The recent development of high-throughput profiling, sequencing technology and bioinformatics has revolutionized cancer research in general. By utilizing the publicly available datasets and advanced bioinformatics tools, we aim to systematically explore biomarkers for PCa prognosis. The weighted gene co-expression network analysis (WGCNA) is an R package for weighted correlation network analysis and can be used as a data exploratory tool or a gene screening (ranking) method to find clusters (modules) of highly correlated genes [19]. It has been widely used to find hub genes in various cancers [20–22]. The least absolute shrinkage and selection operator (LASSO) method was originally designed for regression analysis. Lately, it has been applied to many fields, including the construction of prognostic models for various cancers [23–26].

In this study, we firstly used WGCNA to identify hub genes associated with Gleason score. By using LASSO method, we then constructed a 3-mRNA signature and extensively tested for predicting patient disease free survival. Meanwhile, the outcome model was validated in various independent datasets. Next, we built a nomogram based on the 3-mRNA signature combined with other clinical factors. We have confirmed the clinical utility by decision curve analysis. Furthermore, to explore tumor associated immune cell infiltration and risk score

derived from outcome model, we used CIBERSORT to calculate the immune cell infiltration abundance and performed combination survival analysis.

## Materials and methods

### Clinical samples and data acquisition

Gene expression profiles of prostate cancer and corresponding clinical information of patients were obtained from UCSC Xena database (<http://xena.ucsc.edu/>), Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) and cbiportal database (<http://www.cbiportal.org/>). Detailed information for the 8 independent datasets including MSKCC, GSE116918, GSE46602, GSE54460, GSE70768, GSE70769, GSE16560 and GSE53922 were listed in Additional file 1: Table S1.

### Differentially expressed genes (DEGs) screening

Based on the pan-cancer normalized TCGA-PRAD data (498 prostate cancer samples and 52 non-tumor samples) in UCSC Xena database, we applied “limma” package in R to screen differentially expressed genes. Here, we set  $|\log_2(\text{Fold change})| > 1$  and false discovery rate (FDR)  $< 0.05$  as the cutoff.

### Weighted gene co-expression network construction and progression related genes identification

For WGCNA analysis, we used the gene expression matrix after the DEGs screening. Using “WGCNA” R package, we first deleted the outliers in each dataset [19]. Then, proper soft-thresholding parameter  $\beta$  was chosen to ensure a scale-free network, and genes with similar expression pattern were clustered into the same module. We then combined the modules with different clinical features (overall survival (OS) time, overall survival status, disease free survival (DFS) time, disease free survival status, age, biochemical recurrence, clinical M stage, clinical T stage, total Gleason score, laterality, number of positive lymphnode, pathological T stage, pathological N stage, PSA value, radiation therapy and targeted molecular therapy), we could finally identify the key genes related to total Gleason score.

### Establishment of outcome signature with LASSO regression model

LASSO (least absolute shrinkage and selection operator) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model. Here, using “glmnet” package in R, we applied the LASSO Cox regression analysis to build an optimal prognostic signature for PCa by using key genes related to total Gleason score [27]. Due to few dead patients in the overall survival cohort of TCGA-PRAD, we therefore

chose the disease free survival of TCGA-PRAD ( $n=436$ ) as the training set to set up the outcome model. Firstly, we excluded the patients without complete disease free survival information, then, optimal values of the penalty parameter lambda were determined through 10-times cross-validations. The minimum mean cross-validated error of the best lambda value was screened out. The risk score of prognostic signature for each sample was calculated by the relative expression of each prognostic gene in the signature and its associated coefficient. The risk score of the outcome signature =  $\sum_{i=1}^n (\text{coef}_i \times \text{Expr}_i)$ , where  $\text{Expr}_i$  is the relative expression of the gene in the signature for patient  $i$ ,  $\text{coef}_i$  is the LASSO coefficient of the gene  $i$ .

#### Estimation of outcome signature for patients' prognosis

According to the optimal  $p$  value, patients from different datasets were divided into low risk group and high risk group separately. Then we used "survivalROC" R package to perform time-dependent receiver operating characteristic curve (ROC) analysis and calculate the area under curve (AUC) for 1-year, 3-year and 5-year disease free survival, recurrence free survival (RFS), overall survival and metastasis free survival (MFS) for further evaluating the prediction accuracy for our model [28].

#### Gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA)

According to the optimal separation, we divided the TCGA-PRAD samples into high risk and low risk groups. Here, to further identify the role of our model in tumor metastasis, we used "clusterProfiler" and "fgsea" packages in R to visualize the correlation.  $P$  value and  $\text{padj}$  both  $< 0.05$  were set as the cutoff. For the metastasis phenotype identification, we chose prostate cancer metastasis related gene sets as the reference; for the functional analysis, we chose Hallmark gene sets as the reference. For the GSVA analysis, we used "GSVA" R package and set  $t$  value  $> 2$  and  $\text{FDR} < 0.05$  as the cutoff [29]. Then the common gene sets were identified via GSEA and GSVA analysis.

#### Construction and assessment of the nomogram

Excluding all missing information would lead to not enough patient samples. Therefore, we only firstly perform univariate cox regression analysis to identify the proper terms to build the nomogram. The forest was used to show the  $p$  value, HR and 95% CI of each variable through "forestplot" package in R. The nomogram, calibration plots and decision curve were generated using "rms" package. Afterwards, the calibration curves and decision curve analysis (DCA) were united to see

whether our established nomogram was suitable for clinical utility.

#### Evaluation of infiltrating immune cells and immune checkpoint with risk score

To evaluate the infiltration levels of immune cells in the PCa samples, we used the CIBERSORT algorithm [30], which provided an estimation of the abundances of member cell types by using gene expression data. B cells naïve, B cells memory, Plasma cells, T cells CD8, T cells CD4 naïve, T cells CD4 memory resting, T cells CD4 memory activated, T cells follicular helper, T cells regulatory (Tregs), T cells gamma delta, NK cells resting, NK cells activated, Monocytes, Macrophages M0, Macrophages M1, Macrophages M2, Dendritic cells resting, Dendritic cells activated, Mast cells resting, Mast cells activated, Eosinophils and Neutrophils were investigated. Furthermore, the correlation between risk score and immune cell infiltration was calculated and combination survival analyses were performed based on risk score and significantly-correlated infiltrating immune cells. Besides, the correlation between risk score and immune checkpoints were also investigated.

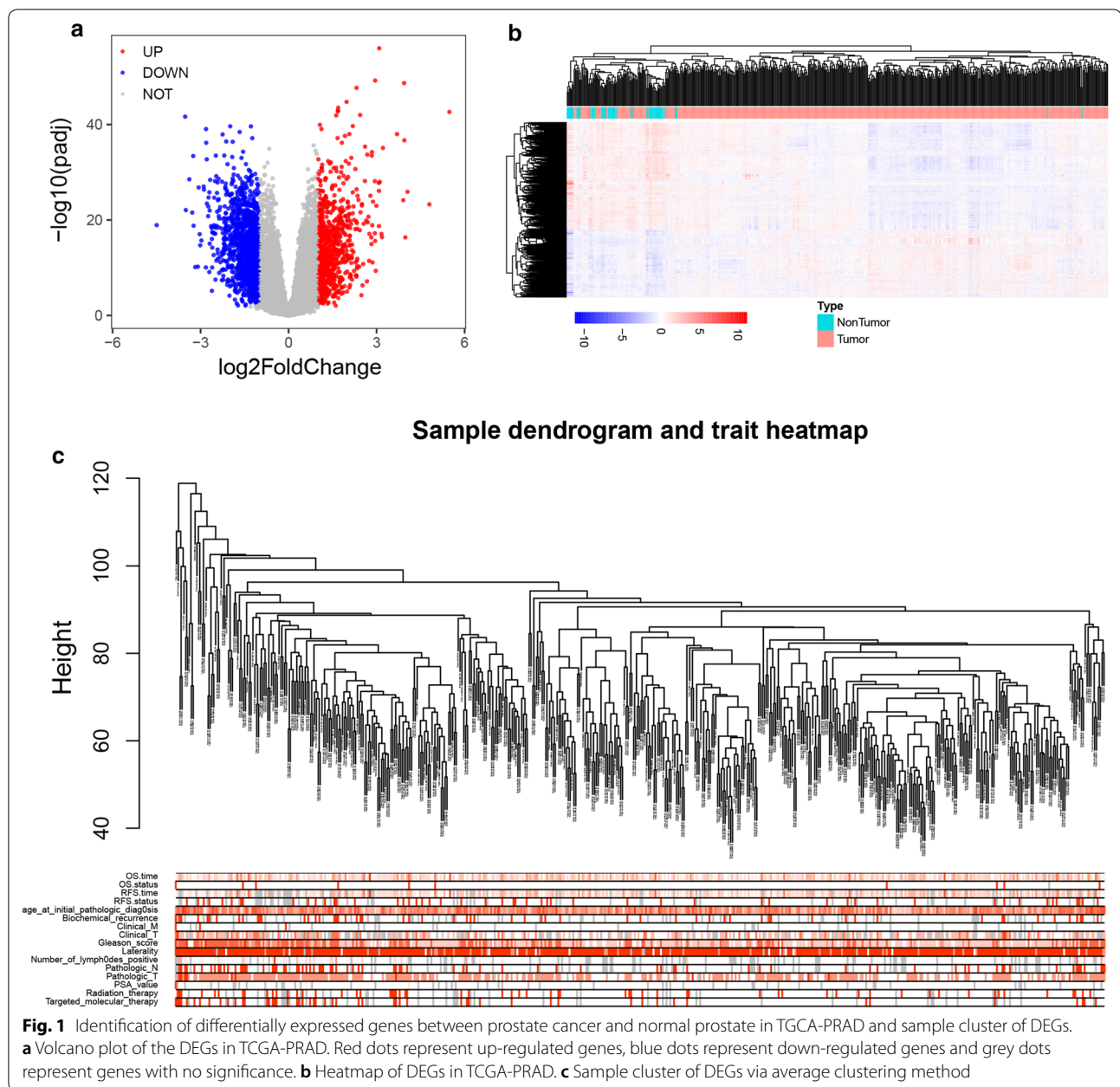
#### Statistical analysis

R software 3.5.0 was used for all statistical analyses. Package details were listed in Additional file 2: Table S2. Statistical significance was set at probability values of  $p < 0.05$ . Two-tailed Student's  $t$ -test was used for significance of differences between subgroups. One-way Anova test or Student  $t$  test were applied to analyze the correlation between risk score and clinicopathological parameters. Kaplan-Meier survival curves were built to analyze survival differences between the high risk group and low risk group. The ROC, calibration curve and DCA were compared for the predictive accuracy of the prognostic models.

## Results

#### Differentially expressed genes (DEGs) screening and selection of candidate genes via WGCNA analysis

We identified 2245 DEGs (858 up-regulated and 1387 down-regulated genes) (Additional file 3: Table S3). Differentially expressed genes were visualized via volcano plot and heatmap (Fig. 1a, b) and further, we used DEGs to construct co-expression network. Here, based on the average clustering, we detected the outlier samples and no one was removed (Fig. 1c). Then, we chose  $\beta=4$  as the proper soft-thresholding parameter and built a scale-free network (Additional file 4: Figure S1). Combined with the module and clinical information,



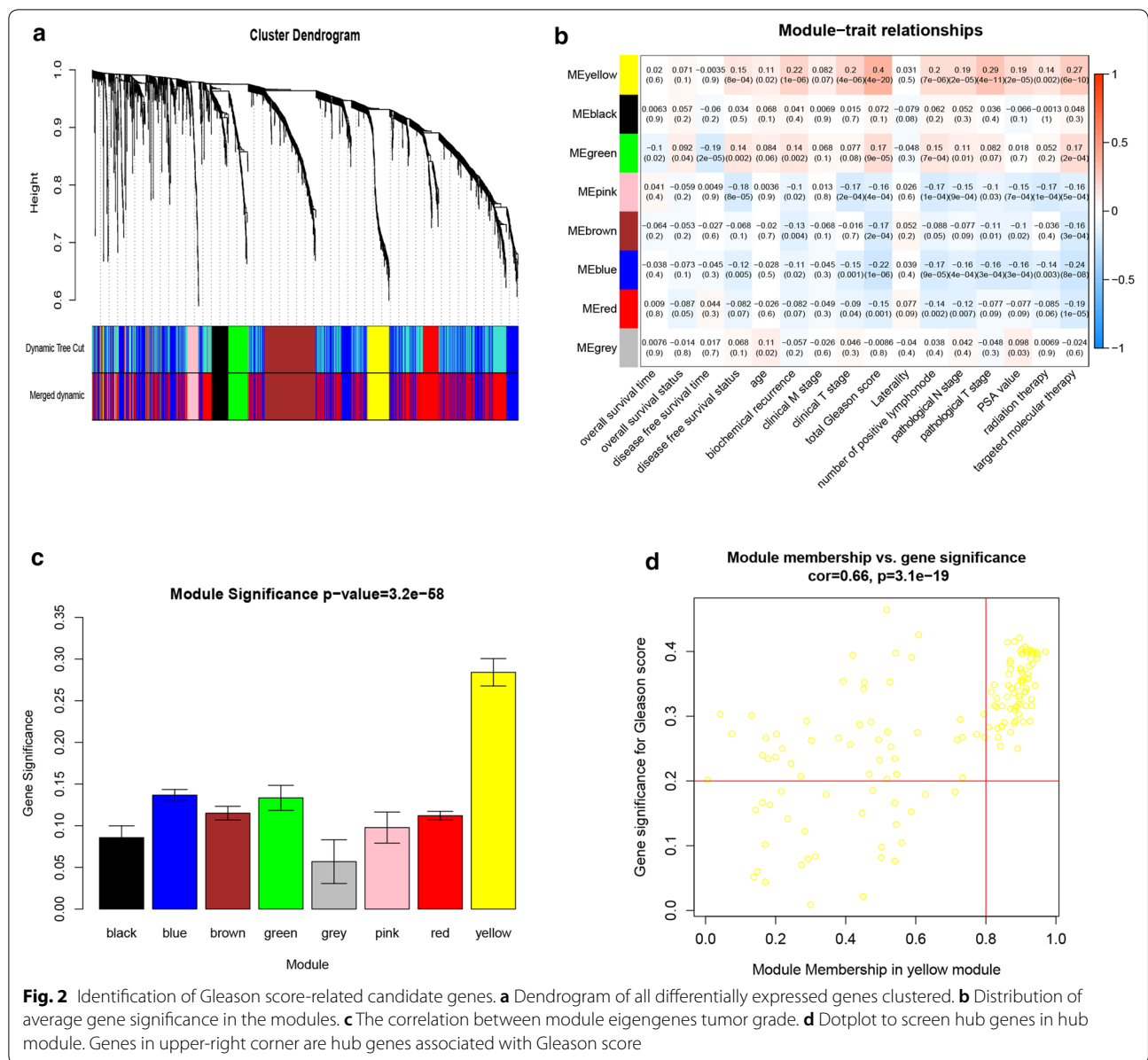
we eventually identified the yellow module which was highly correlated with total Gleason score ( $R=0.4$ ,  $p=4E-20$ ) and we subsequently selected the 73 hub genes in yellow module (Fig. 2 and Additional file 5: Table S4).

#### Construction of outcome model

Excluding the samples with incomplete disease free survival data, we finally used 436 TCGA-PRAD patients as the training set. Using LASSO cox regression, we

then generated a formula based on the expression of the three genes to predict DFS in TCGA-PRAD training dataset, and the formula for the risk score in each sample was calculated as follows:  $\text{risk score} = 1.562E-01 * \text{Exp}_{\text{CDC45}} + 2.850E-02 * \text{Exp}_{\text{ESPL1}} + 1.148E-04 * \text{Exp}_{\text{RAD54L}}$  (Additional file 6: Figure S2). Meanwhile, expression levels of these 3 genes were shown in Additional file 7: Figure S3. According to the risk score, patients were stratified into low-risk and high-risk groups at the best separation cut-off. KM curves indicated that high-risk groups were significantly associated with poorer



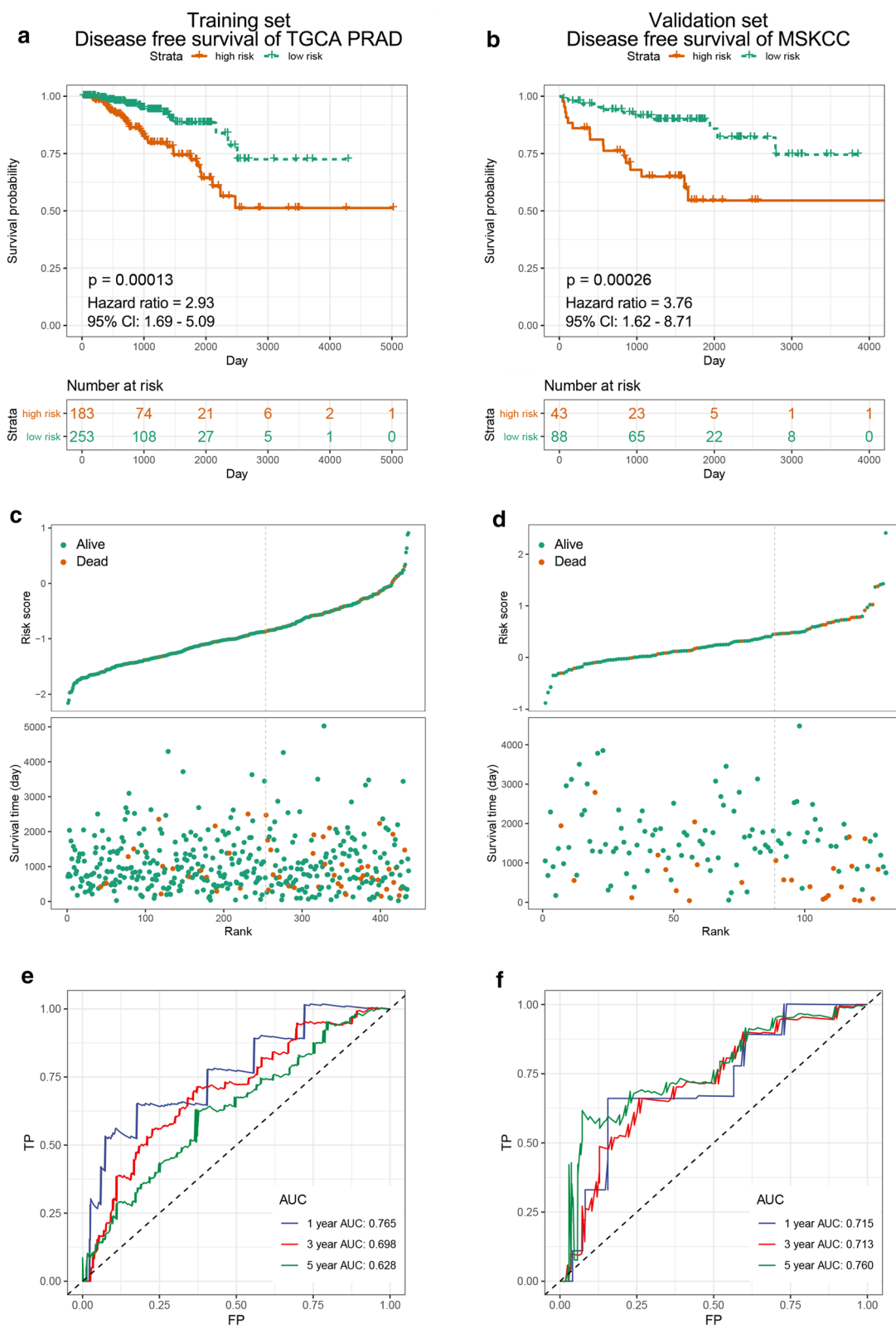


DFS and low-risk groups were associated with better DFS ( $p=0.00013$ ) (Fig. 3a, c). Time-dependent ROC analysis showed that the area under the ROC curve (AUC) for DFS in TCGA-PRAD cohort was 0.765 at 1 year, 0.698 at 3 years and 0.628 at 5 years (Fig. 3e).

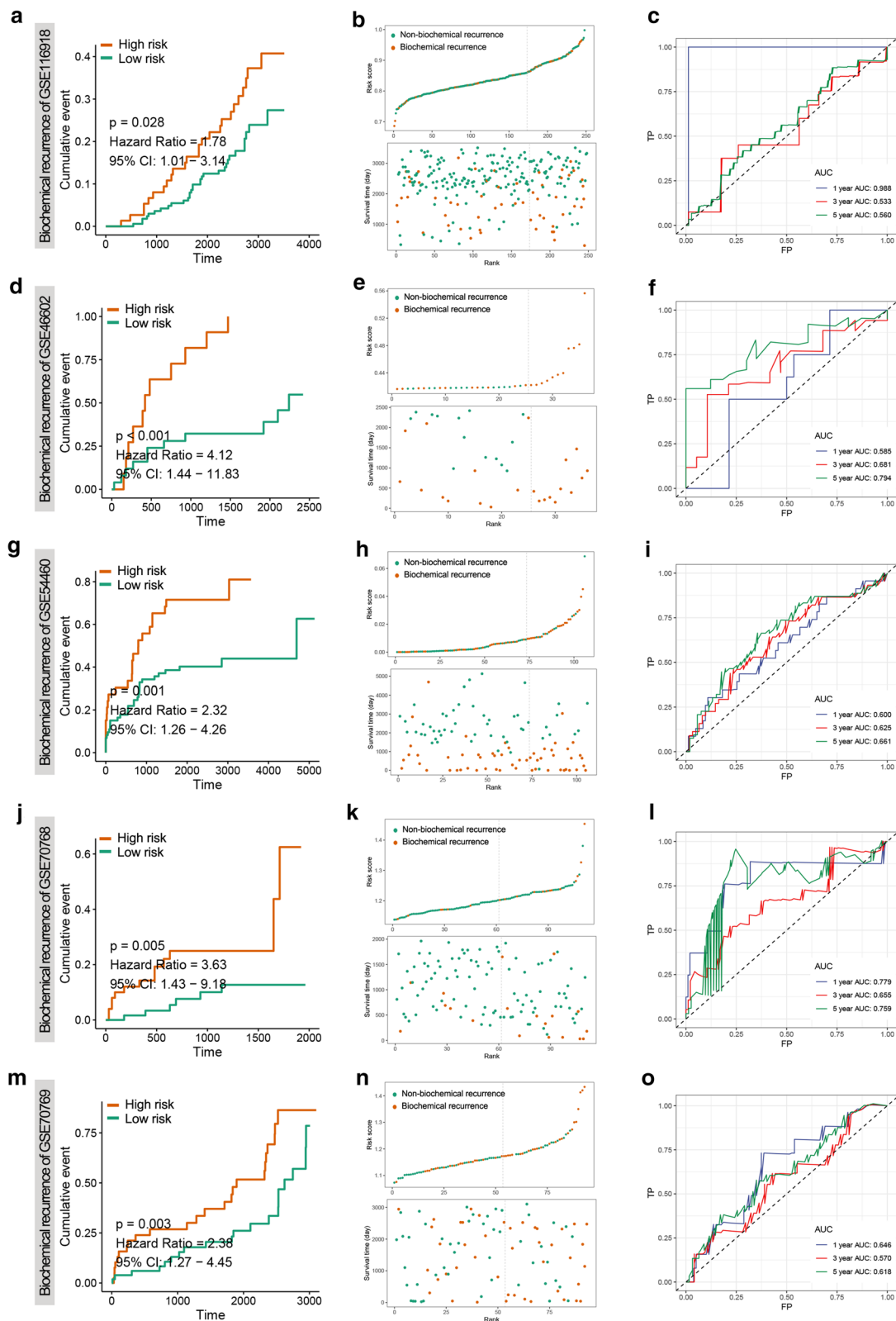
**Validation of prognostic model**

To validate our outcome model, MSKCC cohort was used for the DFS validation. We could observe the same significant prognostic value with  $p=0.00026$  and AUC with 1-, 3- and 5-year prognostic accuracies were 0.715,

0.713, 0.760, respectively (Fig. 3b, d, e). Since the total Gleason score was correlated with tumour behaviour, we then investigated our model’s role in overall survival and recurrence free survival. Due to few dead patients in TCGA-PRAD OS cohort, we didn’t include this cohort and used 2 independent cohorts to prove its utility in OS prediction. The results from the two OS validation datasets (GSE16560 and GSE53922) showed significant prognostic values were  $p=0.005$  and  $p=0.032$ , respectively. The AUC of each dataset was 0.606 and 0.585 at 1 year; 0.562 and 0.552 at 3 years; 0.608 and 0.495 at



**Fig. 3** Risk score derived from 3-gene signature is a prognostic biomarker for disease free survival (DFS). **a, c, e** KM survival, risk score and time-dependent ROC curves of DFS in TCGA-PRAD training cohort. **b, d, f** KM survival, risk score and time-dependent ROC curves of DFS in MSKCC validation cohort. The high-risk and low-risk groups were stratified at optimal cut-off due to the risk score. The AUC was assessed at 1, 3 and 5 years



(See figure on previous page.)

**Fig. 4** Risk score derived from 3-gene signature is a prognostic biomarker for recurrence free survival (RFS). **a–c** KM survival, risk score and time-dependent ROC curves of RFS in GSE116918 validation cohort. **d–f** KM survival, risk score and time-dependent ROC curves of RFS in GSE46602 validation cohort. **g–i** KM survival, risk score and time-dependent ROC curves of RFS in GSE54460 validation cohort. **j–l** KM survival, risk score and time-dependent ROC curves of RFS in GSE70768 validation cohort. **m–o** KM survival, risk score and time-dependent ROC curves of RFS in GSE70769 validation cohort. The high-risk and low-risk groups were stratified at optimal cut-off due to the risk score. The AUC was assessed at 1, 3 and 5 years

5 years, respectively (Additional file 8: Figure S4). Moreover, in the 5 validation cohorts (GSE116918, GSE46602, GSE54460, GSE70768 and GSE70769) for RFS, we could obviously see the significant outcomes that all the high-risk groups were associated with the poorer prognosis (Fig. 4a, d, g, j, m). The significant prognostic values of 5 RFS validation cohorts were  $p=0.028$ ,  $p<0.0001$ ,  $p=0.001$ ,  $p=0.005$  and  $p=0.003$ , respectively. The AUC of each dataset was 0.988, 0.585, 0.600, 0.779 and 0.646 at 1 year; 0.533, 0.681, 0.625, 0.655 and 0.570 at 3 years; 0.560, 0.794, 0.661, 0.759 and 0.618 at 5 years, respectively (Fig. 4).

#### Metastasis phenotype identification and metastasis free survival validation

Since Gleason score was correlated with tumor invasion, thus we used GSEA analysis to evaluate the correlation between metastasis and our model. We could find that high risk score was positively correlated with metastasis up and negatively correlated with metastasis down gene sets (Fig. 5a). Then the MFS validation set GSE116918 revealed that high-risk groups were significantly associated with poorer MFS and low-risk groups were associated with better MFS ( $p=0.009$ ) (Fig. 5b). Time-dependent ROC analysis showed that the area under the ROC curve (AUC) for MFS in GSE116918 cohort was 0.988 at 1 year, 0.896 at 3 years and 0.659 at 5 years (Fig. 5c).

#### Correlation between the risk score with other clinicopathological characteristics

Clinicopathological data, including age, clinical M stage, clinical T stage, total Gleason score, laterality, number of positive lymphonodes, pathological N stage, pathological T stage, PSA value, radiation therapy and targeted molecular therapy were collected from TCGA-PRAD dataset. The detailed information of patients' clinicopathological **characteristics** in TCGA-PRAD cohort were displayed in the Additional file 9: Table S5. Comparison results between risk score and different clinicopathological characters were shown in Additional file 10: Figure S5. In terms of clinical features, risk score was robustly increased in patients with lymphovascular invasion, more advanced stage and Gleason score and additional therapy, which indicated risk score was significantly positive correlated with tumour malignancy.

#### Subgroup analysis of prognostic value of the outcome model

To check the good applicability of our outcome model, the stratification survival analyses were performed. The prognosis of the high-risk group in different age, early clinical stage, late pathological stage, pathological N-stage and different lymphonodes positive subgroups were still worse than the low-risk group (Additional file 11: Figure S6a, b, e, h, i, k, l). Although there was no significance between different Gleason score groups, late clinical stage, early pathological stage and pathological N+ stage, we could still either find a worse prognostic trend in high risk group or more high-risk patients classified into high risk level group (Additional file 11: Figure S6c, d, f, g, j).

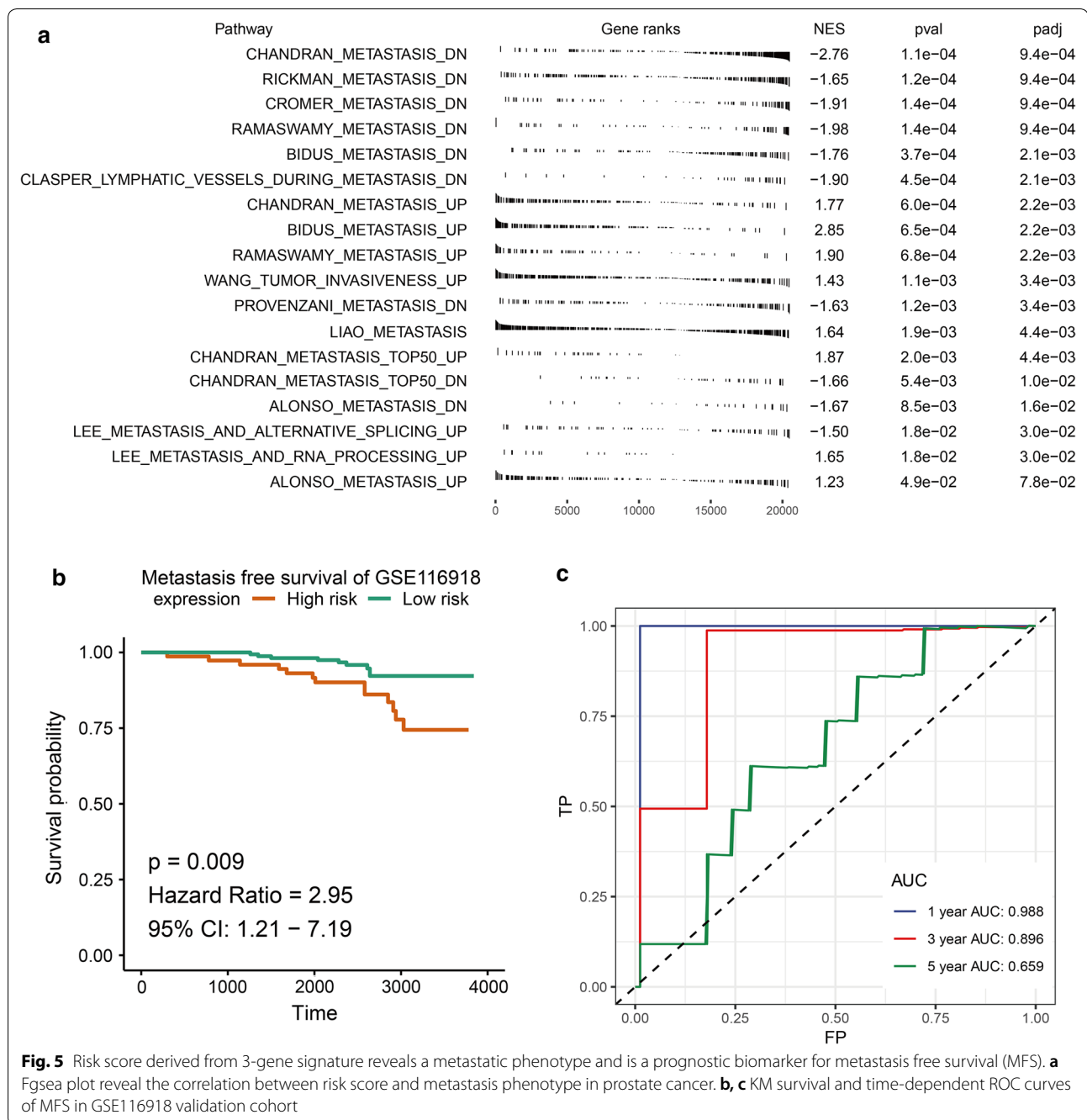
#### Univariate cox regression analyses for the model prognostic ability and nomogram construction

We performed univariate cox regression analysis to investigate whether our model was a clinically independent prognostic factor for PCa patients. And from the univariate regression analysis, Gleason score, risk score, pathological T stage, clinical T stage, PSA value, targeted molecular therapy, radiation therapy and number of lymph nodes positive were significant (Fig. 6a). When we merged all clinical features to perform the multivariate cox regression, we found it would lose almost half of our included patients. Thus we skipped this step and used all significant variables to construct the nomogram which could provide a quantitative method for the clinicians to predict the probability of 3-, 5- and 8-year DFS in PCa patients (Fig. 6b). Every patient would get a total point by plus the each prognostic parameters point, and the higher total points mean a worse outcome for that patient. Moreover, the calibration curve indicated that good performance in the estimation of 3-, 5- and 8-year DFS of the nomogram compared with the estimation of Kaplan–Meier (Fig. 6c–e). The results of DCA analysis also demonstrated that our nomogram was of high potential for clinical usefulness (Fig. 6f–h).

#### Functional enrichment analysis

GSVA analysis showed that 30 gene sets were significantly changed, meanwhile 41 changed significantly via GSVA analysis (Fig. 7a, b and Additional file 12, 13: Tables S6, S7). We then chose the common activated





or suppressed gene sets and eventually 11 activated gene sets (DNA repair, G2M checkpoint, MYC targets V1/2, oxidative phosphorylation, E2F targets, glycolysis, mitotic spindle, MTORC1 signaling, spermatogenesis and unfolded protein response) and 17 suppressed gene sets (allograft rejection, inflammatory response, interferon gamma response, myogenesis, TNFA signaling via NFkB, apical junction, complement, epithelial

mesenchymal transition, estrogen response early, KRAS signaling UP/DN, apical surface, apoptosis, hypoxia, IL2/STAT5 signaling, IL6/JAK/STAT3 signaling and UV response DN) were identified (Fig. 7c–f).

#### The landscape of immune infiltration in prostate cancer

Based on the CIBERSORT algorithm, we obtained an estimation of the abundances of 22 immune cells infiltrating in prostate cancer (Fig. 8a and Additional file 14:

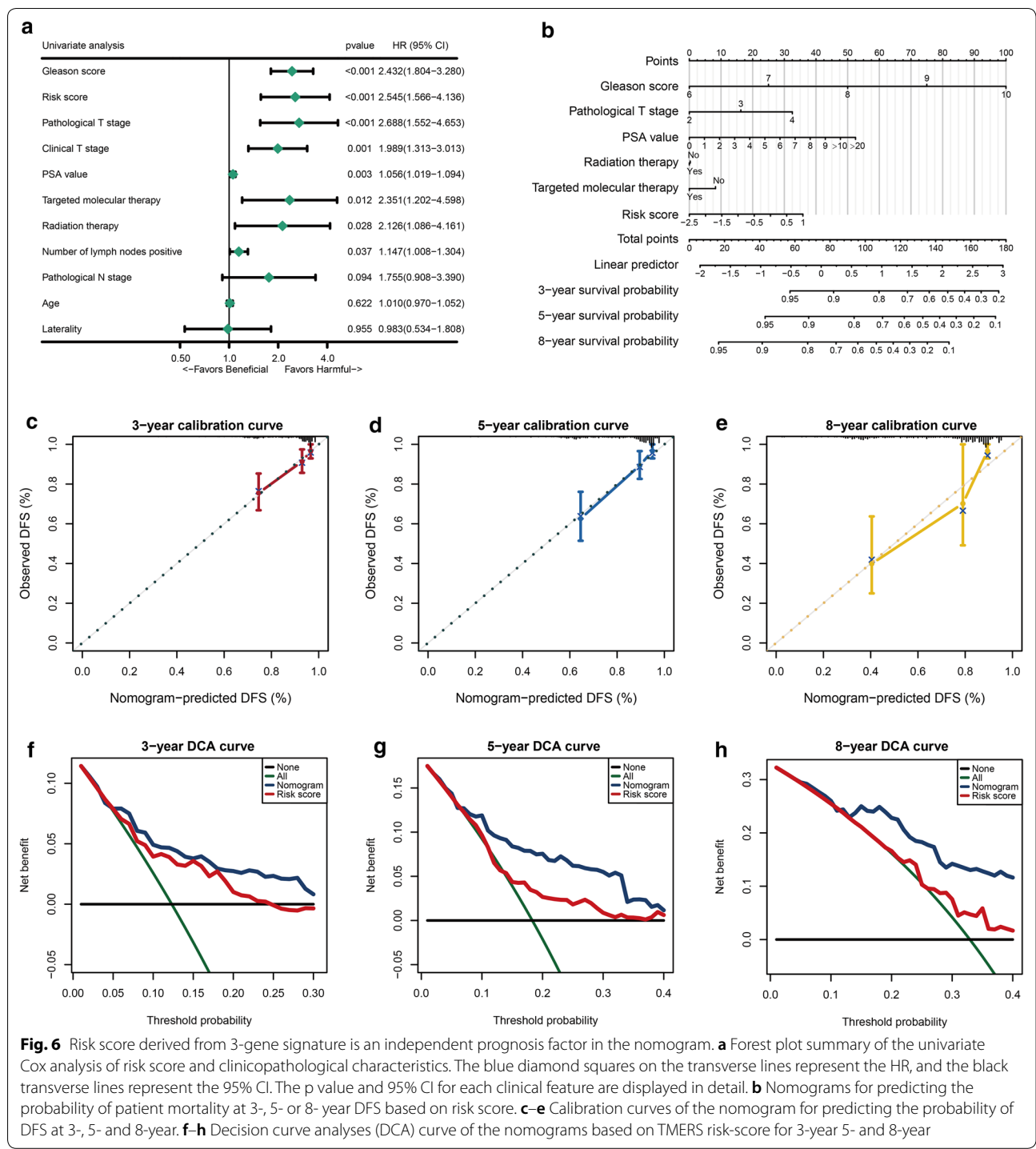
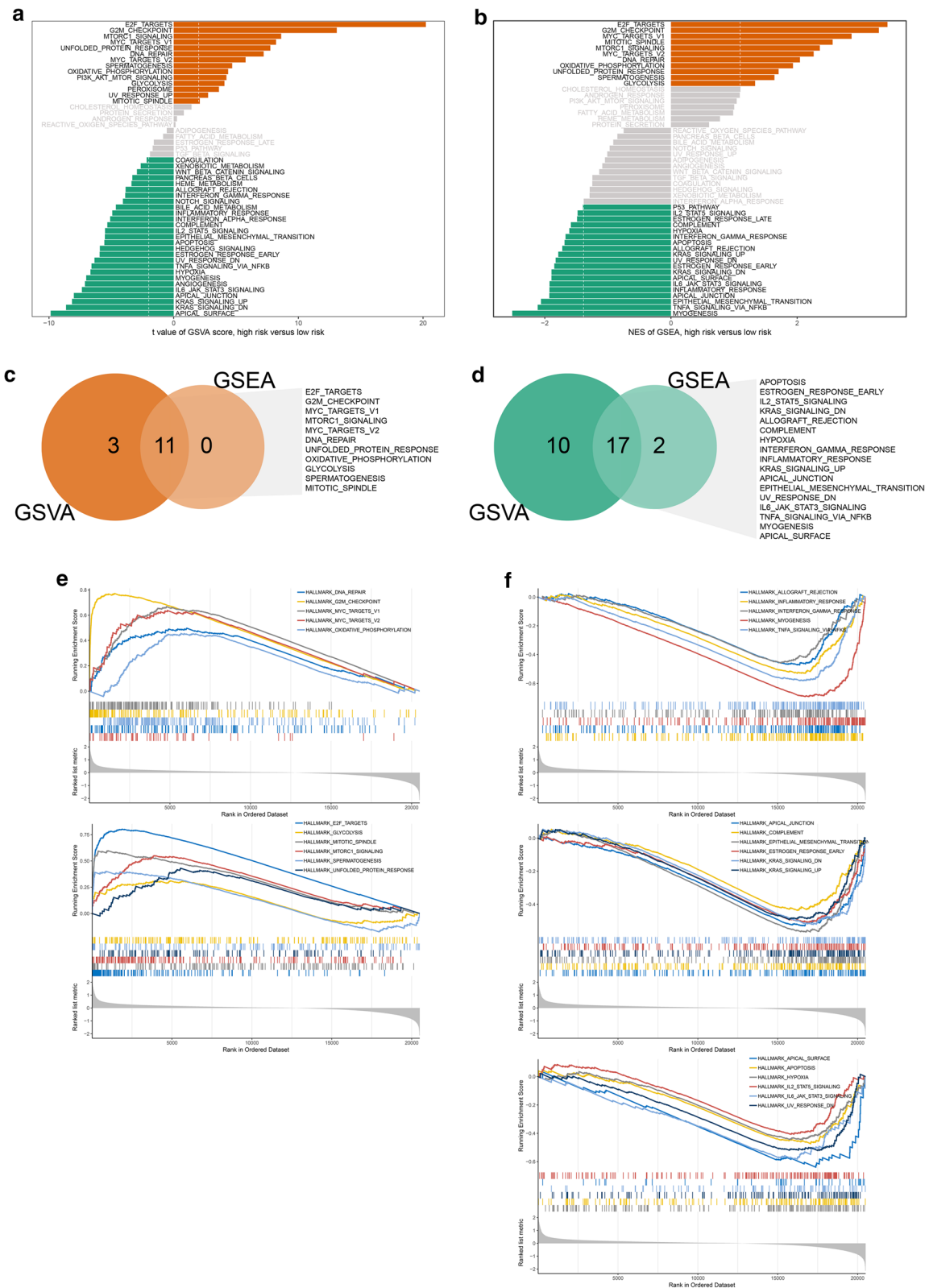
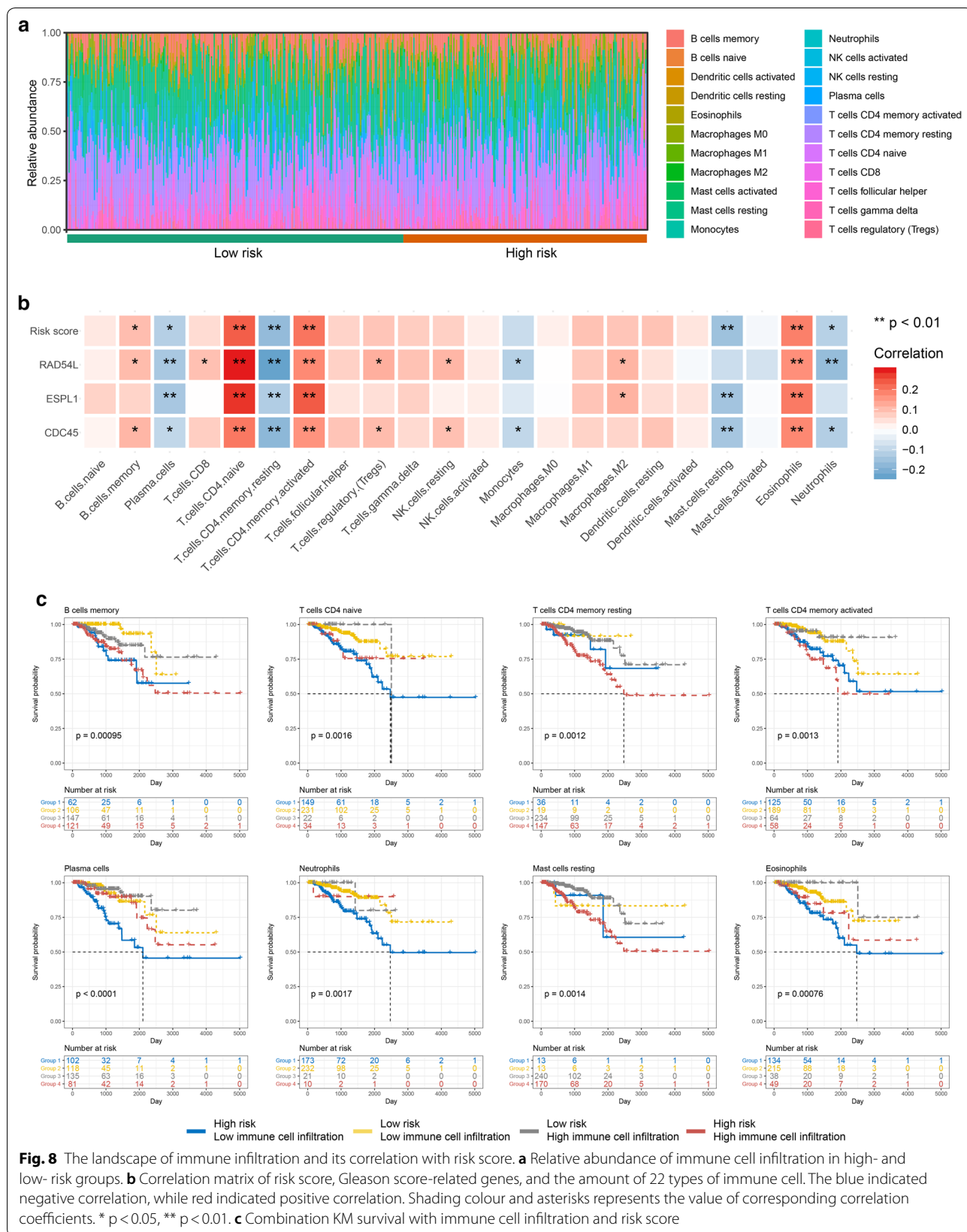


Table S8). We then calculated the correlation between immune cell infiltrations, risk score as well as 3 genes in the outcome model (Additional file 15: Table S9). The results showed that B cells memory, T cells CD4 native, T cells CD4 memory activated and eosinophil were positively correlated with our risk score and 3 genes in

outcome model; while, plasma cells, T cells CD4 memory resting, mast cells resting and neutrophil were negatively correlated with our risk score and 3 genes in outcome model (Fig. 8b). Furthermore, we used those 8 immune cells to perform combination survival analysis with our risk score and we could find that each of them could



**Fig. 7** Functional analysis based on TCGA-PRAD. **a** Barplot of GSEA results. **b** Barplot of GSEA results. **c** Common activated gene sets. **d** Common suppressed gene sets. **e** GSEA plot of common activated gene sets. **f** GSEA plot of common suppressed gene sets





divide patients into 4 groups which showed significant prognosis, especially plasma cell infiltration (Fig. 8c). As for the correlation between risk score and immune checkpoints, we found that risk score is significantly negative correlated with GAL9, LAG3, PD1LG2 and PDL1 (Additional file 16: Figure S7).

## Discussion

In recent years, quite a lot of promising biomarkers either for diagnosis or prognosis prediction were identified via high-throughput transcriptome profiling techniques. Although some gene signatures or nomograms have been established to predict the outcome for human prostate cancer, few of them used different independent validation cohorts [31–33].

In this study, we first identified differentially expressed genes between normal prostate and prostate cancer and then by using WGCNA analysis, we discovered a module highly-correlated with Gleason score and selected hub genes in this module. Based on the hub genes, we then constructed a 3-gene outcome model via LASSO cox regression analysis to predict the disease free survival of human PCa in TCGA-PRAD. Simultaneously, the patients were divided into high- and low-risk groups based on optimal cutoff and high risk group revealed a poor prognosis than low risk group, which could be validated in MSKCC cohort. We also successfully validated the effectiveness of model in predicting OS, RFS and MFS in 7 independent datasets. In addition, the nomogram based on the model exhibits an impressive performance and clinical applicability.

The genes (CDC45, ESPL1 and RAD54L) in our prognostic model have been previously reported to be associated with various cancers. However, only few papers revealed their roles in human prostate cancer. Li et al. reported that cell division cycle 45 (CDC45) might be useful markers for predicting tumor metastasis and therapeutic targets for the treatment of PCa patients via protein–protein network analysis [34]. Zhang et al. demonstrated that extra spindle pole bodies like 1 (ESPL1) could encode separate protein, which was up-regulated in numerous human cancers including breast, bone, brain, and prostate [35]. As for RAD54L, Li et al. found that castration-resistant prostate cancer (CRPC) cells showed a set of homologous recombination (HR)-associated genes, including BRCA1, RAD54L, and RMI2 were elevated [36].

Since our outcome model showed considerable power in risk stratification, the potential biological process and signaling pathways need to be investigated. By using GSEA and GSVA analysis, we identified several activated pathways which were highly correlated with cell cycle such as MYC targets, E2F targets, mitotic

spindle and G2M checkpoint. Therefore, we supposed that outcome model derived cell cycle alteration might play a critical role in cancer progression which could lead to poorer prognosis in PCa patients. We also found some pathways associated with immune response and inflammation were suppressed which might suggest high-risk patients could have higher risk of immunosuppression.

The crosstalk of tumor and immune cells from the tumor microenvironment (TME) is essential for tumor progression and metastasis development [37, 38]. In this study, we also put emphasis on tumor immune cells infiltration. We calculated the correlation between immune cell abundance and risk score derived from outcome model. The 4 positive correlated immune cells were memory B cells, native CD4 T cells, activated CD4 memory T cells and eosinophil. It was reported that B-cells were activated, differentiate to a memory B-cell phenotype in high-grade serous ovarian cancer (HGSOC); they could be activated by DCs and promote a cytotoxic response and cause HGSOC metastases [39]. CD4<sup>+</sup> T cells were reported to play a central role in initiating and maintaining anticancer immune responses in human head and neck cancer [40]; meanwhile, they were also reported to correlate with lymph node involvement and unfavorable prognosis in human breast cancers [41]. Therefore, we cannot just easily draw a conclusion as the immune cells infiltration was dynamic. The other 4 negative correlated immune cells were plasma cells, resting CD4 memory T cells, resting mast cells and neutrophil. Mast cells were immune cells that accumulated in the tumors and their microenvironment during disease progression. For instance, mast cells could secrete pro-angiogenic and growth factors but also pro- and anti-inflammatory mediators, thus it was really difficult to explain clearly its role in tumor progression [42]. As for the tumor-associated neutrophils (TAN), it functioned like a “double-edged sword” [43]. On one hand, it could acquire a tumor-promoting phenotype via induction by TGF- $\beta$  [44]. On the other hand, it could mediate immunosuppression by inhibiting the tumor-killing function of cytotoxic T cells [45].

## Conclusions

In conclusion, our study was the first to construct an outcome model based on WGCNA analysis in human prostate cancer. Consisting of Gleason score related genes, our outcome model had promising potential in predicting prognosis of PCa. Meanwhile, some limitations of our study should be acknowledged. First, this is a retrospective bioinformatics analysis; large prospective clinical trials were needed to prove its real utility.

Second, experimental studies are required to further explore underlying mechanisms of 3 candidate genes in our outcome model.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12935-020-01230-x>.

**Additional file 1: Table S1.** Information of all datasets in our study.

**Additional file 2: Table S2.** Detail information for packages.

**Additional file 3: Table S3.** Differentially expressed genes in TCGA-PRAD.

**Additional file 4: Figure S1.** Determination of soft-thresholding power in the weighted gene co-expression network analysis (WGCNA). (a) Analysis of the scale-free fit index for various soft-thresholding powers ( $\beta$ ). (b) Analysis of the mean connectivity for various soft-thresholding powers. (c) Histogram of connectivity distribution when  $\beta = 4$ . (d) Checking the scale free topology when  $\beta = 4$ .

**Additional file 5: Table S4.** Hub genes correlated with Gleason score via WGCNA analysis.

**Additional file 6: Figure S2.** Construction of LASSO Cox regression model in TCGA-PRAD.

**Additional file 7: Figure S3.** Expression level of 3 Gleason score related genes. (a) Expression of 3 Gleason score related genes between prostate cancer and normal prostate in TCGA-PRAD. (b) Expression of 3 Gleason score related genes between different Gleason scores in TCGA-PRAD.

**Additional file 8: Figure S4.** Risk score derived from 3-gene signature is a prognostic biomarker for overall survival (OS). (a, c, e) KM survival, risk score and time-dependent ROC curves of OS in GSE16560 validation cohort. (b, d, f) KM survival, risk score and time-dependent ROC curves of OS in GSE53922 validation cohort. The high-risk and low-risk groups were stratified at optimal cut-off due to the risk score. The AUC was assessed at 1, 3 and 5 years.

**Additional file 9: Table S5.** Clinical information with DFS data of TCGA-PRAD.

**Additional file 10: Figure S5.** Association between risk score and clinicopathological characteristics. The survival rate of indicated subgroups in different clinicopathological characteristics was measured. Boxplots indicate the correlation between risk score and the indicated subtype of each clinicopathological characteristics by the t-test or one-way ANOVA. The patients are stratified into different subtypes based on (a) age: elder: age  $\geq 65$ , younger: age  $< 65$ . (b) clinical M stage: cM0 and cM1. (c) clinical T stage: cT1, cT2, cT3 and cT4. (d) total Gleason score: 6, 7, 8, 9, 10. (E) laterality: left, right and bilateral. (f) number of positive lymph nodes by HE: number of positive lymph nodes by HE = 0 and number of positive lymph nodes by HE  $> 0$ . (g) pathological N stage: pN0 and pN1. (h) pathological N stage: pT2, pT3 and pT4. (i) PSA value:  $< 10$ , 10–20 and  $> 20$ . (j) Radiation therapy: yes and no. (k) targeted molecular therapy: yes and no.

**Additional file 11: Figure S6.** KM survival subgroup analyses of all patients in TCGA-PRAD cohort according to the risk score stratified by clinical characteristics. (a) Age  $< 65$ . (b) Age  $\geq 65$ . (c) Gleason score 6–7. (d) Gleason score 8–10. (e) Clinical stage I–II. (f) Clinical stage III–IV. (g) Pathological stage I–II. (h) Pathological stage III–IV. (i) Pathological N stage - (N0). (j) Pathological N stage + (N+). (k) Number of positive lymph nodes by HE = 0. (l) Number of positive lymph nodes by HE  $> 0$ .

**Additional file 12: Table S6.** The results of GSEA analysis.

**Additional file 13: Table S7.** The results of GSVA analysis.

**Additional file 14: Table S8.** CIBERSORT Output matrix.

**Additional file 15: Table S9.** Correlation between risk score and immune cells infiltration.

**Additional file 16: Figure S7.** Correlation between risk score and immune checkpoints.

## Abbreviations

PCa: Prostate cancer; DEGs: Differentially expressed genes; PSA: Prostate specific antigen; PHI: Prostate health indexes; 4 K: Four-kallikrein panel; WGCNA: Weighted gene co-expression network analysis; GEO: Gene Expression Omnibus; FDR: False discovery rate; LASSO: Least absolute shrinkage and selection operator; ROC: Receiver operating characteristic curve; AUC: Area under curve; GSEA: Gene set enrichment analysis; GSVA: Gene set variation analysis; DCA: Decision curve analysis; Tregs: Regulatory T cells; OS: Overall survival; RFS: Recurrence free survival; MFS: Metastasis free survival; CDC45: Cell division cycle 45; ESPL1: Extra spindle pole bodies like 1; RAD54L: DNA repair and recombination protein RAD54-like; CRPC: Castration-resistant prostate cancer; HR: Homologous recombination; TME: Tumor microenvironment; HGSOC: High-grade serous ovarian cancer; TAN: Tumor-associated neutrophils.

## Acknowledgements

The authors are grateful for the invaluable support and useful discussions with other members of the Department of Urology.

## Authors' contributions

YW and ZY made substantial contributions to conception and design of the research. YW carried out data collection and analysis. YW and ZY wrote the paper. YW and ZY edited the manuscript and provided critical comments. Both authors read and approved the final manuscript.

## Funding

This work was supported by grants from Science, Technology and Innovation Seed Fund of Zhongnan Hospital of Wuhan University (Grant No. cxy20160036) and Joint foundation of Health Commission of Hubei Province (Grant No. WJ2019H088).

## Availability of data and materials

All data generated or analysed during this study are included in this published article and Additional files.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

Received: 10 January 2020 Accepted: 23 April 2020

Published online: 11 May 2020

## References

- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68(1):7–30.
- Schatten H. Brief overview of prostate cancer statistics, grading, diagnosis and treatment strategies. *Adv Exp Med Biol*. 2018;1095:1–14.
- Palmbo PL, Hussain M. Non-castrate metastatic prostate cancer: have the treatment options changed? *Semin Oncol*. 2013;40(3):337–46.
- Ost P, Bossi A, Decaestecker K, De Meerleer G, Giannarini G, Karnes RJ, Roach M 3rd, Briganti A. Metastasis-directed therapy of regional and distant recurrences after curative treatment of prostate cancer: a systematic review of the literature. *Eur Urol*. 2015;67(5):852–63.
- Lotan TL, Epstein JI. Clinical implications of changing definitions within the Gleason grading system. *Nat Rev Urol*. 2010;7(3):136–42.
- Fizazi K, Flaig TW, Stockle M, Scher HI, de Bono JS, Rathkopf DE, Ryan CJ, Kheoh T, Li J, Todd MB, et al. Does Gleason score at initial diagnosis predict efficacy of abiraterone acetate therapy in patients with metastatic castration-resistant prostate cancer? an analysis of abiraterone acetate phase III trials. *Ann Oncol*. 2016;27(4):699–705.
- Hugosson J, Carlsson S, Aus G, Bergdahl S, Khatami A, Lodding P, Pihl CG, Stranne J, Holmberg E, Lilja H. Mortality results from the Goteborg randomised population-based prostate-cancer screening trial. *Lancet Oncol*. 2010;11(8):725–32.

8. Preston MA, Batista JL, Wilson KM, Carlsson SV, Gerke T, Sjoberg DD, Dahl DM, Sesso HD, Feldman AS, Gann PH, et al. Baseline prostate-specific antigen levels in midlife predict lethal prostate cancer. *J Clin Oncol*. 2016;34(23):2705–11.
9. Loeb S, Bjurlin MA, Nicholson J, Tammela TL, Penson DF, Carter HB, Carroll P, Etzioni R. Overdiagnosis and overtreatment of prostate cancer. *Eur Urol*. 2014;65(6):1046–55.
10. Garraway LA, Verweij J, Ballman KV. Precision oncology: an overview. *J Clin Oncol*. 2013;31(15):1803–5.
11. Mano R, Eastham J, Yossepowitch O. The very-high-risk prostate cancer: a contemporary update. *Prostate Cancer Prostatic Dis*. 2016;19(4):340–8.
12. Kretschmer A, Tilki D. Biomarkers in prostate cancer—current clinical utility and future perspectives. *Crit Rev Oncol Hematol*. 2017;120:180–93.
13. Stephan C, Vincendeau S, Houlgatte A, Cammann H, Jung K, Semjonow A. Multicenter evaluation of [-2]prostate-specific antigen and the prostate health index for detecting prostate cancer. *Clin Chem*. 2013;59(1):306–14.
14. Vickers A, Cronin A, Roobol M, Savage C, Peltola M, Pettersson K, Scardino PT, Schroder F, Lilja H. Reducing unnecessary biopsy during prostate cancer screening using a four-kallikrein panel: an independent replication. *J Clin Oncol*. 2010;28(15):2493–8.
15. Nordstrom T, Vickers A, Assel M, Lilja H, Gronberg H, Eklund M. Comparison between the four-kallikrein panel and prostate health index for predicting prostate cancer. *Eur Urol*. 2015;68(1):139–46.
16. Esfahani M, Ataei N, Panjehpour M. Biomarkers for evaluation of prostate cancer prognosis. *Asian Pac J Cancer Prev*. 2015;16(7):2601–11.
17. Wise HM, Hermida MA, Leslie NR. Prostate cancer, PI3K, PTEN and prognosis. *Clin Sci (Lond)*. 2017;131(3):197–210.
18. Larne O, Hagman Z, Lilja H, Bjartell A, Edsjo A, Ceder Y. miR-145 suppress the androgen receptor in prostate cancer cells and correlates to prostate cancer prognosis. *Carcinogenesis*. 2015;36(8):858–66.
19. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. 2008;9:559.
20. Zhou Z, Cheng Y, Jiang Y, Liu S, Zhang M, Liu J, Zhao Q. Ten hub genes associated with progression and prognosis of pancreatic carcinoma identified by co-expression analysis. *Int J Biol Sci*. 2018;14(2):124–36.
21. Yuan L, Shu B, Chen L, Qian K, Wang Y, Qian G, Zhu Y, Cao X, Xie C, Xiao Y, et al. Overexpression of COL3A1 confers a poor prognosis in human bladder cancer identified by co-expression analysis. *Oncotarget*. 2017;8(41):70508–20.
22. Clarke C, Madden SF, Doolan P, Aherne ST, Joyce H, O'Driscoll L, Gallagher WM, Hennessy BT, Moriarty M, Crown J, et al. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. *Carcinogenesis*. 2013;34(10):2300–8.
23. Zhang JX, Song W, Chen ZH, Wei JH, Liao YJ, Lei J, Hu M, Chen GZ, Liao B, Lu J, et al. Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. *Lancet Oncol*. 2013;14(13):1295–306.
24. Yang K, Hou Y, Li A, Li Z, Wang W, Xie H, Rong Z, Lou G, Li K. Erratum: identification of a six-lncRNA signature associated with recurrence of ovarian cancer. *Sci Rep*. 2017;7(1):11481.
25. Jiang Y, Zhang Q, Hu Y, Li T, Yu J, Zhao L, Ye G, Deng H, Mou T, Cai S, et al. ImmunoScore signature: a prognostic and predictive tool in gastric cancer. *Ann Surg*. 2018;267(3):504–13.
26. Chen L, Luo Y, Wang G, Qian K, Qian G, Wu CL, Dan HC, Wang X, Xiao Y. Prognostic value of a gene signature in clear cell renal cell carcinoma. *J Cell Physiol*. 2019;234(7):10324–35.
27. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1–22.
28. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics*. 2005;61(1):92–105.
29. Hanzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform*. 2013;14:7.
30. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–7.
31. Shi R, Bao X, Weischenfeldt J, Schaefer C, Rogowski P, Schmidt-Hegemann NS, Unger K, Lauber K, Wang X, Buchner A, et al. A novel gene signature-based model predicts biochemical recurrence-free survival in prostate cancer patients after radical prostatectomy. *Cancers (Basel)*. 2019;12(1):E1.
32. Pinskaya M, Saci Z, Gallopin M, Gabriel M, Nguyen HT, Firlje V, Describes M, Rapinat A, Gentien D, Taille A, et al. Reference-free transcriptome exploration reveals novel RNAs for prostate cancer diagnosis. *Life sci Alliance*. 2019;2(6):e201900449.
33. Zhao Z, Weickmann S, Jung M, Lein M, Kilic E, Stephan C, Erbersdobler A, Fendler A, Jung K. A novel predictor tool of biochemical recurrence after radical prostatectomy based on a five-microRNA tissue signature. *Cancers (Basel)*. 2019;11(10):1603.
34. Li HY, Jin N, Han YP, Jin XF. Pathway crosstalk analysis in prostate cancer based on protein-protein network data. *Neoplasma*. 2017;64(1):22–31.
35. Zhang N, Pati D. Biology and insights into the role of cohesin protease separase in human malignancies. *Biol Rev Camb Philos Soc*. 2017;92(4):2070–83.
36. Li L, Karanika S, Yang G, Wang J, Park S, Broom BM, Manyam GC, Wu W, Luo Y, Basourakos S, et al. Androgen receptor inhibitor-induced “BRCA-ness” and PARP inhibition are synthetically lethal for castration-resistant prostate cancer. *Science Signal*. 2017;10(480):eaam7479.
37. Dirat B, Bochet L, Dabek M, Daviaud D, Dauvillier S, Majed B, Wang YY, Meulle A, Salles B, Le Gonidec S, et al. Cancer-associated adipocytes exhibit an activated phenotype and contribute to breast cancer invasion. *Cancer Res*. 2011;71(7):2455–65.
38. Motrescu ER, Rio MC. Cancer cells, adipocytes and matrix metalloproteinase 11: a vicious tumor progression cycle. *Biol Chem*. 2008;389(8):1037–41.
39. Lee GT, Srivastava A, Kwon YS, Kim IY. Immune reaction by cytoreductive prostatectomy. *Am J Clin Exp Urol*. 2019;7(2):64–79.
40. Badoual C, Hans S, Rodriguez J, Peyrard S, Klein C, Agueznay Nel H, Mosseri V, Laccourreye O, Bruneval P, Fridman WH, et al. Prognostic value of tumor-infiltrating CD4+ T-cell subpopulations in head and neck cancers. *Clin Cancer Res*. 2006;12(2):465–72.
41. Matkowski R, Gisterek I, Halon A, Lacko A, Szweczyk K, Staszek U, Pudelko M, Szynglarewicz B, Szelachowska J, Zolnieriek A, et al. The prognostic role of tumor-infiltrating CD4 and CD8 T lymphocytes in breast cancer. *Anticancer Res*. 2009;29(7):2445–51.
42. Maciel TT, Moura IC, Hermine O. The role of mast cells in cancers. *F1000 Prime Rep*. 2015;7:09.
43. Fridlender ZG, Sun J, Kim S, Kapoor V, Cheng G, Ling L, Worthen GS, Albelda SM. Polarization of tumor-associated neutrophil phenotype by TGF-beta: “N1” versus “N2” TAN. *Cancer Cell*. 2009;16(3):183–94.
44. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646–74.
45. Casbon AJ, Reynaud D, Park C, Khuc E, Gan DD, Schepers K, Passegue E, Werb Z. Invasive breast cancer reprograms early myeloid differentiation in the bone marrow to generate immunosuppressive neutrophils. *Proc Natl Acad Sci USA*. 2015;112(6):E566–75.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.