Identification of Potential Predictor of Biochemical Recurrence in Prostate Cancer

Jingchao Wei, Xiaohang Wu, Yuxiang Li, Xiaowu Tao, Bo Wang, Guangming Yin

Department of Urology, The Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, People’s Republic of China; Department of Urology, The Third Xiangya Hospital of Central South University, Changsha, People’s Republic of China

Correspondence: Guangming Yin, Email yingm75@126.com

Background: Prostate cancer is a common malignancy in men. Radical prostatectomy is one of the primary treatment modalities for patients with prostate cancer. However, early identification of biochemical recurrence is a major challenge for post-radical prostatectomy surveillance. There is a lack of reliable predictors of biochemical recurrence. The purpose of this study was to explore potential biochemical recurrence indicators for prostate cancer.

Materials and Methods: We analyzed transcriptomic data of cases with biochemical recurrence in The Cancer Genome Atlas (TCGA). Then, we performed integrative bioinformatics analyses to establish a biochemical recurrence predictor model of prostate cancer.

Results: There were 146 differentially expressed genes (DEGs) between prostate cancer and normal prostate, including 12 upregulated and 134 downregulated genes. Comprehensive pathway enrichment analyses revealed that these DEGs were associated with multiple cellular metabolic pathways. Subsequently, according to the random assignment principle, 208 patients were assigned to the training cohort and 205 patients to the validation cohort. Univariate Cox regression analysis showed that 7 genes were significantly associated with the biochemical recurrence of prostate cancer. A model consisting of 5 genes was constructed using LASSO regression and multivariate Cox regression to predict biochemical recurrence of prostate cancer. Expression of PAH and AOC1 decreased with an increasing incidence of prostate cancer, whereas expression of DDC, LINC01436 and ORM1 increased with increasing incidence of prostate cancer. Kaplan–Meier curves and receiver operating characteristic (ROC) curves indicated that the 5-gene model had reliable utility in identifying the risk of biochemical recurrence of prostate cancer.

Conclusion: This study provides a model for predicting prostate cancer recurrence after surgery, which may be an optional indicator for postoperative follow-up.

Keywords: prostate cancer, biochemical recurrence, predictor, signature, follow-up

Introduction

Prostate cancer is a considerable health threat to middle-aged and older men, and its mortality is ranked fifth among all tumors in males. It is well known that the growth and progression of prostate cancer depend on androgen. Androgen deprivation therapy is an effective treatment strategy and is widely used for treating prostate cancer. Also, there has been increasing interest in the role of localized therapies in recent years. However, the biological characteristics of prostate cancer could be either indolent or highly aggressive. Therefore, biochemical recurrence (BCR) is a leading concern for patients after radical prostatectomy or androgen deprivation therapy, as it often represents a progression of the cancer to a stage that is more difficult to be treated. The incidence of prostate cancer biochemical recurrence is around 15% to 30% within 5 years and about 40% within 10 years after radical prostatectomy. Once biochemical recurrence occurs, many prostate cancer patients are inclined to progress to be resistant to androgen deprivation therapy, which is also called castration-resistant prostate cancer (CRPC).

CRPC usually leads to death within 2 to 4 years. Massive efforts have resulted in new biomarkers for early detection and prognosis of prostate cancer, but few reliable predictive markers of prostate cancer progression. Multiple factors, including the Gleason score, pathological clinic staging, prostate-specific antigen (PSA), and surgical margin, fail to predict biochemical recurrence accurately. PSA is a vital biomarker of prostate cancer for postoperative diagnosis, screening, and follow-up. Previous studies investigated...
the significance of PSA above 0.2 ng/mL post radical prostatectomy. PSA testing facilitates earlier prediction of PSA recurrence. However, not all patients with detectable PSA levels will experience clinical progression. In general, patients with advanced prostate cancer are no longer suitable for surgical treatment. Therefore, the prognostic value of T stage for biochemical recurrence of prostate cancer is not very clear. Furthermore, traditional imaging technology has limited value in identifying BCR at extremely low PSA levels. To better differentiate the risk of recurrence, reliable predictors are needed. A personalized management approach to post-treatment prostate cancer requires efficacious recurrence risk prediction to implement timely care and circumvention of overtreatment.

With the continuous progress of gene detection technology and the increasing ability to define the molecular characteristics of prostate cancer, there is a growing interest in using molecular markers to distinguish the severity of cancer. In this study, leveraging prostate cancer transcriptome data from The Cancer Genome Atlas (TCGA), we identified a 5-gene signature, which might allow for early identification of patients at increased risk for biochemical recurrence of prostate cancer. It has the potential to serve as a beneficial predictor for appropriate follow-up and timely treatment after radical prostatectomy. In addition, it may be used for risk stratification of patients who have not yet reached the PSA risk threshold for biochemical recurrence.

Materials and Methods

Ethical Statement

The authors are responsible for all aspects of the work to ensure that issues related to the accuracy or integrity of the work are properly investigated. This research was conducted in accordance with the Declaration of Helsinki (revised in 2013). TCGA projects provided a set of policies to protect the privacy of participants donating specimens to TCGA.

Determination of DEGs Between Normal Prostate and Prostate Cancer Samples

Patients from the original cohort have provided consent and that this study was conducted in line with declaration of Helsinki. The transcriptome raw count data of prostate cancer in TCGA (UCSC XENA) were used for analysis. The corresponding phenotype data, including biochemical recurrence events, were matched to the samples. The samples with missing biochemical recurrence data were excluded from this study. R Annotables package was leveraged for identifier conversion (version GRCh38). The average count was calculated when a gene corresponded to several Ensembl gene identifiers. Determination of DEGs between normal prostate and prostate cancer samples was conducted leveraging the R DESeq2 package using the cutoff: p < 0.05 and the absolute value of log2 fold change ≥ 1. Expression data were normalized using variance stabilizing transformation.

KEGG and GO Enrichment Analyses of DEGs

DEGs were used as input for the enrichment of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology biological processes (GO_BP). R clusterProfiler package was used for enrichment analyses. For KEGG enrichment analysis, all 12 significant (adjusted p value < 0.05) pathways were shown, while in GO_BP enrichment analysis, the top 20 most significant (according to ranked adjusted p value) biological processes were demonstrated. P values were adjusted through the Benjamini-Hochberg Procedure.

Construction of the Prediction Model for Biochemical Recurrence in Prostate Cancer

The 413 patients were randomly divided into a training cohort used for model construction and a validating cohort used for model validation. Univariate Cox regression was conducted in the training cohort using the DEGs and biochemical recurrence data to identify biochemical recurrence-related genes. The least absolute shrinkage and selection operator (LASSO) regression was leveraged to avoid model overfitting by eliminating some variables based on penalty parameter lambda, which was determined by cross-validation using the R glmnet package. The minimum of the cross-validation error means was treated as the best lambda value. Multivariate Cox regression analysis was leveraged for construction of the prediction model for biochemical recurrence in prostate cancer.
Validation of the Predictor for Prostate Cancer Biochemical Recurrence

Risk score obtained from multivariate Cox regression analysis was used for patients grouping. The risk score constructed according to the theory of the Cox model is based on the following equation: risk score = coefficient1 * gene1 expression + ... + coefficientN * geneN expression. The coefficients corresponding to each gene were generated by cox model calculations as mentioned above. Risk scores smaller than the median value of all risk scores in the training cohort were artificially defined as low risk, and correspondingly, higher scores (>50% median) were defined as high risk. Firstly, patients in the training and validating cohorts were ranked according to the risk score, and counts of biochemical recurrence events in low-risk and high-risk groups were calculated for comparison. An expression heatmap of the 5 genes in the predictor was plotted to check the difference between high and low-risk groups. Kaplan–Meier curves for the time to biochemical recurrence and ROC curves were drawn to validate the predictive accuracy of the predictor.

Data Processing, Statistical Analysis and Visualization

Data were analyzed using R Programming software. The correlations of different variables were assessed using Pearson correlation. Differences between variables were evaluated with independent t-tests. P < 0.05 was considered to be statistically significant. Data visualization was implemented leveraging R and Adobe Illustrator. R packages including DESeq2, annotables, clusterProfiler, glmnet, survival, survivalROC, enrichplot, ggplot2, ggbiplot, pheatmap packages were used for analysis and plotting.

Results

DEGs Between Were Correlated with Cellular Metabolism

A total of 413 prostate cancer samples, including 375 prostate cancer samples and 38 normal prostate samples, were selected from TCGA (UCSC XENA) using the following criteria: having complete biochemical recurrence data; not paraffin-embedded. The overall normalized gene expressions of all samples were roughly similar (Figure 1A). Principal component analysis (PCA) was conducted using the expression data of all genes and showed the relationships between the samples. Although the samples were not clustered by group (prostate cancer or normal prostate), the PCA plot still showed no unknown batch effect causing outliers or weird clustering of samples (Figure 1B). A total of 146 DEGs, including 12 upregulated and 134 downregulated genes, were identified in prostate cancer compared to normal prostate (Figure 1C and D). KEGG and GO_BP enrichment analyses showed that the DEGs were correlated with several cellular metabolism pathways (Figure S1A and B).

A Model Comprising 5 Genes Was Constructed via Univariate and Multivariate Cox Regression Analysis

208 patients were distributed into the training cohort, and 205 patients were distributed into the validating cohort. Univariate Cox regression analysis showed that 7 genes were significantly correlated with biochemical recurrence of prostate cancer. LASSO regression was conducted, and the minimum of the cross-validation error means was treated as the best lambda value (Figure 2A and B). According to multivariate Cox regression analysis, a model comprising 5 genes was constructed to predict prostate cancer biochemical recurrence (Figure 2C). The 5 genes included amine oxidase copper containing 1 (AOC1), DOPA decarboxylase (DDC), long intergenic non-protein coding RNA 1436 (LINC01436), orosomucoid 1 (ORM1) and phenylalanine hydroxylase (PAH). The computational formula was as follows: Training cohort risk score = (0.85 × expression of AOC1) + (1.20 × expression of DDC) + (1.26 × expression of LINC01436) + (1.18 × expression of ORM1) + (0.80 × expression of PAH) (Table S1).

The 5-Gene Model Was Correlated with Biochemical Recurrence and Was Speculated to Be a Potential Biochemical Recurrence Predictor of Prostate Cancer

To test the efficacy of the 5-gene model in predicting biochemical recurrence of prostate cancer, a risk score was calculated for each patient in training and validating cohorts. Then patients in the training and validating cohorts were ranked according to the risk score (Figure 3A and B). The count of patients with biochemical recurrence was higher in
the high-risk score group than the relative low-risk score group in training and validating cohorts (Figure 3C and D). The gene expression heatmap of the 5 genes showed that the expressions of PAH and AOC1 were decreased with increasing risk score, while the expressions of DDC, LINC01436 and ORM1 were increased with increasing risk score. These expression trends were found in both training and validating cohorts (Figure 3E and F). Kaplan–Meier curves to biochemical recurrence events showed that patients with low-risk scores had better biochemical recurrence-free survival than patients with high-risk scores in both training and validating cohorts (Figure 4A and B). The area under the curve (AUC) values were 0.739 and 0.625 in the training and validating cohorts, respectively, indicating that the 5-gene model
had a reliable utility in identifying the biochemical recurrence risk of prostate cancer (Figure 4C and D). These data together suggested that the 5-gene was a potential biochemical recurrence predictor of prostate cancer.

**Discussion**

Radical prostatectomy is an effective choice for the majority of localized prostate cancer patients. However, about 15–40% of patients who underwent radical prostatectomy experienced biochemical recurrence within 5 years.\textsuperscript{12–14} Once biochemical recurrence occurs, radiotherapy is the main treatment option.\textsuperscript{15} Follow-up imaging sometimes failed in detecting recurrence. Radiotherapy is usually initiated in patients with PSA levels below 1 ng/mL, which is not sensitive in standard care.
imaging. The serum PSA tracking is the only way to monitor biochemical recurrence in patients after primary treatment, indicating a critical unmet need for biomarkers for risk stratification in patients after treatment. Traditionally, it has been recommended to keep active monitoring until PSA has reached 0.2 ng/mL. Although early postoperative radiotherapy may be beneficial for patients at potential risk of biochemical recurrence, the lack of practical markers has the potential to allow for side effects from overly aggressive radiotherapy. Therefore, identification of practical markers of cancer progression is important for predicting the risk of recurrence and for active surveillance using PSA.

Leveraging Cox regression, we established a 5-gene model to identify patients with prostate cancer at increased risk for biochemical recurrence. The LASSO regression validation showed that this novel model might serve as a valuable tool. LASSO is a regression-based method, which allows a large number of covariates in the model. Importantly, it has the unique function of penalizing the absolute value of the regression coefficient. Therefore, the adjustment coefficient may impact the overall regression. The greater the penalization, the greater the coefficient shrinkage, some reaching 0, to automatically remove unnecessary covariates.

Intriguingly, the expression levels of PAH, AOC1, DDC, LINC01436 and ORM1 at the early stage of prostate cancer were correlated with late biochemical recurrence. Since PAH and AOC1 were negatively associated with biochemical recurrence risk, while DDC, LINC01436 and ORM1 were positively correlated with biochemical recurrence risk, PAH and AOC1 might play different roles with DDC, LINC01436 and ORM1 in the late relapse of prostate cancer. To date, the functions of the 5 genes in prostate cancer biochemical recurrence are not elucidated yet. Compared to the other 4 genes, there are relatively more studies focusing on the roles of DDC in prostate cancer.

There is evidence that DDC is highly expressed in prostate cancer and might be a relevant biomarker for prostate cancer. Besides, DDC is revealed to enhance androgen receptor transcriptional activity. The antitumor effect of
Carbidopa (an inhibitor of DDC) in prostate cancer is related to ubiquitination and degradation of androgen receptors.\(^\text{20}\) It is well established that long noncoding RNAs (lncRNAs) are involved in a panel of human diseases, including various types of tumors.\(^\text{21,22}\) LINCO01436 is reported to promote gastric cancer and non-small cell lung cancer (NSCLC).\(^\text{23,24}\) ORM1 is a kind of acute-phase protein, increasing in serum with inflammatory and immunomodulatory properties. ORM1 has been reported as a biomarker in various advanced cancers, such as bladder cancer and ovarian cancer.\(^\text{25–27}\) PAH encodes for phenylalanine hydroxylase and lack of this enzyme activity leads to the autosomal recessive disorder phenylketonuria.\(^\text{28}\) Whether PAH is involved in prostate cancer biologic progression, especially postoperative recurrence, remains unknown. AOC1 is reported to encode a metal-binding membrane glycoprotein regulating polyamine breakdown and is inhibited by amiloride.\(^\text{29,30}\) Kirschner et al showed that AOC1 was a downstream target gene of the Wilms tumor protein, promoting the proliferation of human Wilms tumor.\(^\text{31}\) Likewise, AOC1 promotes colorectal cancer and gastric cancer progression.\(^\text{32,33}\) In this study, we provided evidence of the biological significance of the 5 genes in prostate cancer.

Figure 4 Validation of reliability of the 5-gene model. (A and B) Kaplan–Meier curves showed that biochemical recurrence-free survival was better in the low-risk group compared with the high-risk group in both training cohorts and validating cohorts. (C) ROC curve of training cohorts. The area under the curve (AUC) value was 0.739. (D) ROC curve of validating cohorts. AUC was 0.625, indicating the 5-gene model had high specificity and sensitivity.
Nonetheless, several limitations still existed, and further explorations were recommended. First, prostate cancer consist of several different subtypes, which may have different pathological mechanisms, and the conclusions of this study may not be applicable to all subtypes. In addition, the conclusions of this study are predictive and require more follow-up validation.

To date, the application of the majority of recognized prostate cancer biomarkers belongs to pre-treatment detections. Specific biomarkers indicating post-treatment progress of prostate cancer remain lacking besides serum PSA level, which has some limitations to reach extensive use in predicting biochemical recurrence. In this study, we sought to identify a potential predictor of biochemical recurrence in prostate cancer. PAH and AOC1 were negatively correlated with biochemical recurrence of prostate cancer, while DDC, LINC01436 and ORM1 were positively correlated with biochemical recurrence of prostate cancer. Besides, it might be used to risk-stratify patients who do not yet reach the risk threshold of PSA for biochemical recurrence. It might be beneficial to identify better patients for appropriate follow-up and timely treatment. In addition, these 5 genes might help to reveal potential mechanisms of biochemical recurrence of prostate cancer after radical prostatectomy. Based on these findings, it might be rewarding to investigate more details of the 5 genes in recurrent prostate cancer.

Data Sharing Statement
The TCGA database presented in this study can be found in online repositories.

Ethical Approval
The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by The Institutional Research Ethics Committee of The Third Xiangya Hospital of Central South University, Changsha, China (NO. 22009).

Acknowledgments
The authors appreciate study investigators and staff who participated in this study.

Author Contributions
All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding
This study was supported by Hunan Provincial Science and Technology Innovation Program Project (#2020SK53602 to Guangming Yin).

Disclosure
The authors have no conflicts of interest in this work.

References


