



A Robust Prognostic Gene Signature Based on eRNAs-Driven Genes in Prostate Cancer

Shuaishuai Fan^{1†}, Zheng Wang^{1,2†}, Li Zhao^{3†}, ChenHui Zhao^{1,4**}, DaJiang Yuan^{3**} and Jingqi Wang^{1**}

¹ First Clinical Medical College, Shanxi Medical University, First Hospital of Shanxi Medical University, Taiyuan, China,

² People's Hospital of Zezhou County, Jincheng, China, ³ Department of Anesthesia, Shanxi Medical University, Taiyuan, China, ⁴ The First People's Hospital of Jinzhong, Jinzhong, China

OPEN ACCESS

Edited by:

Ming Tang,
Dana–Farber Cancer Institute,
United States

Reviewed by:

Yang Liu,
Dana–Farber Cancer Institute,
United States
Nathan Lack,
Koç University, Turkey

*Correspondence:

ChenHui Zhao
605207579@qq.com
DaJiang Yuan
yuandajiang@sina.com
Jingqi Wang
drwangjq@126.com

[†]These authors have contributed equally to this work and share first authorship

[‡]These authors have contributed equally to this work

Specialty section:

This article was submitted to
Cancer Genetics and Oncogenomics,
a section of the journal
Frontiers in Genetics

Received: 06 March 2021

Accepted: 01 June 2021

Published: 29 June 2021

Citation:

Fan S, Wang Z, Zhao L, Zhao C,
Yuan D and Wang JQ (2021) A
Robust Prognostic Gene Signature
Based on eRNAs-Driven Genes
in Prostate Cancer.
Front. Genet. 12:676845.
doi: 10.3389/fgene.2021.676845

Prostate cancer (PCa) is the second most common malignancy in men, but its exact pathogenetic mechanisms remain unclear. This study explores the effect of enhancer RNAs (eRNAs) in PCa. Firstly, we screened eRNAs and eRNA-driven genes from The Cancer Genome Atlas (TCGA) database, which are related to the disease-free survival (DFS) of PCa patients; screening methods included bootstrapping, Kaplan–Meier (KM) survival analysis, and Pearson correlation analysis. Then, a risk score model was established using multivariate Cox analysis, and the results were validated in three independent cohorts. Finally, we explored the function of eRNA-driven genes through enrichment analysis and analyzed drug sensitivity on datasets from the Genomics of Drug Sensitivity in Cancer database. We constructed and validated a robust prognostic gene signature involving three eRNA-driven genes namely *MAPK15*, *ZNF467*, and *MC1R*. Moreover, we evaluated the function of eRNA-driven genes associated with tumor microenvironment (TME) and tumor mutational burden (TMB), and identified remarkable differences in drug sensitivity between high- and low-risk groups. This study identified a prognostic gene signature, which provides new insights into the role of eRNAs and eRNA-driven genes while assisting clinicians to determine the prognosis and appropriate treatment options for patients with PCa.

Keywords: prostate cancer, enhancer RNA, prognostic gene signature, risk score model, drug sensitivity

INTRODUCTION

Prostate cancer (PCa) is the second most common malignancy in men (No Authors Listed, 2020). At present, the main treatment for primary PCa is radical prostatectomy (Albertsen, 2020), and although most patients with PCa benefit from this procedure, the recurrence rate following this treatment remains high (27–53%) (Van den Broeck et al., 2019). Almost all recurrent PCa will

Abbreviations: PCa, Prostate cancer; eRNAs, Enhancer RNAs; lncRNAs, Long non-coding RNAs; SCCHN, head and neck squamous cell carcinoma; GO, Gene ontology; GSEA, Gene Set Enrichment Analysis; DEGs, differentially expressed genes; TCGA, The Cancer Genome Atlas; NIH, National Institutes of Health; GDC, Genomic Data Commons; RNA-seq, RNA sequencing; GEO, Gene Expression Omnibus; TIMER2.0, Tumor Immune Estimation Resource 2.0; PRAD, prostate adenocarcinoma; GDSC, Genomics of Drug Sensitivity in Cancer; EMT, Epithelial-mesenchymal transformation; TGF- β , transforming growth factor β ; ECM, extracellular matrix; MSigDB, Molecular Signatures Database; DFS, disease-free survival; KM, Kaplan–Meier; LASSO, least absolute shrinkage and selection operator; ROC, receiver operator characteristic; AUC, area under the curve; FDR, false discovery rate; FC, fold change; C-index, concordance index; TME, tumor microenvironment; TMB, tumor mutational burden; MSI, microsatellite instability; IC50, half-inhibitory concentration.

develop into advanced PCa and further progress into castration-resistant prostate cancer (CRPC). For advanced PCa and CRPC, targeted therapy and chemotherapy are the main treatment options of disease management (Connor et al., 2021). Therefore, exploring the mechanisms involved in PCa and identifying effective PCa biomarkers, are important measures to improve the prognosis of patients with PCa.

Enhancer RNAs (eRNAs)—important components of long non-coding RNAs (lncRNAs)—are transcribed from enhancer regions and important in gene transcriptional regulation (Sedano et al., 2020). Numerous eRNAs have been found in human cells, which occupy a high position in the transcriptional circuit by mediating the activation of target genes (Ding et al., 2018). eRNAs, induced by tumor suppressors, could suppress tumors (Wang et al., 2020). On the contrary, oncogene-induced eRNAs can directly promote tumorigenesis (Liang et al., 2016). Zhang et al. (2019) also showed that eRNAs play an important role in the development and treatment of tumors. These findings illustrate the importance of eRNAs in the process of human oncogenesis.

With the advent of high-throughput sequencing, it is possible to explore the function of eRNAs across the whole genome. Previous studies have conducted preliminary investigations of the functions of eRNAs. For example, Xiaolian et al. found that AP001056.1—a key immune-related eRNA in head and neck squamous cell carcinoma (SCCHN)—has a positive effect on clinical outcomes in patients with SCCHN (Liang et al., 2016). However, the role of eRNAs in PCa is not clear. Therefore, we used high-throughput sequencing technology and bioinformatics to screen eRNAs and target genes that may influence the prognosis of patients with PCa and constructed a robust prognostic model. Moreover, we explored the mechanism of eRNAs' influence on PCa development, as well as the predictive power of the model for PCa drug sensitivity and potential PCa drug identification. We also investigated the functions of prognostic genes associated with PCa using a multi-omics, multi-database approach involving various methods. The workflow of our research is shown in **Figure 1A**. In total, 835 prostate samples were included, and the pathogenesis and treatment of PCa was explored.

MATERIALS AND METHODS

Data Collection and Pre-processing

The transcriptome data of PCa and paracancerous tissues, as well as clinical data of patients with PCa as per The Cancer Genome Atlas (TCGA) database, were obtained from the National Institutes of Health (NIH) Genomic Data Commons (GDC Data Portal¹). Using the Gene Expression Omnibus (GEO²) and the cBioPortal (cBio³), we searched for PCa datasets that included: (1) PCa and RNA-seq or micro-array dataset types; (2) more

than 100 PCa samples in survival data; and (3) information on the expression of the genes in the gene signature. Per these parameters we obtained two datasets, namely GSE54460 and GSE70768 from the GEO and one dataset, MSKCC (Cerami et al., 2012), from cBio. We also obtained data on the lncRNA-target gene relationships, as predicted by PreSTIGE (Predicting Specific Tissue Interactions of Genes and Enhancers), and the lncRNA expressions from active tissue-specific enhancers. Using the human gene annotation file and Perl, we converted the Ensembl ID to its corresponding transcript ID and gene name, respectively. Using the Tumor Immune Estimation Resource 2.0 (TIMER2.0), we downloaded the immune cell infiltration data of prostate adenocarcinoma (PRAD) tissues from the Cistrome Project⁴. We obtained the gene sets associated with epithelial-mesenchymal transformation (EMT), transforming growth factor β (TGF- β), and extracellular matrix (ECM) from the Molecular Signatures Database (MSigDB⁵) (Zhang et al., 2020). The copy number alterations (CNAs) analysis was performed by cBio.

Screening of eRNAs and Target Genes Associated With the Prognosis of Patients With PCa

To preliminarily screen eRNAs related to the prognosis of patients with PCa, we analyzed the eRNAs and their expression levels along with their corresponding clinical disease-free survival (DFS) rates. Through application of bootstrapping methods and the Kaplan–Meier (KM) survival analysis, candidate eRNAs were narrowed down as follows: 70% of the samples were randomly selected from the TCGA data cohort for gene survival analysis. The process was repeated 1,000 times. Genes that appeared more than 700 times among the samples (robustness test, $P < 0.05$) were regarded as robust prognostic genes (Zhou et al., 2019). The target genes related to prognosis were screened using the same method. Based on the correlation analysis between the eRNAs and target genes, we suggested that the target genes ($\text{cor} > 0.3$; $P < 0.05$; CNAs of target genes were no significant correlation with its corresponding eRNAs) are defined as eRNA-driven genes.

Building the Predictive Signature for Patients With PCa

For the training cohort from TCGA dataset, we adopted a least absolute shrinkage and selection operator (LASSO) regression analysis using the R package glmnet 4.0-2 (Friedman et al., 2010; Simon et al., 2011), to narrow down the range of eRNA-driven genes related to the prognosis of patients with PCa. Using the R package survival 3.2-7, we developed a risk model through performing multivariate Cox regression analysis. Applying the formula:

$$\text{risk score} = \sum_{k=1}^n (\text{risk genes expression}_k \times \text{coefficient}_k),$$

we were able to obtain the prognostic risk scores.

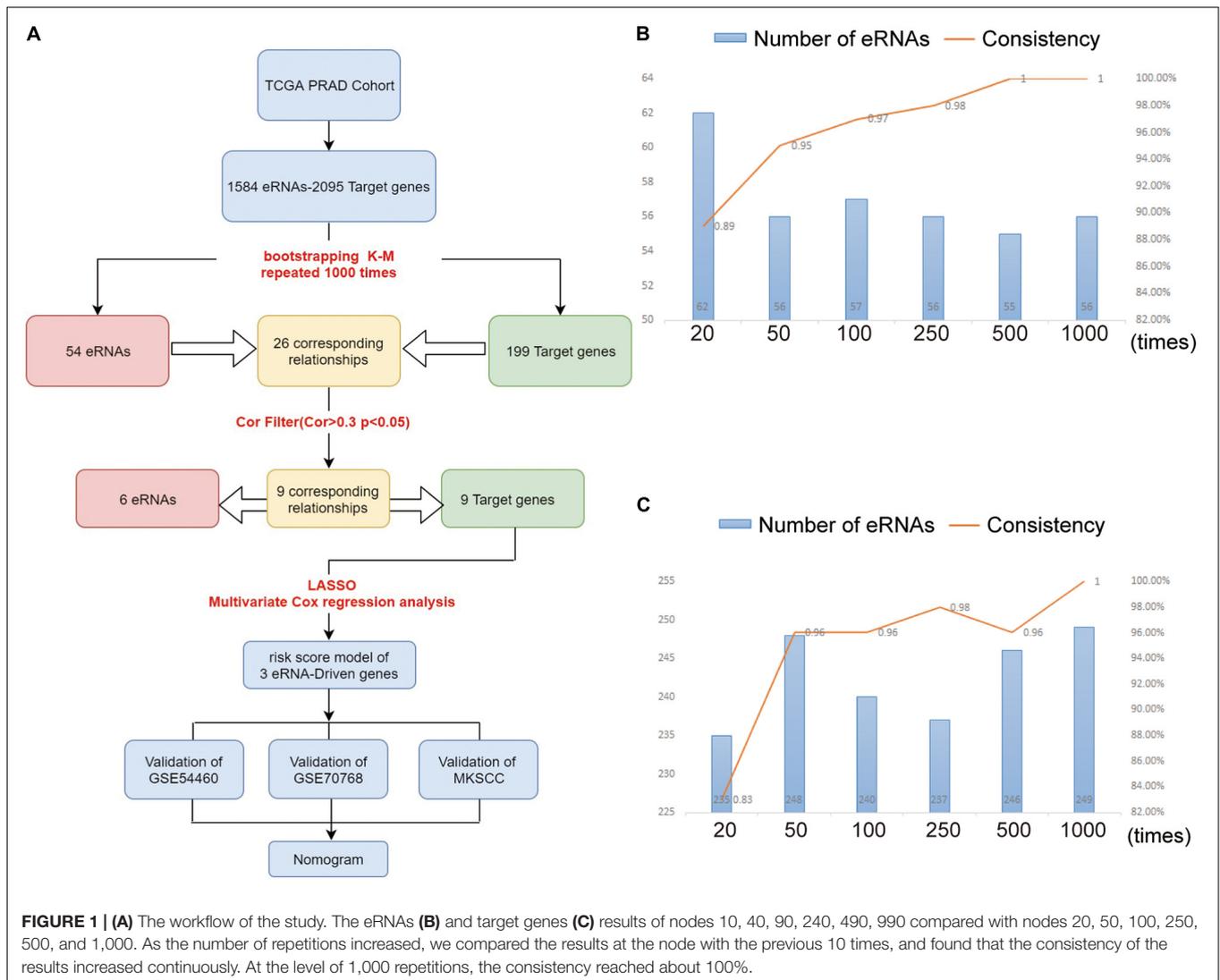
¹<https://portal.gdc.cancer.gov/>

²<http://www.ncbi.nlm.nih.gov/geo/>

³<http://www.cbioportal.org/>

⁴<http://www.cistrome.org/>

⁵<http://www.gsea-msigdb.org/>



Assessment and Validation of the Risk Score Model

Due to the differences between the sequencing data and the chip data, the risk scores of the training set (TCGA) and the Validation sets (GSE55460, GSE70768, and MSKCC) were calculated, respectively, according to the above risk scoring calculation formula. Then, using the median value of their risk scores, we divided the training set and the Validation sets into high- and low-risk groups, respectively. We further used the survival data of the two groups of patients with PCa, to generate KM survival curves through a log-rank test (the model has predictive value when $P < 0.05$). Using the R package survivalROC 1.0.3 (Kamarudin et al., 2017), we conducted a time-dependent receiver operating characteristic (ROC) analysis to determine the accuracy of the prediction model. An area under the curve (AUC) of > 0.60 was considered to reflect minimum predictive value, while $AUC > 0.75$ showed good predictive value.

Enrichment Analysis of Prognostic Genes in PCa

We identified the differentially expressed genes (DEGs) in high- and low-risk groups from TCGA cohort using the R package limma 3.42.2. An absolute \log_2 fold change ($|FC| > 1$) and an adjusted false discovery rate (FDR) < 0.05 were set as cutoff criteria. Using the R package limma 3.42.2, 609 DEGs were subjected to Gene Ontology (GO) enrichment analysis. Furthermore, Gene Set Enrichment Analysis (GSEA) was conducted using GSEA 3.0, per the methods specified in the user guide (gsea-3.0.jar⁶).

Screening of Prognostic Clinical Factors in PCa

Using the R package survival 3.2-7, we performed univariate Cox regression analysis to appraise the predictive value of the

⁶<http://software.broadinstitute.org/gsea/index.jsp>

risk score and other clinical factors for DFS in patients with PCa. Then, we conducted multivariate Cox regression analysis to remove confounders. The bilateral significance level was 0.05, and the hazard ratio (HR) and 95% confidence interval (CI) were calculated.

Establishment and Evaluation of the Nomogram

Combining the prognostic factors analyzed using the above methods, we used the R package rms 6.1-0 to construct a nomogram. Using the R package pec v2019.11.3 (Mogensen et al., 2012), we performed Harrell's concordance index (C-index) analysis to assess the performance of the nomogram. A robust C-index was obtained using 1,000 bootstrap resamples. The closer the score was to 0.5, the less discriminative it was deemed to be, and vice versa for scores closer to 1. The efficiency of the nomogram was evaluated using the time-dependent ROC curve.

Drug Sensitivity Analysis

We used R package pRRophetic to compare half maximal inhibitory concentrations (IC50) of two drugs (docetaxel, bicalutamide) used in treatment of PCa, as per the Genomics of Drug Sensitivity in Cancer database. Results were recorded for, and compared between, the high- and low-risk groups. In addition, we used the Connectivity Map (CMap)⁷ to screen small molecule drugs that are beneficial to the prognosis of PCa. We obtained the *P* value of the association between each small molecule compound and the differential genes, by using the CMap online tool (Ma et al., 2021); *P* < 0.05 was statistically significant.

Statistical Analysis

All statistical analyses were performed using R software (version 3.6.3). The Mann-Whitney *U* test was used to compare two groups with non-normally distributed variables. Furthermore, Kruskal-Wallis analysis of variance was applied as non-parametric method to compare the three groups. All statistical tests were two-sided; *P* < 0.05 was statistically significant.

RESULTS

Identification of Prognosis-Associated eRNAs and Target Genes in PCa

Table 1 summarizes the important clinical factors from TCGA data involved in the study. Based on previous findings (Vučićević et al., 2015), we constructed a table of the relevant eRNAs and their predicted target genes. Through bootstrapping, we performed KM survival analysis on 1,584 eRNAs and 2,095 eRNA-driven genes. Finally, we identified 54 eRNAs and 199 target genes as prognosis-associated genes (Supplementary Table 1). With this screening method, randomness was significantly removed and the stability of results improved.

In order to verify that the Bootstrap method is more rational, we set six nodes (20, 50, 100, 250, 500, and 1,000) and compared

TABLE 1 | Clinical information from the 545 PCa patients of TCGA.

| Clinical parameters | Variable | Number (total) | Percentage | Hazard ratio |
|---------------------|----------|----------------|------------|---------------------|
| Age (years) | ≤60 | 242 | 44.40% | 1.338 (0.874–2.047) |
| | >60 | 303 | 55.60% | |
| Survival status | Dead | 10 | 1.83% | Not applicable |
| | Alive | 482 | 88.44% | |
| | Unknown | 53 | 9.72% | |
| Gleason score | 6 | 48 | 8.81% | 3.907 (2.471–6.178) |
| | 7 | 285 | 52.29% | |
| | 8 | 66 | 12.11% | |
| | 9 and 10 | 146 | 26.78% | |
| T grade | T2 | 188 | 34.50% | 4.018 (2.224–7.256) |
| | T3 | 295 | 54.13% | |
| | T4 | 10 | 1.83% | |
| | Unknown | 52 | 9.54% | |
| N grade | N0 | 348 | 63.85% | 1.820 (1.112–2.980) |
| | N1 | 79 | 14.50% | |
| | Unknown | 118 | 21.65% | |

the results at each node as shown in Figures 1B,C. Node 20 indicates the 20th time of Bootstrap, while node 50 indicates the 50th time, and so on. Taking node 20 as an example, we compared the result at the 20th-time Bootstrap with that of 10th time. For each the 10th and 20th times, we identified 66 genes for further screening and found that intersection for 62 genes when comparing the two nodes (we defined the intersection gene number as N). The unique gene number of the 10th time is defined as X, while that of the 20th time as Y. From these results, we defined the value of $N / (X + Y + N)$ as the consistency of the two results. As shown in Figures 1B,C, as the number of repeats increases, the consistency of data at the nodes increases to 100%, while the number of intersected genes gradually tends to be stable.

Acquisition of Candidate Genes and Labeling of Gene Function

Through comparative examinations, we screened 26 regulatory relationships and performed correlation analysis of these relationships ($\text{cor} > 0.3$, $P < 0.05$). We identified nine relationships consisting of six eRNAs and nine target genes (Table 2 and Supplementary Figure 1), from which we obtained nine candidate genes of which the functions were annotated using the Uniport database⁸ (Supplementary Table 2).

Building the Predictive Signature for PCa

The nine target genes were filtered using LASSO analysis. The filter condition was set to repeat 1,000 times (Figure 2A). Therefore, the number of candidate genes was narrowed down from nine to three. We used the three screened genes in the multivariate Cox regression analysis (without bidirectional stepwise regression) to establish a prognostic model for PCa (Table 3). For each of the three genes, the product of the correlation coefficient and gene expression

⁷<https://portals.broadinstitute.org/cmap/>

⁸<https://www.uniprot.org/>

TABLE 2 | List of eRNAs and eRNAs-driven genes associated with overall survival derived from enhancers.

| Ensembl ID | Symbol | TANRIC Overall Survival Analysis, Log-Rank p -Value | Target gene | Correlation between lncRNA and the Neighboring Target | |
|-----------------|------------|---|-------------|---|-------------------------|
| | | | | p -value | Correlation coefficient |
| ENSG00000223959 | AFG3L1P | 0.004207 | FANCA | $P < 0.001$ | 0.418196 |
| ENSG00000223959 | AFG3L1P | 0.004207 | MC1R | $P < 0.001$ | 0.70217 |
| ENSG00000223959 | AFG3L1P | 0.004207 | SPIRE2 | $P < 0.001$ | 0.556318 |
| ENSG00000161912 | ADCY10P1 | 4.72E-05 | UNC5CL | $P < 0.001$ | 0.64206 |
| ENSG00000206417 | H1FX-AS1 | 7.75E-06 | RPL32P3 | $P < 0.001$ | 0.372866 |
| ENSG00000242258 | LINC00996 | 0.002828 | GIMAP4 | $P < 0.001$ | 0.678696 |
| ENSG00000254812 | ACO67930.3 | 0.006284 | MAPK15 | $P < 0.001$ | 0.435871 |
| ENSG00000197558 | SSPOP | 0.000238 | ZNF467 | $P < 0.001$ | 0.603343 |
| ENSG00000227297 | SSPOP | 0.000863 | ZNF862 | $P < 0.001$ | 0.68106 |

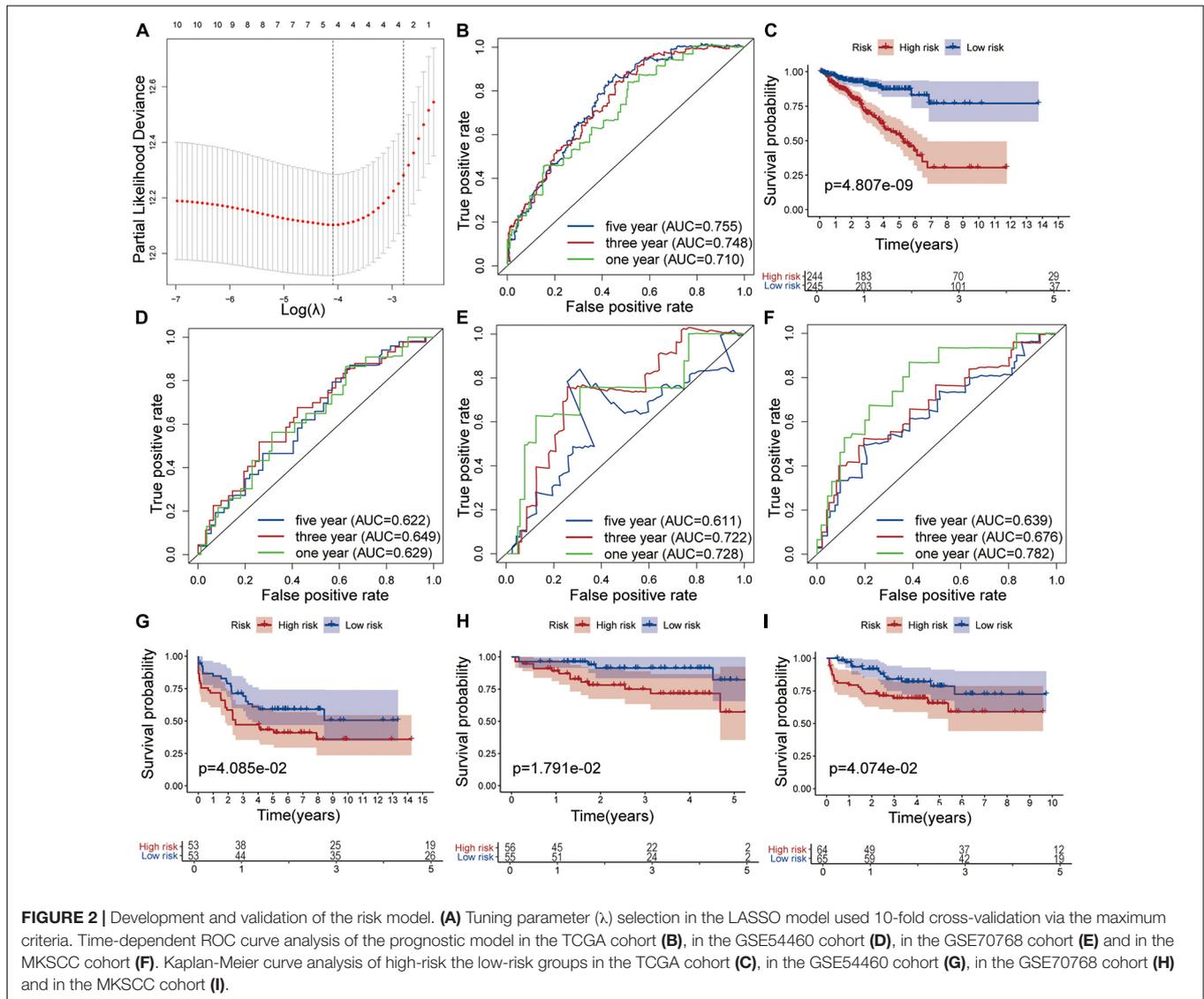


TABLE 3 | Risk genes in the prognostic risk model.

| Gene | Coef | HR | HR.95L | HR.95H | p value |
|--------|----------|----------|----------|----------|----------|
| MAPK15 | 0.186338 | 1.204829 | 0.977376 | 1.485215 | 0.080882 |
| ZNF467 | 0.523366 | 1.687699 | 1.345957 | 2.116209 | 5.80E-06 |
| MC1R | 0.624495 | 1.867304 | 1.332502 | 2.616749 | 0.000286 |

was used to determine the relevant risk score. Thus, we established a risk score model that calculated the risk score as follows: risk score = (0.186338 × *MAPK15* gene expression) + (0.523366 × *ZNF467* gene expression) + (0.624495 × *MC1R* gene expression). Correlation analysis between CNVs of model genes and its corresponding eRNAs showed no significant correlation among them (cor < 0.3) (Supplementary Figures 3A–C).

Evaluation of the Prognostic Risk Score Model for PCa

The AUC values for 1-, 3-, and 5-year survival were 0.710, 0.748, and 0.755, respectively (Figure 2B). Using the median risk score (0.951), we distributed the 489 samples with complete clinical data in TCGA cohort into a high-risk group consisting of 244 samples and a low-risk group of 245. The DFS of the low-risk group was significantly higher than that of the high-risk group (log-rank test, $P < 0.05$, Figure 2C).

We also analyzed the performance of individual model genes in the training set and showed that the AUC values for 1-, 3-, and 5-year survival of *MAPK15* were 0.671, 0.718, and 0.731, respectively (Supplementary Figure 4A), while those of *ZNF467* were 0.631, 0.715, and 0.719, respectively (Supplementary Figure 4B), and of *MC1R* were 0.669, 0.634, and 0.626, respectively (Supplementary Figure 4C). No significant correlation between the three genes was observed (Supplementary Figure 4D). The results suggest that the predictive value of the model is better than that of a single gene.

Validation of the Prognostic Risk Score Model for PCa

The prognostic risk model was further validated in the GEO datasets GSE54460, GSE70768, and MSKCC. The clinical information of the validation cohort is summarized in Table 4. As shown in Figure 2D, the prognostic risk score model performed well in GSE54460, and the AUC values for 1-, 3-, and 5-year survival were 0.629, 0.649, and 0.622, respectively. Using the median risk score for segmentation (0.502), we distributed the 106 samples into high- and low-risk groups, containing 53 samples each. This result was similar to that obtained in the KM analysis of the training cohort (Figure 2G). In GSE70768, the respective AUC values for 1-, 3-, and 5-year survival were 0.728, 0.722, and 0.611, respectively (Figure 2E). Similarly, using the median risk score for segmentation (9.773), we divided the 111 samples into a low-risk group of 55 and a high-risk group of 56. In the KM analysis of GSE70768, the results of the low-risk group were significantly different from those of the high-risk group (log-rank test, $P < 0.05$, Figure 2H). In MSKCC, the

AUC values for 1-, 3-, and 5-year survival were 0.782, 0.676, and 0.639, respectively. Using the median risk score (4.210) as discriminator, we divided the 129 samples into two groups of which 64 were deemed as low-risk, and 65 as high-risk. In the KM analysis of MSKCC, the results of the low-risk group were also significantly different from those of the high-risk group (log-rank test, $P < 0.05$, Figures 2F,I).

To explore the single predictive value of the three-gene signature, we visualized the differential expression of each gene in samples with different clinical features. These included normal tissue and tumor samples, samples with T- and N-category characteristics, and samples differentiated according to their Gleason score and risk (TCGA dataset). Among these groups, those involving tumor samples, specifically category T3/T4 tumors, with N1 nodal status, high-risk, and Gleason score > 7 exhibited the highest gene expression levels (Figures 3A–E).

Establishment and Appraisal of a Nomogram for DFS Prediction in PCa

We screened clinical variables with independent prognostic value using univariate and multivariate Cox regression analyses (Supplementary Table 3) and applied the screening results based on the Gleason score, T-stage, and risk score to construct a nomogram (Supplementary Figure 3D). We then assessed the clinical significance of the nomogram using ROC analysis and C-index calculations. The respective AUC values of the nomogram for 1-, 3-, and 5-year survival were 0.780, 0.783, and 0.777, respectively. Compared to the individual clinical variables, the nomogram exhibited superior predictive ability, suggesting that the clinical nomogram offered better predictive value in the prognosis of patients with PCa than the individual clinical variables or risk score models (Supplementary Figures 5A–H).

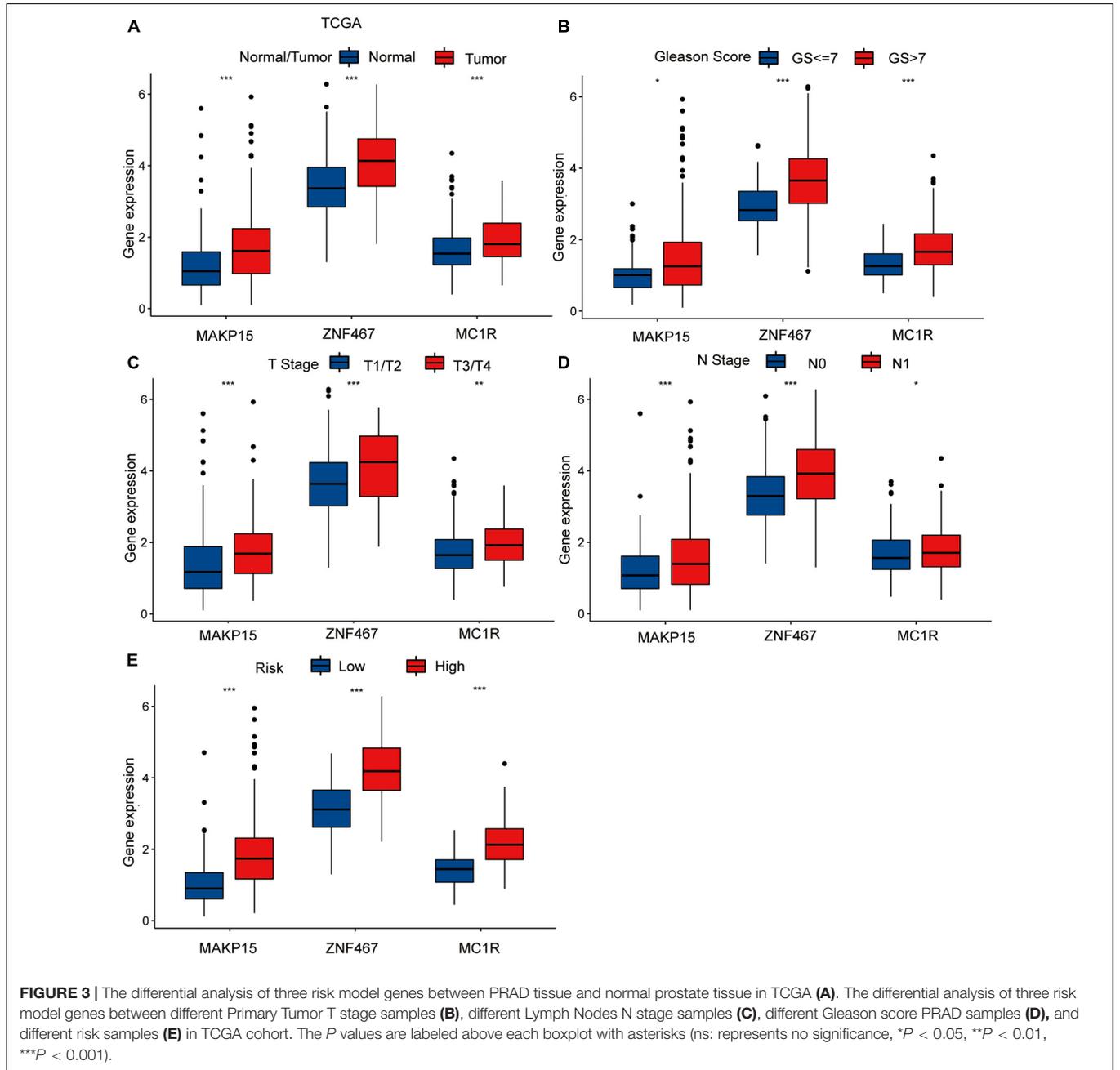
Exploration of the Functions of eRNAs and Target Genes

By comparing the gene expression levels in samples from the high- and low-risk groups of the training cohort, 609 DEGs ($|FC| \geq 1$, adjusted $P < 0.05$) were further analyzed. GO enrichment analysis was used to explore the functions of DEGs in high- and low-risk groups, and 130 GO terms were obtained ($P < 0.05$). The results showed that the GO functions with significant DEG enrichment were “humoral immune response,” “complement activation, classical pathway,” “humoral immune response mediated by circulating immunoglobulin,” “immunoglobulin mediated immune response,” “immunoglobulin complex,” and “B cell mediated immunity” ($P < 0.01$) (Supplementary Figure 6C and Supplementary Table 4).

GSEA was used to explore the potential signaling pathways that influence the risk score model in the high- and low-risk groups ($P < 0.05$). In the high-risk group, the highest-ranking signaling pathways were base excision repair, DNA replication, Drug metabolism—other enzymes, homologous recombination, and spliceosome. In the low-risk group, the highest-ranking signaling pathways were adherens junction, propanoate metabolism, sphingolipid metabolism, tryptophan

TABLE 4 | Clinical recurrence rate in TCGA, GES54460, GES70768, and MKSCC database.

| Dataset | Clinical parameter | | | | | |
|----------------------------|--------------------|------------|------------|---------------|-------------|-------------|
| | Recurrence | | | No recurrence | | |
| | 1 year | 3 years | 5 years | 1 year | 3 years | 5 years |
| TCGA(<i>n</i> = 489) | 29(5.93%) | 67(17.30%) | 81(16.56%) | 460(94.07%) | 422(82.70%) | 408(83.44%) |
| GES54460 (<i>n</i> = 106) | 24(22.64%) | 45(42.45%) | 51(48.11%) | 82(77.36) | 61(57.55%) | 55(51.89%) |
| GES70768 (<i>n</i> = 111) | 8(7.20%) | 16(14.41%) | 19(17.11%) | 103(92.79%) | 95(85.59%) | 92(82.89%) |
| MKSCC (<i>n</i> = 129) | 15(11.63%) | 28(21.71%) | 31(24.03%) | 114(88.37%) | 101(78.29%) | 98(75.97%) |



metabolism, and valine, leucine, and isoleucine degradation (**Supplementary Table 5** and **Supplementary Figures 6A,B**).

Comparison of TMEs Between High- and Low-Risk Groups

Based on the GO enrichment analysis results, the genes included in the risk model may affect the occurrence and development of tumors, through their impact on certain immune-related functions such as immunoglobulin complex. Therefore, we compared the tumor microenvironments (TMEs) in the high- and low-risk groups.

We calculated the single-sample GSEA (ssGSEA) scores of EMT, TGF- β , and ECM to compare the differences between the high- and low-risk groups, based on stromal cells and showed that the ECM and TGF- β scores of the low-risk group were higher than those of the high-risk group. No significant difference was seen in the EMT score between the high- and low-risk groups (**Figure 4A**).

We further explored immune cell infiltration to determine the immune status of the high- and low-risk groups and showed that CD4⁺ T cell and CD8⁺ T cell counts differed significantly between the two groups (**Figure 4B**). In addition to the TIMER algorithm, we used epic and xcell algorithms to compare the degree of immune infiltration in the two groups. The results of the EPIC algorithm showed that cancer associated fibroblast, T cell CD4+, T cell CD8+, macrophage, and uncharacterized cells were significantly different (**Supplementary Figure 7A**). The results of the xcell algorithm show that T cell CD4+ central memory, T cell CD4+ effector memory, T cell CD8+, T cell CD8+ effector memory, eosinophil, cancer associated fibroblast, neutrophil, T cell NK, plasmacytoid dendritic cell, immune score, and stroma scores were significantly different between the high and low risk groups (**Supplementary Figure 7B**).

We also compared the levels of expression of 14 immune checkpoints between the high- and low-risk groups and showed that the expression of nine of these immune checkpoints (LAG3, CD80, CD276, PDCD1, IL6, CD4, TGFB1, CD86, CTLA4) were significantly different between the high- and low-risk groups (**Figure 4C**). All except IL6 showed higher expression in the high-risk group than in low-risk group.

Comparison of the DNA Mutation Burden, TMB, and MSI Between High- and Low-Risk Groups

Based on the GSEA results, base excision repair was associated with the heterogeneity between the high- and low-risk groups. We then explored the differences in the DNA mutation burden, tumor mutational burden (TMB), and microsatellite instability (MSI), between the two groups.

By analyzing the 20 genes with the highest mutation frequency, we found a marked difference in mutation burden between the high- and low-risk groups. The genes with the highest mutational frequency in the high- and low-risk groups, were TP53 and SPOP, respectively (**Figure 5A**).

Comparing the differences in TMB between the high- and low-risk groups showed that the TMB of the high-risk group was

significantly different to that of the low-risk group ($P < 0.05$), and the sensitivity of the low-risk group to immunotherapy was lower than that of the high-risk group (**Figure 5B**).

Differential analysis of the MSI of each PRAD sample obtained from a previous study (Zhang et al., 2020) showed that the low-risk group had a lower level of MSI than the high-risk group ($P < 0.05$, **Figure 5C**). We considered this to be MSI high in the high-risk group and MSI low in the low-risk group, due to the tumor's disrupted function during DNA damage repair which increased gene instability. This is consistent with the results of TMB in the high and low risk groups.

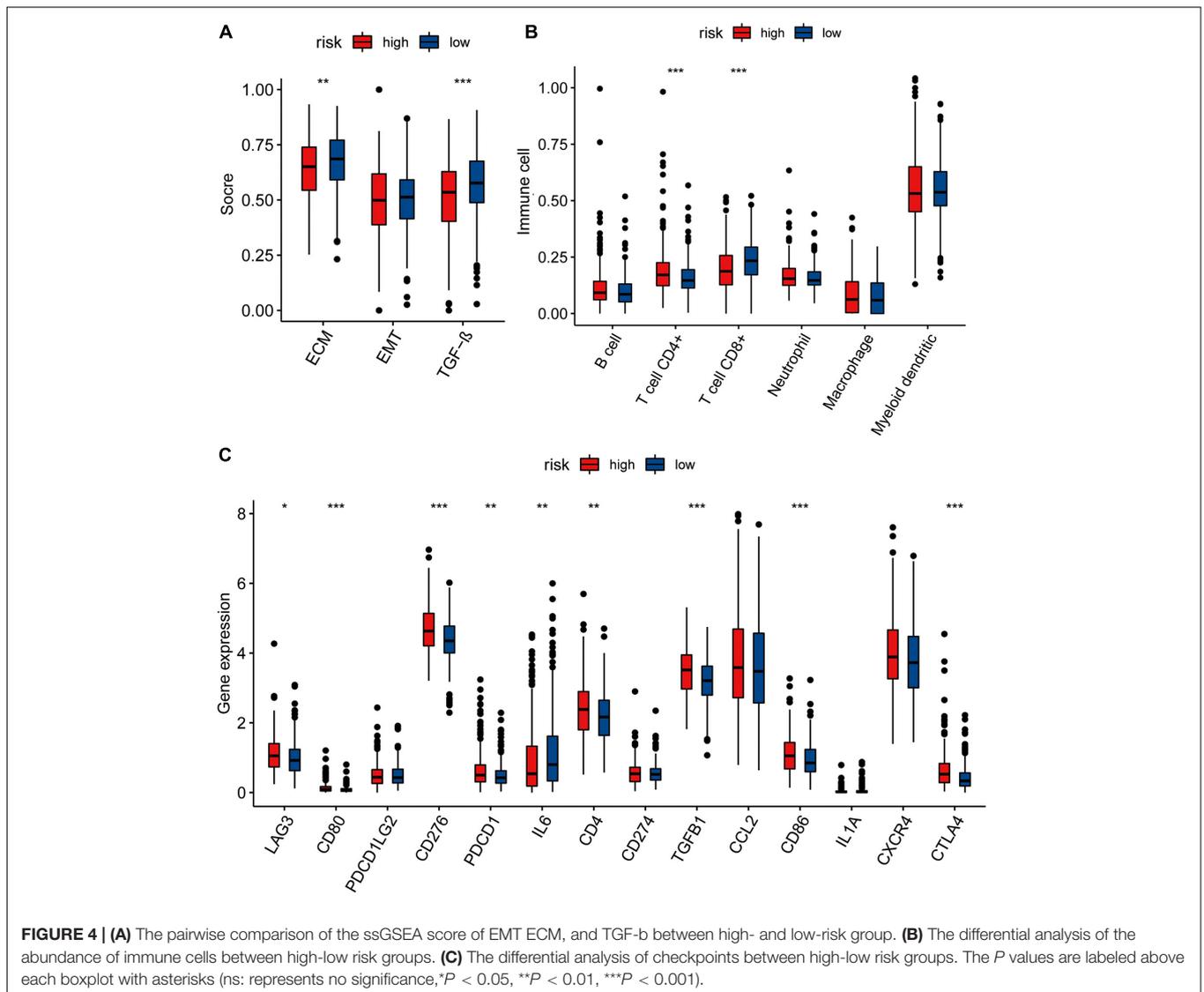
Drug Sensitivity Analysis

We assessed differences in drug sensitivity between the high- and low-risk groups by analyzing the IC50 of two chemotherapeutic agents and showed that, among the drugs for which significantly different sensitivities were perceived between the high- and low-risk groups, patients in the high-risk group were more sensitive to docetaxel. Conversely, patients in the low-risk group were comparatively more sensitive to bicalutamide (**Figures 6A,B**). Through online CMap analysis, we identified 10 of the most significant small molecule candidate drugs that could improve prognosis in PCa, of which megestrol and tiletamine were the most likely to favorably alter gene expression (**Supplementary Table 6**).

DISCUSSION

Genomic and epigenomic changes in tumor cells are associated with certain tumor factors, including cellular proliferation and oncogenic transformation (Rhie et al., 2016). In recent years, several studies have illustrated the importance of lncRNAs (including eRNAs) in regulating gene transcription and protein synthesis (Ouyang et al., 2014; Engreitz et al., 2016). Compared with other types of lncRNA, eRNAs are ideal anticancer targets, as they affect the occurrence of many cancers by regulating the expression of multiple genes (van Velzen et al., 2020). However, few reports have described the influence of eRNAs on the development of PCa. Therefore, we identified eRNAs and eRNA-driven genes related to the prognosis of patients with PCa, to explore the possible mechanisms of these eRNAs' influence on the occurrence and development of PCa. Our model showed positive values for the coefficients of the relevant three genes, indicating that these genes are promoters of PCa. Thus, we inferred a direct correlation between expression of the signature genes and risk of poor prognosis in PCa.

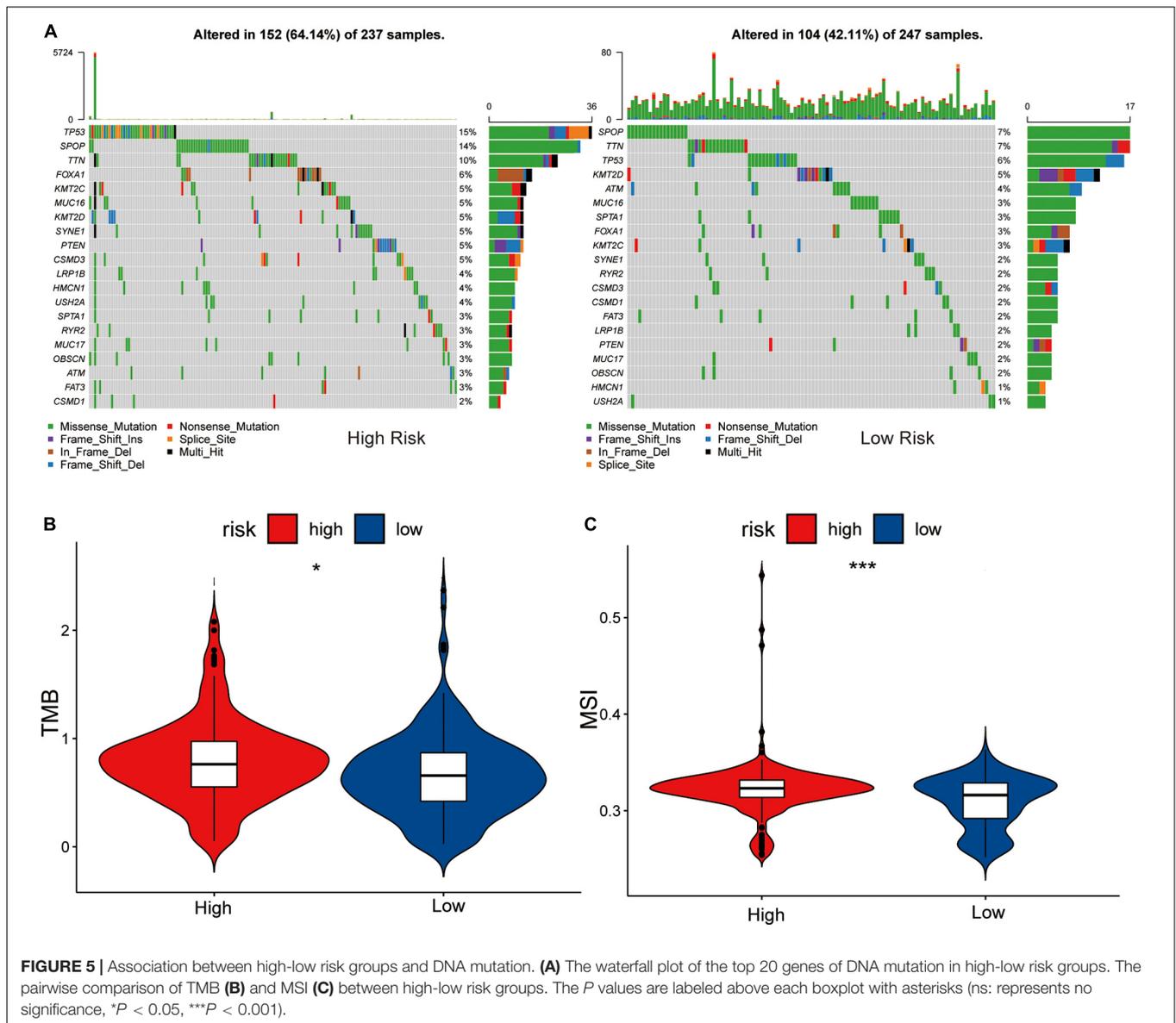
Through functional analysis, we found that the eRNAs and corresponding eRNA-driven genes may affect tumor development through certain immune processes. Previous studies have shown that immune cells are important in hindering tumor progression, and that the TME plays a key role in the occurrence and development of PCa and other cancers (Fortis et al., 2017; Togashi et al., 2019). For example, natural killer cells are regulated by the TME in the process of killing lung cancer tumor cells (Ren et al., 2015). Previous studies have also shown that ECM, TGF- β , and EMT are important factors



affecting the TME and, resultantly, tumor progression (Park et al., 2019; Yang et al., 2020). In our study, EMT played no significant role in PCa progression. Fibrogenesis of ECM can promote tumor development (Barry et al., 2020). In the high-risk group of our study, this process might have been exaggerated, resulting in a poor prognosis. The TGF- β signaling pathway can inhibit tumor development by inducing the expression of multiple tumor-suppressor genes (Papoutsoglou and Moustakas, 2020). In this study, the function of the tumor-suppressor arm of the TGF- β signaling pathway might have been disabled in the high-risk group, thereby altering gene expression in a direction that is favorable to tumor development. Our study also revealed that the levels of CD8⁺ T cells were decreased in the high-risk group, whereas levels of CD4⁺ T cells were increased in the low-risk group. This suggests that patients in the high-risk group were in a state of immunodepletion and that the three genes of our model may promote tumor progression by accelerating immunodepletion. The high expression of immune

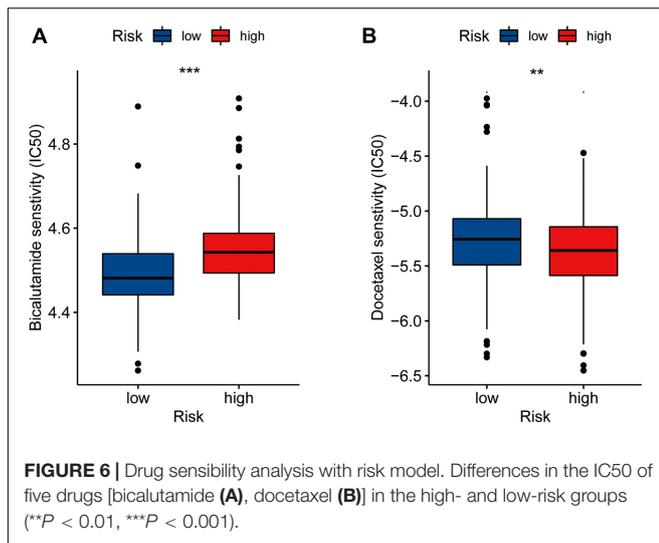
checkpoint genes is an important mechanism of immune evasion in tumorigenesis (Zhang and Zheng, 2020). In our study, the levels of expression of most immune checkpoint genes were lower in the low-risk group than in the high-risk group, which may partially explain the poor prognosis of patients in the high-risk group. Thus, we can surmise that the three signature genes we have identified in this study may contribute to cancer development by increasing immune checkpoint expression.

The results of the functional analysis indicated that the eRNAs and corresponding eRNA-driven genes also influence the progression of PCa by affecting the base excision repair process. We compared the tumor mutation spectrum, TMB, and MSI between the high- and low-risk groups and showed that the gene with the highest mutation frequency in the high-risk group was TP53, whereas in the low-risk group it was SPOP. TP53 mutations, which can significantly promote tumor growth, are sequentially associated with the development of PCa (Kaur et al., 2019), while SPOP is an important tumor-suppressor gene



in PCa cells (Geng et al., 2017). Our results also showed that the high expression levels of eRNAs and corresponding eRNA-driven genes were associated with PCa prognosis, and could promote occurrence of TP53 mutations, which further aids the progression of PCa. It has been shown that MSI is associated with many genetic diseases and is a key indicator of genomic instability (Niu et al., 2014). TMB, defined as the total number of somatic gene coding errors, base substitutions, and gene insertion or deletion errors detected per million bases, is closely related to tumor heterogeneity (Merino et al., 2020; Salem et al., 2020). The GSEA results showed that the scores of TMB and MSI in the high-risk group were higher than those in the low-risk group, suggesting that the eRNAs and corresponding eRNA-driven genes could affect the process of base excision repair. We considered this to be MSI high in the high-risk group and MSI low in the low-risk group, due to the tumor's disrupted

function during DNA damage repair leading to increased gene instability, which is consistent with the results of TMB in the high and low risk groups. Jiang et al. (2021) found that patients with high TMB had worse survival outcomes than those with low TMB. We believe that this finding may account for the longer DFS in the low-risk group, compared to the high-risk group. In addition, TMB and MSI are closely associated with the efficacy of immunotherapy and can be used as biomarkers of immunotherapy response in PCa (Merino et al., 2020; van Velzen et al., 2020). Although PCa is an immunodesert tumor, immunity has a significant influence (Bilusic et al., 2017). Moreover, our GO enrichment results also showed that the DEGs in the high and low risk groups were related to immunity, which could explain the immune desertification of PCa. In our study, the high-risk group exhibited high immune cell infiltration and high immune checkpoint expression, suggesting that patients



in this group could significant benefit from immune-targeted therapy. This finding was consistent with our results related to TMB and MSI. Therefore, our calculated risk scores were highly correlated with the prognosis of patients with PCa and the efficacy of immunotherapy. At present, drug therapy is used to a clinical setting for the treatment of PCa, mainly docetaxel and bicalutamide. The results of drug sensitivity analysis have shown that our model, based on the eRNA-driven genes, can better predict the drug sensitivity of patients with PCa, which could guide the selection of appropriate clinical drugs to a certain extent.

Based on the outcomes of our study, our method of constructing a risk score model is more reliable in producing a robust model than existing methods, and that our study was well-founded as the mechanisms by which eRNAs affect the occurrence and development of PCa, as well as their involvement in potential therapeutic options, were explored on many levels and through various methods. However, our study has some limitations. Firstly, the data were not from our own database but rather retrieved from public databases. Secondly, the mechanisms of action of the eRNAs and target genes in PCa, were not further investigated through *in vitro* and *in vivo* experiments.

CONCLUSION

In summary, we used expression profiles from public databases to screen genes closely associated with the prognosis of patients with PCa and established a prognostic risk model, which was validated in multiple datasets. Differential analyses showed significant differences in TME, TMB, and MSI between the high- and low-risk groups. The genes included in the three-gene signature may affect tumor progression, by impacting the aforementioned factors. Therefore, the risk score derived from our model is a good predictor of the prognosis of patients with PCa and can be used to explore how eRNAs affect the occurrence and development of PCa. Drug sensitivity analysis showed that respondents in the high- and low-risk groups differed in their sensitivity to several

chemotherapeutic agents used to treat PCa. Our model could predict drug sensitivity, and several candidate drugs that may inhibit the progression of PCa were identified. These findings can help to design personalized medication regimes for patients and exploring more effective drug options for PCa treatment.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

AUTHOR CONTRIBUTIONS

SF, ZW, and LZ designed the study, analyzed the data, revised the images, performed the literature search, and collected data for the manuscript. CZ, JW, and DY revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This article was supported by National Natural Science Foundation of China (81800668) and Natural Science Foundation of Shanxi province (201801D221276).

ACKNOWLEDGMENTS

The authors would like to thank all the patients who participated in the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.676845/full#supplementary-material>

Supplementary Figure 1 | Expression relationships of six eRNAs and corresponding eRNAs-driven Genes.

Supplementary Figure 2 | KM survival curves for six eRNAs and nine target genes.

Supplementary Figure 3 | Correlation analysis of three model genes' copy number variation regions with target genes [MAPK15 (A), ZNF467 (B), and MC1R (C)].

Supplementary Figure 4 | Time-dependent ROC curve analysis of the model genes [MAPK15 (A), MC1R (B), and ZNF467 (C)] in the TCGA cohort. Correlation analysis of three genes in the model (D).

Supplementary Figure 5 | The correlation of riskscore and Gleason score (≤ 7 , > 7) (A), Lymph nodes (N0, N1) (B), and T-stage (T1&T2, T3&T4) (C) expression. Nomogram to predict the 1-, 3- and 5-year DFS of PCa patients (D). Concordance index of the indicated prognostic model in the training datasets (E). Time-dependent ROC analysis was used to evaluate the accuracy of the DFS nomograms (F-H). The red, blue, green, and orange solid lines represent the T-stage, gleason-score, risk-score, and total-points, respectively.

Supplementary Figure 6 | The rows and columns indicate the genes and tumor samples, respectively. Enrichment plots of the top five KEGG pathways in the high-risk score (A) and low-risk score (B) groups in PRAD. GO analysis of differential expression genes in high-low risk group (C).

Supplementary Figure 7 | Analysis of the difference of immune infiltration between the two algorithms of epic (A) and xcell (B) in the high- and low-risk groups.

Supplementary Table 1 | In 1,000 iterations, eRNAs and target genes appeared more than or equal to 700 times.

Supplementary Table 2 | Functional markers of nine eRNAs-driven genes.

Supplementary Table 3 | Clinical factors were analyzed by univariate and multivariate Cox.

Supplementary Table 4 | 130 GO terms of differential expression genes.

Supplementary Table 5 | GSEA results of the KEGG pathways in the high-risk score and low-risk score groups in PRAD.

Supplementary Table 6 | 10 Important Small Molecule Drugs that may reverse tumor status of PCA.

REFERENCES

- Albertsen, P. C. (2020). Re: radical Prostatectomy or Watchful Waiting in Prostate Cancer-29-Year Follow-up. *Eur. Urol.* 78:471. doi: 10.1016/j.eururo.2020.04.017
- Barry, A., Baldeosingh, R., Lamm, R., Patel, K., Zhang, K., Dominguez, D., et al. (2020). Hepatic Stellate Cells and Hepatocarcinogenesis. *Front. Cell Dev. Biol.* 8:709. doi: 10.3389/fcell.2020.00709
- Bilusic, M., Madan, R. A., and Gulley, J. L. (2017). Immunotherapy of Prostate Cancer: facts and Hopes. *Clin. Cancer Res.* 23, 6764–6770. doi: 10.1158/1078-0432.ccr-17-0019
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B., Sumer, S., Aksoy, B., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2, 401–404. doi: 10.1158/2159-8290.cd-12-0095
- Connor, M., Shah, T., Smigielska, K., Day, E., Sukumar, J., Fiorentino, F., et al. (2021). Additional Treatments to the Local tumour for metastatic prostate cancer-Assessment of Novel Treatment Algorithms (IP2-ATLANTA): protocol for a multicentre, phase II randomised controlled trial. *BMJ Open* 11:e042953. doi: 10.1136/bmjopen-2020-042953
- Ding, M., Liu, Y., Liao, X., Zhan, H., Liu, Y., and Huang, W. (2018). Enhancer RNAs (eRNAs): new Insights into Gene Transcription and Disease Treatment. *J. Cancer* 9, 2334–2340. doi: 10.7150/jca.25829
- Engreitz, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., et al. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* 539, 452–455. doi: 10.1038/nature20149
- Fortis, S., Sofopoulos, M., Sotiriadou, N., Haritos, C., Vaxevanis, C., Anastasopoulou, E., et al. (2017). Differential intratumoral distributions of CD8 and CD163 immune cells as prognostic biomarkers in breast cancer. *J. Immunother. Cancer* 5:39. doi: 10.1186/s40425-017-0248-z
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* 33, 1–22. doi: 10.18637/jss.v033.i01
- Geng, C., Kaochar, S., Li, M., Rajapakshe, K., Fiskus, W., Dong, J., et al. (2017). SPOP regulates prostate epithelial cell proliferation and promotes ubiquitination and turnover of c-MYC oncoprotein. *Oncogene* 36, 4767–4777. doi: 10.1038/onc.2017.80
- Jiang, A., Ren, M., Liu, N., Gao, H., Wang, J., Zheng, X., et al. (2021). Tumor Mutation Burden, Immune Cell Infiltration, and Construction of Immune-Related Genes Prognostic Model in Head and Neck Cancer. *Int. J. Med. Sci.* 18, 226–238. doi: 10.7150/ijms.51064
- Kamarudin, A., Cox, T., and Kolamunnage-Dona, R. (2017). Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC Med. Res. Methodol.* 17:53. doi: 10.1186/s12874-017-0332-6
- Kaur, H., Lu, J., Guedes, L., Maldonado, L., Reitz, L., Barber, J., et al. (2019). TP53 missense mutation is associated with increased tumor-infiltrating T cells in primary prostate cancer. *Hum. Pathol.* 87, 95–102. doi: 10.1016/j.humpath.2019.02.006
- Liang, J., Zhou, H., Gerd, C., Tan, M., Colson, T., Kaye, K. M., et al. (2016). Epstein-Barr virus super-enhancer eRNAs are essential for MYC oncogene expression and lymphoblast proliferation. *Proc. Natl. Acad. Sci. U. S. A.* 113, 14121–14126. doi: 10.1073/pnas.1616697113
- Ma, J., Cai, X., Kang, L., Chen, S., and Liu, H. (2021). Identification of novel biomarkers and candidate small-molecule drugs in cutaneous melanoma by comprehensive gene microarrays analysis. *J. Cancer* 12, 1307–1317. doi: 10.7150/jca.49702
- Merino, D. M., McShane, L. M., Fabrizio, D., Funari, V., Chen, S. J., White, J. R., et al. (2020). Establishing guidelines to harmonize tumor mutational burden (TMB): in silico assessment of variation in TMB quantification across diagnostic platforms: phase I of the Friends of Cancer Research TMB Harmonization Project. *J. Immunother. Cancer* 8:e000147. doi: 10.1136/jitc-2019-000147
- Mogensen, U., Ishwaran, H., and Gerds, T. (2012). Evaluating Random Forests for Survival Analysis using Prediction Error Curves. *J. Stat. Softw.* 50, 1–23. doi: 10.18637/jss.v050.i11
- Niu, B., Ye, K., Zhang, Q., Lu, C., Xie, M., McLellan, M. D., et al. (2014). MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* 30, 1015–1016. doi: 10.1093/bioinformatics/btt755
- No Authors Listed. (2020). Erratum: global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 70:313. doi: 10.3322/caac.21609
- Ouyang, J., Zhu, X., Chen, Y., Wei, H., Chen, Q., Chi, X., et al. (2014). NRAV, a long noncoding RNA, modulates antiviral responses through suppression of interferon-stimulated gene transcription. *Cell Host Microbe* 16, 616–626. doi: 10.1016/j.chom.2014.10.001
- Papoutsoglou, P., and Moustakas, A. (2020). Long non-coding RNAs and TGF- β signaling in cancer. *Cancer Sci.* 111, 2672–2681. doi: 10.1111/cas.14509
- Park, J., Kim, D. H., Shah, S. R., Kim, H. N., Kshitiz, Kim, P., et al. (2019). Switch-like enhancement of epithelial-mesenchymal transition by YAP through feedback regulation of WT1 and Rho-family GTPases. *Nat. Commun.* 10:2797. doi: 10.1038/s41467-019-10729-5
- Ren, J., Nie, Y., Lv, M., Shen, S., Tang, R., Xu, Y., et al. (2015). Estrogen upregulates MICA/B expression in human non-small cell lung cancer through the regulation of ADAM17. *Cell. Mol. Immunol.* 12, 768–776. doi: 10.1038/cmi.2014.101
- Rhie, S. K., Guo, Y., Tak, Y. G., Yao, L., Shen, H., Coetzee, G. A., et al. (2016). Identification of activated enhancers and linked transcription factors in breast, prostate, and kidney tumors by tracing enhancer networks using epigenetic traits. *Epigenetics Chromatin* 9:50. doi: 10.1186/s13072-016-0102-4
- Salem, M., Bodor, J., Puccini, A., Xiu, J., Goldberg, R., Grothey, A., et al. (2020). Relationship between MLH1, PMS2, MSH2 and MSH6 gene-specific alterations and tumor mutational burden in 1057 microsatellite instability-high solid tumors. *Int. J. Cancer* 147, 2948–2956. doi: 10.1002/ijc.33115
- Sedano, M. J., Harrison, A. L., Zilaie, M., Das, C., Choudhari, R., Ramos, E., et al. (2020). Emerging Roles of Estrogen-Regulated Enhancer and Long Non-Coding RNAs. *Int. J. Mol. Sci.* 21:3711. doi: 10.3390/ijms21103711
- Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2011). Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J. Stat. Softw.* 39, 1–13. doi: 10.18637/jss.v039.i05
- Togashi, Y., Shitara, K., and Nishikawa, H. (2019). Regulatory T cells in cancer immunosuppression - implications for anticancer therapy. *Nat. Rev. Clin. Oncol.* 16, 356–371. doi: 10.1038/s41571-019-0175-7
- Van den Broeck, T., van den Bergh, R. C. N., Arfi, N., Gross, T., Moris, L., Briens, E., et al. (2019). Prognostic Value of Biochemical Recurrence Following Treatment with Curative Intent for Prostate Cancer: a Systematic Review. *Eur. Urol.* 75, 967–987. doi: 10.1016/j.eururo.2018.10.011
- van Velzen, M. J. M., Derks, S., van Grieken, N. C. T., Haj Mohammad, N., and van Laarhoven, H. W. M. (2020). MSI as a predictive factor for treatment outcome of gastroesophageal adenocarcinoma. *Cancer Treat. Rev.* 86:102024. doi: 10.1016/j.ctrv.2020.102024

- Vučičević, D., Corradin, O., Ntini, E., Scacheri, P. C., and Ørom, U. A. (2015). Long ncRNA expression associates with tissue-specific enhancers. *Cell Cycle* 14, 253–260. doi: 10.4161/15384101.2014.977641
- Wang, S., Somisetty, V. S., Bai, B., Chernukhin, I., Niskanen, H., Kaikkonen, M. U., et al. (2020). The proapoptotic gene interferon regulatory factor-1 mediates the antiproliferative outcome of paired box 2 gene and tamoxifen. *Oncogene* 39, 6300–6312. doi: 10.1038/s41388-020-01435-4
- Yang, J., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., et al. (2020). Guidelines and definitions for research on epithelial-mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* 21, 341–352. doi: 10.1038/s41580-020-0237-9
- Zhang, Y., Zhang, R., Liang, F., Zhang, L., and Liang, X. (2020). Identification of Metabolism-Associated Prostate Cancer Subtypes and Construction of a Prognostic Risk Model. *Front. Oncol.* 10:598801. doi: 10.3389/fonc.2020.598801
- Zhang, Y., and Zheng, J. (2020). Functions of Immune Checkpoint Molecules Beyond Immune Evasion. *Adv. Exp. Med. Biol.* 1248, 201–226. doi: 10.1007/978-981-15-3266-5_9
- Zhang, Z., Lee, J. H., Ruan, H., Ye, Y., Krakowiak, J., Hu, Q., et al. (2019). Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer. *Nat. Commun.* 10:4562. doi: 10.1038/s41467-019-12543-5
- Zhou, R., Zeng, D., Zhang, J., Sun, H., Wu, J., Li, N., et al. (2019). A robust panel based on tumour microenvironment genes for prognostic prediction and tailoring therapies in stage I-III colon cancer. *EBioMedicine* 42, 420–430. doi: 10.1016/j.ebiom.2019.03.043

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Fan, Wang, Zhao, Zhao, Yuan and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.