

Original Research

A Machine Learning Method for Predicting Biomarkers Associated with Prostate Cancer

Yanqiu Tong^{1,2,*}, Zhongle Tan³, Pu Wang⁴, Xi Gao⁵¹Laboratory of Forensic Medicine and Biomedical Informatics, Chongqing Medical University, 400016 Chongqing, China²School of Tourism and Media, Chongqing Jiaotong University, 400074 Chongqing, China³School of Traditional Chinese Medicine, Chongqing Three Gorges Medical College, 404120 Chongqing, China⁴Department of Rehabilitation, Southwest Hospital, Third Military Medical University (Army Medical University), 400038 Chongqing, China⁵Department of Traditional Chinese Medicine, University-Town Hospital, Chongqing Medical University, 400016 Chongqing, China*Correspondence: yqtong@126.com (Yanqiu Tong)

Academic Editor: Taeg Kyu Kwon

Submitted: 1 September 2023 Revised: 12 October 2023 Accepted: 24 October 2023 Published: 6 December 2023

Abstract

Background: Prostate cancer (PCa) is a prevalent form of malignant tumors affecting the prostate gland and is frequently diagnosed in males in Western countries. Identifying diagnostic and prognostic biomarkers is not only important for screening drug targets but also for understanding their pathways and reducing the cost of experimental verification of PCa. The objective of this study was to identify and validate promising diagnostic and prognostic biomarkers for PCa. **Methods:** This study implemented a machine learning technique to evaluate the diagnostic and prognostic biomarkers of PCa using protein-protein interaction (PPI) networks. In addition, multi-database validation and literature review were performed to verify the diagnostic biomarkers. To optimize the prognosis of our results, univariate Cox regression analysis was utilized to screen survival-related genes. This study employed stepwise multivariate Cox regression analysis to develop a prognostic risk model. Finally, receiver operating characteristic analysis confirmed that these predictive biomarkers demonstrated a substantial level of sensitivity and specificity when predicting the prognostic survival of patients. **Results:** The hub genes were *UBE2C* (Ubiquitin Conjugating Enzyme E2 C), *CCNB1* (Cyclin B1), *TOP2A* (DNA Topoisomerase II Alpha), *TPX2* (TPX2 Microtubule Nucleation Factor), *CENPM* (Centromere Protein M), *F5* (Coagulation Factor V), *APOE* (Apolipoprotein E), *NPY* (Neuropeptide Y), and *TRIM36* (Tripartite Motif Containing 36). All of these hub genes were validated by multiple databases. By validation in these databases, these 10 hub genes were significantly involved in significant pathways. The risk model was constructed by a four-gene-based prognostic factor that included *TOP2A*, *UBE2C*, *MYL9*, and *FLNA*. **Conclusions:** The machine learning algorithm combined with PPI networks identified hub genes that can serve as diagnostic and prognostic biomarkers for PCa. This risk model will enable patients with PCa to be more accurately diagnosed and predict new drugs in clinical trials.

Keywords: machine learning; prostate cancer; prognostic biomarker; prognostic model; drug targets

1. Introduction

Prostate cancer (PCa), the leading cause of cancer-related death in Western countries, predominantly affects men between the ages of 40 and 60 [1,2]. Early detection of the disease relies greatly on clinical indicators (biomarkers) and drug targets, which are crucial for monitoring disease development and successful therapy. Urinary liquid biopsy is attractive and promising for PCa detection. Apart from the specific biomarkers from urine, potential serum biomarkers that may allow the precision medicine revolution to take place include androgen receptor (AR) variants, bone metabolism, and neuroendocrine and metabolite biomarkers. In the subset of patients with bone metastases, bone sialoprotein (BSP) and osteopontin (OPN) have prognostic value. Higher BSP levels are related to a shorter time to develop bone metastases in patients with PCa. OPN may be useful for assessing the treatment response after chemotherapy in patients with castration-resistant PCa [3]. Older men over age of 50 are most likely to develop PCa.

The rate of PCa diagnosis has been increasingly growing, and certain patients may experience extended survival even following the metastasis of cancer to distant locations.

PCa is a heterogeneous disease of the male reproductive system associated with gene aberrations, cellular context, and the environment [4,5]. Lack of understanding pertaining to the exact molecular mechanisms that drive PCa advancement hampers the potential to effectively manage advanced conditions. Hence, a comprehensive understanding of the PCa biomarkers implicated in the proliferation, apoptosis, and invasion of PCa holds utmost significance for enhancing the efficacy of diagnostic and therapeutic approaches.

Advances in molecular biological, high-throughput platforms, and machine learning techniques have enhanced the identification of novel biomarkers and the screening of potential drug targets for PCa. Germline or somatic aberrations in the DNA damage repair genes are found in 19% of primary PCa and almost 23% of metastatic castration-resistant PCa, and compromise genomic integrity. Patients



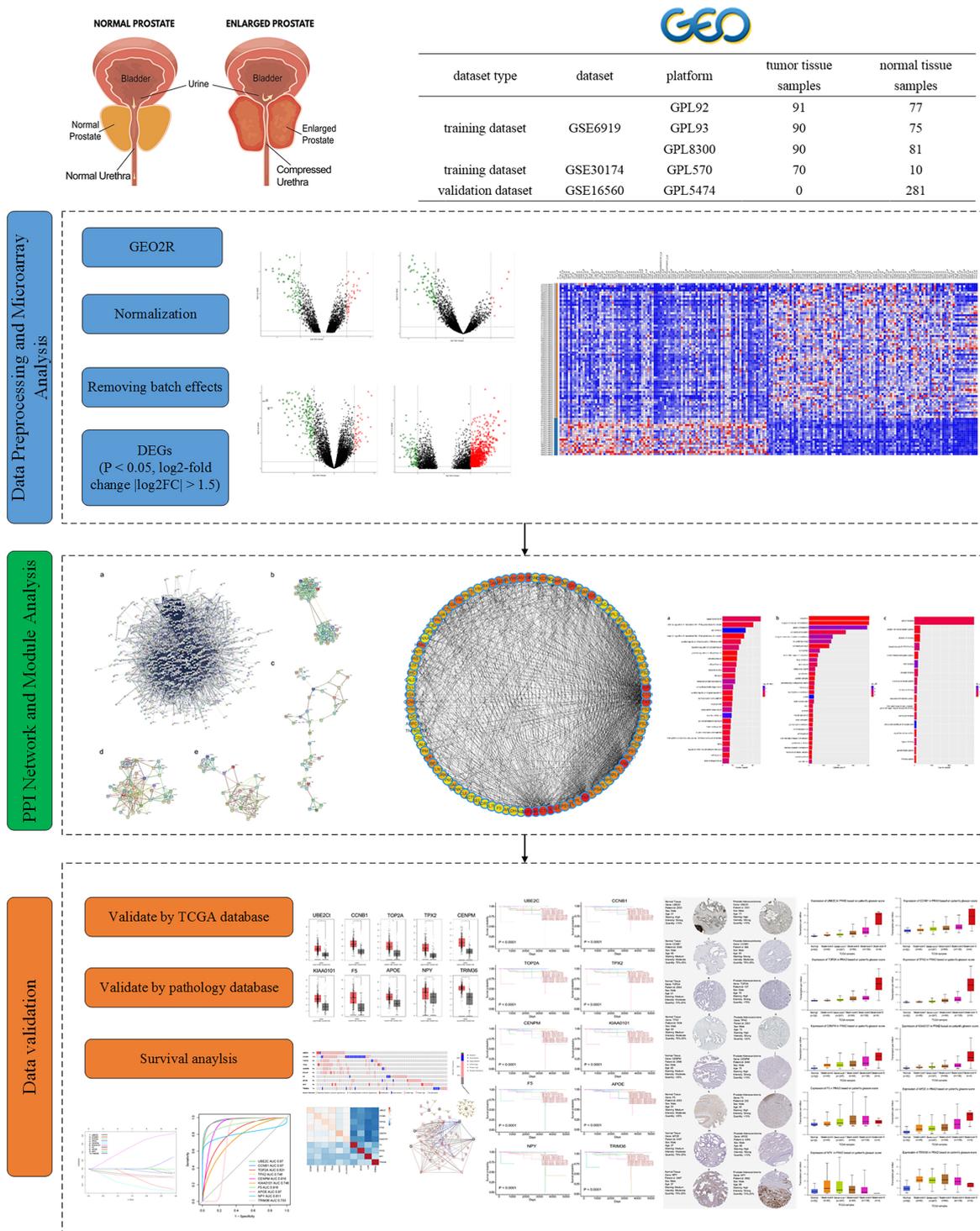


Fig. 1. Workflow for identifying hub genes and drug targets in prostate cancer. TCGA, The Cancer Genome Atlas; DEGs, differentially expressed genes; PPI, protein–protein interaction.

with breast cancer 2 (BRCA2) pathogenic sequence variants have increased levels of serum prostate-serum albumin (PSA) at diagnosis, an increased proportion of tumors with high Gleason score (GS), elevated rates of nodal and distant metastases, and high recurrence rates. Therapeutically, several poly (ADP-ribose) polymerase inhibitors have been investigated in patients with metastatic castration-resistant

PCa and are effective against germline BRCA2-mutant tumors [6]. This study assessed the potential of PCA biomarkers to both diagnose and predict disease status, and evaluated their effectiveness in predicting the response to drugs and the occurrence of treatment-related toxicities. The scope of this research ranged from molecular diagnosis to the classification of cancers at a molecular level. It also en-

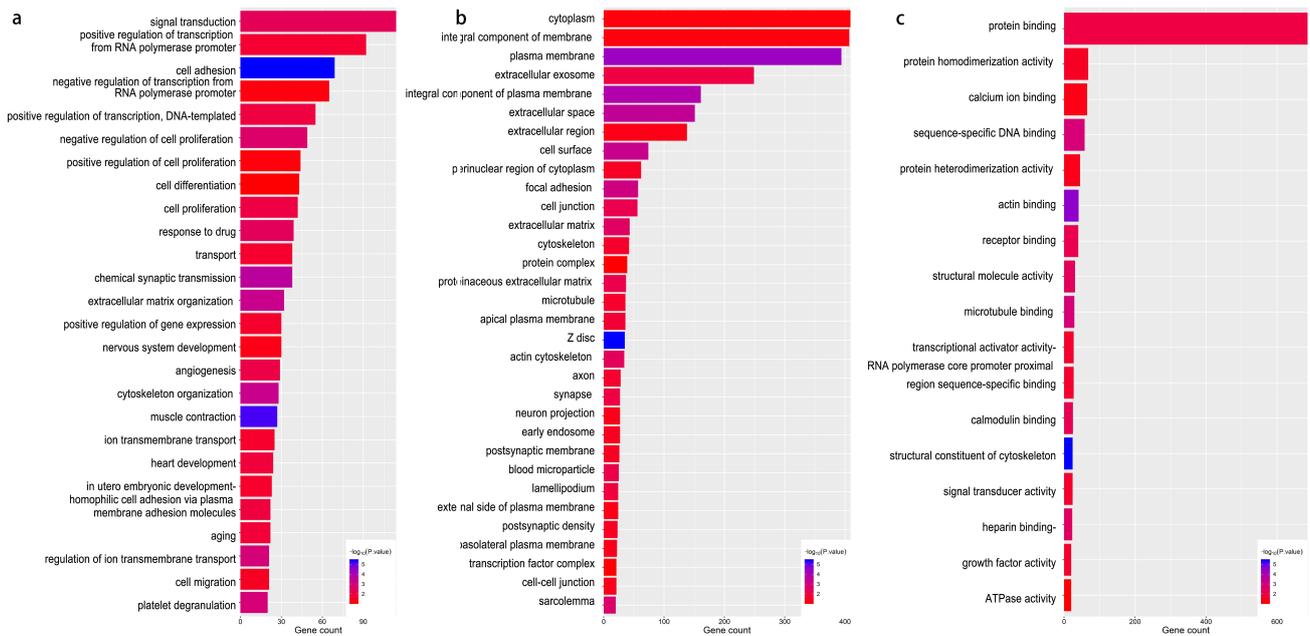


Fig. 2. Significant enriched genes. (a) Biological process. (b) Cellular component. (c) Molecular function.

compacted patient stratification, prognosis prediction, and the identification of novel targets for drug development as well as predicting the tumor response [7–9]. Consequently, high-throughput technologies, such as RNA sequencing (RNA-Seq) and microarray, have made it possible to identify PCa biomarkers for tumor prognosis in the development and progression of PCa. The integration of a machine learning approach with RNA-Seq data for PCa analysis is proving to be significant in unraveling the intricacy of the transcriptome [10,11]. RNA-Seq technology possesses numerous distinctive benefits including elevated sensitivity. However, it also has several novel challenges, the most notable of which involves the management of voluminous data. Despite the potential of machine learning technology to diminish superfluous and unrelated information, it necessitates substantial computational resources. Following years of advancement, microarrays have established a comprehensive framework, approach, and utilities. Consequently, microarray data can now be analyzed utilizing a personal computer, instead of relying on a workstation.

In this study, two types of microarray datasets, GSE6919 (public on Jan. 30, 2007) and GSE30174 (public on Dec. 21, 2012), were obtained from Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>). Then differentially expressed genes (DEGs) associated with PCa were calculated. Subsequently, the results of hub genes were generated by a machine learning algorithm and visualized by Cytoscape software (version 3.5.1, University of California San Diego, 9500 Gilman Drive, LaJolla, CA, USA). Finally, 10 candidate hub genes were found to be associated with PCa. Gene Ontology (GO) and pathway enrichment analyses were performed by Gene Set Enrichment Analysis (GSEA). By examining the biological functions and

Table 1. Gene expression profile information associated with PCa.

Dataset type	Dataset	Platform	Tumor samples	Normal samples
Training	GSE6919	GPL92	91	77
		GPL93	90	75
		GPL8300	90	81
	GSE30174	GPL570	70	10
Validation	GSE16560	GPL5474	0	281

PCa, Prostate cancer.

pathways, potential biomarkers that could be used for diagnosis, prognosis, and as drug targets can be identified and explored. Finally, all of these target genes and drug targets were validated by different types of bioinformatics database.

Graphs such as molecules, atoms, and proteins are considered chemical bonds and nodes are treated as edges. Numerous strategies have been suggested to carry out the classification of nodes, classification of graphs, and generation of graphs with the aim of focusing on molecular/protein graphs to anticipate molecular characteristics [12] and deduce protein interfaces [13]. In this short review, we discuss graph application on biological function prediction. There are three primary categories of graph embedding methods: factorization-based, random walks-based, and deep learning-based methods [14–16].

Analyzing key biomarkers in disease progression can benefit from the utilization of various algorithms that carry out intricate graph mining tasks. Some algorithms are semi-supervised learning for node-level classification [17], supervised learning for graph-level classification [18], and unsupervised learning for graph embedding [19]. We iden-

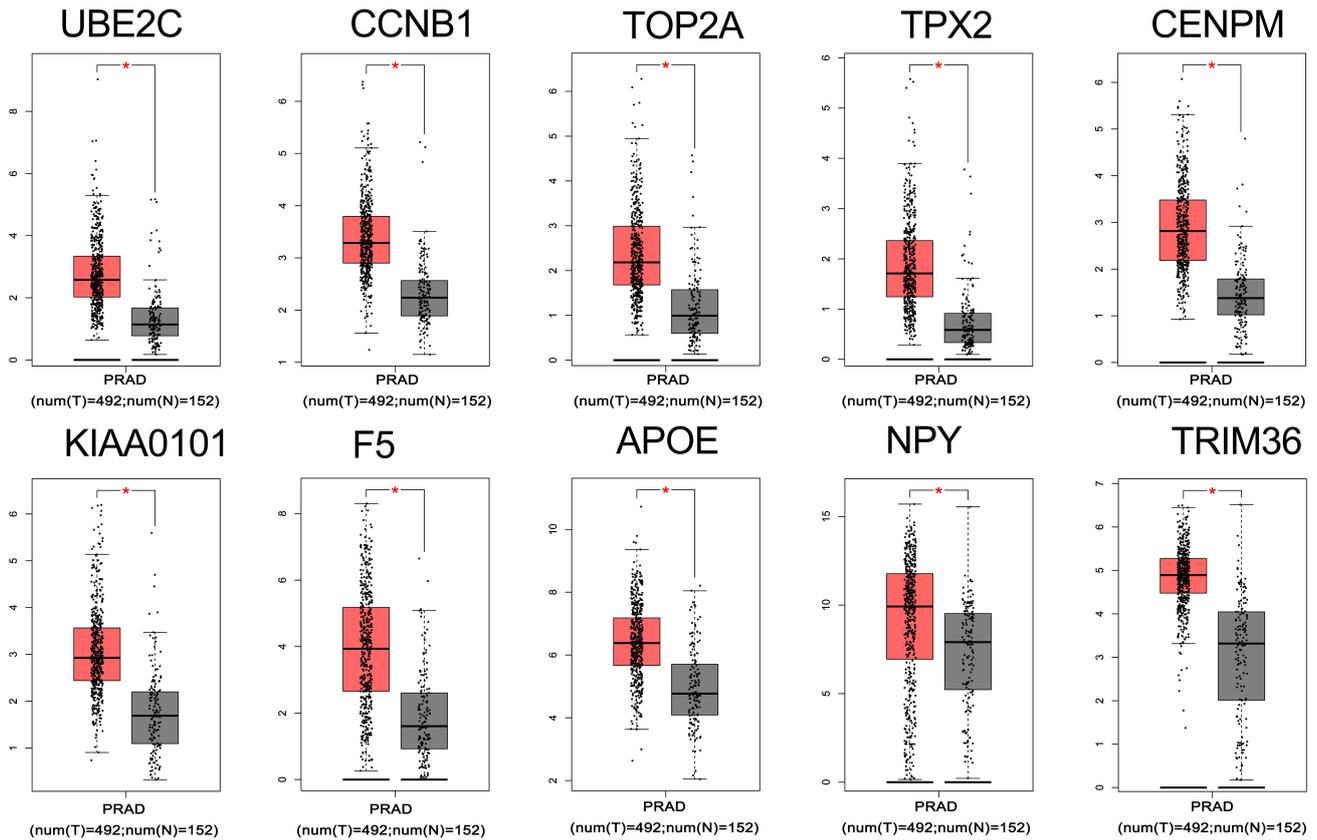


Fig. 3. Validation of hub genes by Gene Expression Profiling Interactive Analysis (red: tumor sample; grey: normal sample). PRAD, Prostate adenocarcinoma.

tified the diagnostic and prognostic biomarkers associated with PCa using the graph autoencoder (GAE) algorithm. Then prognostic biomarkers were verified in The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) dataset. TCGA, a landmark cancer genomics program, molecularly characterized over 20000 primary cancer and matched normal samples spanning 33 cancer types. The methodology employed in this study is presented in Fig. 1.

2. Materials and Methods

2.1 Definitions

Definition 1. Let graph $G(V, E)$ denote a graph with a collection of protein nodes, where $V = \{v_1, \dots, v_{12}\}$ and $E = \{e_{ij}\}$. G and E corresponds to the graph and edges set for protein-protein interaction (PPI) networks.

Definition 2. Given graph $G(V, E)$, graph embedding is a mapping $f : v_i \rightarrow y_i \in \mathbf{R}^d \forall i \in [n]$ such that $d \ll |V|$ and the function f preserve some proximity measure defined on graph G .

Each node in the graph embedding could be mapped to a low-dimensional feature vector.

2.2 Datasets

Three cohorts were utilized in our study; the descriptions of these cohorts are as follows. GSE6919

Table 2. Validation database for PCA.

Dataset name	Data link	References
GEPIA	http://gepia.cancer-pku.cn/	[27]
Human Protein Atlas	https://www.proteinatlas.org/	[28–30]
cBioPortal	https://www.cbioportal.org/	[31,32]
GTEx	https://www.gtexportal.org/home/	[33]
Ualcan	http://ualcan.path.uab.edu/	[34]
Oncomine	https://www.oncomine.org/	[35]
DrugBank	https://www.drugbank.ca/	[36]
systemsDock	http://systemsdock.unit.oist.jp/ (accessed November 2022)	[37,38]
GTEx	https://gtexportal.org/home/	[39]

GEPIA, Gene Expression Profiling Interactive Analysis; GTEx, Genotype-Tissue Expression.

and GSE30174 were obtained from the GEO database as training datasets. GSE6919 was based on the Agilent GPL92, GPL93, and GPL8300 platforms (Affymetrix Human Genome U95 Version 2 Array), submitted by Federico Alberto Monzon (2018). The GSE6919 dataset contained 504 samples, including 233 normal prostate tissues and 271 metastatic prostate tumors. GSE30174 was based on the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array), submitted by Jennifer Barb (2019). The GSE30174 dataset contained 80 samples, including 10

Table 3. KEGG pathway analysis of DEGs associated with PCa.

Pathway	Count	<i>p</i> -value
Dilated cardiomyopathy	23	1.28×10^{-7}
Hypertrophic cardiomyopathy	19	1.25×10^{-5}
ECM-receptor interaction	19	6.00×10^{-5}
Proteoglycans in cancer	31	1.72×10^{-4}
Arrhythmogenic right ventricular cardiomyopathy	15	3.51×10^{-4}
Morphine addiction	18	3.52×10^{-4}
Adrenergic signaling in cardiomyocytes	23	5.37×10^{-4}
Amoebiasis	19	7.92×10^{-4}
Insulin secretion	16	0.001420903
GABAergic synapse	16	0.001420903
Pathways in cancer	46	0.002427201
Retrograde endocannabinoid signaling	17	0.003179689
Tyrosine metabolism	9	0.003610174
Focal adhesion	27	0.005655205
Vascular smooth muscle contraction	18	0.00596106
Phenylalanine metabolism	6	0.006699088
Transforming growth factor beta signaling pathway	14	0.009099806

KEGG, Kyoto Encyclopedia of Genes and Genomes; ECM, Ecm-receptor-interaction.

healthy peripheral blood and 70 non-metastatic prostate tumors. GSE16560 as a validation dataset was based on the GPL5474 platform (Human 6k Transcriptionally Informative Gene Panel for DASL), submitted by Andrea Sboner (2013), contained 281 samples including primary prostate tumors ordered by different GS. Dataset information is summarized in Table 1.

All of the training datasets were analyzed using the online tool GEO2R. Then the DEGs were calculated using the limma R package (version 3.36.5, WEHI Bioinformatics, Bundoora, Victoria, Australia) between normal and tumor samples [20]. Multiple testing was corrected by the Benjamini and Hochberg (BH) [21] method to obtain the adjusted *p* value. The cutoff values for screening DEGs were set at an adjusted $p < 0.05$ and $|\log_2FC| > 1.5$.

2.3 Functional Enrichment and Pathway Analysis of DEGs

GO analysis is a widely used technique for annotating genes and gene products to identify the biological process (BP), cellular component (CC), and molecular function (MF) [22]. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database serves as a valuable resource for systematically analyzing gene functions and connecting genomic information with higher level functional information [23]. To ensure the success of any high-throughput gene functional analysis, it is crucial to map the user's genes to the appropriate biological annotations within the Database for Annotation, Visualization, and Integrated Discovery (DAVID) database [24]. This serves as a foundational step for the analysis process.

2.4 Transformed Features Using Machine Learning Framework

To construct a graph-embedding network based on the PPI network extracted from gene expression profile data, the Search Tool for the Retrieval of Interacting Genes (STRING) database was utilized to transform gene expression data into vector input. To assess the interactive associations among DEGs, we carried out mapping of the DEGs onto STRING database, exclusively considering experimentally confirmed interactions with a combined score ≥ 0.4 . Graphs are good at denoting information and relationships in PPI networks [25]. Due to the presence of identical protein complexes, proteins that possess similar topology exhibit comparable biological functions [26]. Consequently, the exploration of distinctive structural attributes among PPI networks associated with cancer might offer valuable insights into the advancement of tumors. With the aim of extracting, characterizing, and distinguishing network structures within intricate PPI networks, this research paper introduces an extensive and organized analysis of diverse graph embedding algorithms. Within the realm of graph embedding, numerous studies have been conducted, concentrating on the development of novel algorithms suitable for graphs encompassing millions of nodes and edges.

In this paper, we employed an unsupervised learning framework known as GAEs (graph autoencoders) to encode graphs or nodes into a latent vector space, and subsequently reconstructed the graph data based on the encoded information. GAEs can encode the node features and graph construction into the latent representations and decode the graph construction. So GAEs are suitable to acquire knowledge on network embedding and graph generative distributions from PPI nodes. Regarding network embedding, GAEs primarily focus on learning representations of protein nodes through the PPI graph's structural characteristics such as the graph adjacency matrix. Regarding graph generation, diverse approaches exist, including step-by-step generation of nodes and edges or the simultaneous generation of an entire graph. Finally, the GAE results were visualized by Cytoscape software.

2.5 Cross-Validation of Hub Genes and Drug Targets

To assess the validity of the hub genes and drug targets identified in the training dataset, cross-validation was conducted utilizing the merged databases. The databases utilized in this study are listed in Table 2 (Ref. [27–39]).

Evaluation of the performance of GAEs for 10 hub genes was performed by receiver operating characteristic (ROC) analysis (**Supplementary Material 1**).

2.6 Construction and Validation of the Gene-Related Prognostic Model

To further investigate the relationship between patient overall survival (OS) and the expression levels of individual hub genes, univariate Cox analysis [40,41] was carried

Analysis Type by Cancer	Cancer Cases	Cancer vs. Normal																	
		<i>UBE2C</i>	<i>CCNB1</i>	<i>TOP2A</i>	<i>TPX2</i>	<i>CENPM</i>	<i>F5</i>	<i>APOE</i>	<i>NPY</i>	<i>TRIM36</i>									
Bladder Cancer	288	6	6	7	3	3			1										
Brain and CNS Cancer	1,531	9	5	13	5	1	4	1	3		5	5							
Breast Cancer	1,602	20	17	24	1	16	14	1	1	3	2	2							
Cervical Cancer	200	3	3	4	4	4				2		1							
Colorectal Cancer	881	21	13	1	19	21		1	1		5	15	15						
Esophageal Cancer	132	3	2	2	2	2		5		3		2	2						
Gastric Cancer	637	7	4	8	8	8		2		7									
Head and Neck Cancer	628	10	10	11	10					6		1	1						
Leukemia	392	1	4	1	10		5	1	5		2	10		4					
Liver Cancer	212	4	3	4	4	4		4											
Lung Cancer	1,537	16	14	20	15	10		3		1			1		1				
Melanoma	630	1	2	2			2	2	2	2			1		1				
Ovarian Cancer	1,168	6	7	6	6	6		1											
Prostate Cancer	380	1	2	1	3	3		1		7		4	1	4	4				
Sarcoma	107	7	9	11	11	11		5	1			1							
Significant Unique Analyses		56	4	98	11	132	6	109	6	52	5	22	13	39	12	7	25	7	25
Total Unique Analyses		396		470		461		426		459		452		399		419		443	



Fig. 4. Validation of hub genes in the Oncomine database.

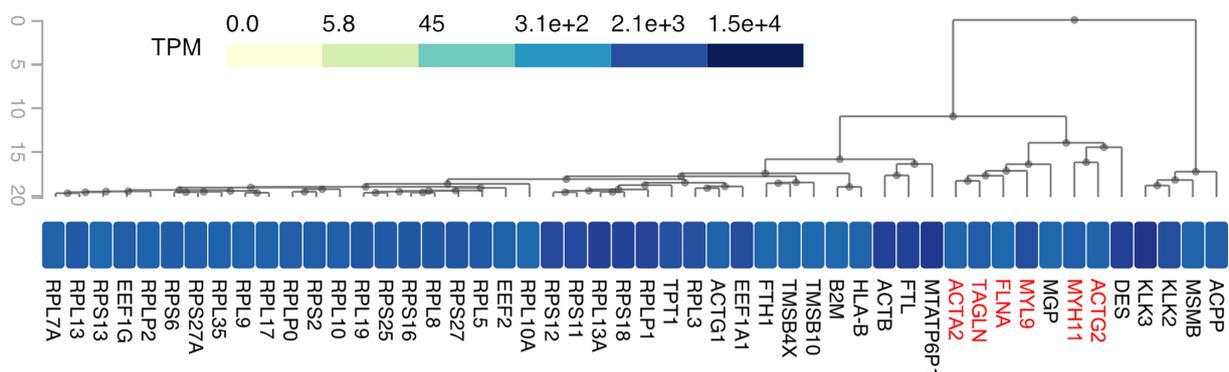


Fig. 5. Significant downregulated genes in the Genotype-Tissue Expression (GTEx) database.

out to construct a prognostic model of PCa. The survival and survminer R package [42] were utilized for this analysis. The screened genes with $p < 0.01$ in univariate Cox regression analysis were considered significant. Next, an independent investigation using the multivariate Cox proportional hazards regression analysis was performed to evaluate the influence of numerous genes as individual prognostic indicators impacting the survival of patients [43]. Finally, the stepwise method was used to select the optimized model. The risk score was calculated as follows:

$$risk\ score = \sum coefficient\ of\ gene_i \times expression\ level\ of\ gene_i.$$

ROC analysis was performed to assess the prognostic risk model, and the area under the curve (AUC) was calcu-

lated. The performance of the prognostic risk model was then validated using the GEO dataset GSE16560.

3. Results

3.1 Identification of DEGs in PCa

After data preprocessing, 6269 of 26,696 DEGs were identified; the top 200 DEGs included 153 upregulated genes and 47 downregulated genes (Supplementary Fig. 1). The DEGs are shown in Supplementary Fig. 2.

3.2 GO Enrichment Analysis and KEGG Pathway Analysis

To identify overrepresented GO categories and KEGG pathways, the obtained DEGs were subjected to analysis using the online software DAVID. The GO analysis results (Supplementary Material 2) revealed the significant en-

richment of DEGs in various BPs such as signal transduction, positive regulation of transcription from RNA polymerase II promoter, and cell adhesion (Fig. 2a). Furthermore, GO analysis demonstrated the significant enrichment of DEGs in cellular compartments such as the cytoplasmic vesicle membrane, integral component of membrane, and plasma membrane (Fig. 2b). Regarding the MF, the DEGs showed enrichment in protein binding, protein homodimerization, and calcium ion binding (Fig. 2c). The overall distribution of the GO results are shown in **Supplementary Fig. 3**.

Table 3 presents the enriched pathways of the upregulated and downregulated DEGs. These pathways were analyzed using KEGG analysis. The upregulated DEGs demonstrated enrichment in pathways such as the cell cycle, DNA replication, progesterone-mediated oocyte maturation, p53 signaling pathway, and extracellular matrix (ECM)–receptor interaction. Conversely, the downregulated DEGs exhibited enrichment in pathways including drug metabolism, metabolism of xenobiotics by cytochrome P450, retinol metabolism, hematopoietic cell lineage, and calcium signaling. The results of KEGG enrichment analysis conducted by GSEA [44] is shown in **Supplementary Fig. 4**. All of the upregulated genes were significantly enriched in dilated cardiomyopathy, hypertrophic cardiomyopathy, ECM–receptor interaction, arrhythmogenic right ventricular cardiomyopathy, focal adhesion, and the transforming growth factor beta signaling pathway.

3.3 Important Structures Resulted in Graphs of PCa-Related Genes

All DEGs were uploaded to the STRING database, with the analysis results shown in **Supplementary Fig. 5**. Based on the information available in the STRING database, a total of 6475 nodes were identified in topological form. In this paper, we present the illustrated GAEs, which utilize an encoder for extracting network embeddings and employ a decoder to ensure the preservation of nodes' topological information through the adjacency matrix [45] as:

$$PPMI_{v_1, v_2} = \max\left(\log\left(\frac{\text{count}(v_1, v_2) \cdot |D|}{\text{count}(v_1) \text{count}(v_2)}\right), 0\right)$$

Where $v_1, v_2 \in V$, $|D| = \sum_{v_1, v_2} \text{count}(v_1, v_2)$ and the $\text{count}(\cdot)$ function returns the frequency that node v and/or node u co-occur/occur in sampled random walks.

The top 100 genes generated by GAEs are visualized in **Supplementary Fig. 6**. The significantly upregulated genes were ubiquitin-conjugating enzyme E2C (*UBE2C*), cyclin B1 (*CCNB1*), topoisomerase II α (*TOP2A*), Xenopus kinesin-like protein 2 targeting protein (*TPX2*), centromere protein M (*CENPM*), coagulation factor V (*F5*), apolipoprotein E (*APOE*), neuropeptide Y (*NPY*), and

Table 4. Hub genes screened by the graph autoencoders (GAEs) algorithm.

Expression	Gene	Score	Expression	Protein-protein interaction (PPI) module
Upregulated	<i>UBE2C</i>	207	up	module 1
	<i>CCNB1</i>	138	up	module 1
	<i>TOP2A</i>	138	up	module 1
	<i>TPX2</i>	121	up	module 1
	<i>CENPM</i>	103	up	module 1
	<i>F5</i>	97	up	module 3
	<i>APOE</i>	96	up	module 3
	<i>NPY</i>	85	up	module 2
	<i>TRIM36</i>	78	up	module 4
Downregulated	<i>MYH11</i>	31	down	module 1
	<i>FLNA</i>	31	down	module 1
	<i>ACTA2</i>	31	down	module 2
	<i>MYL9</i>	27	down	module 2
	<i>TAGLN</i>	22	down	module 3
	<i>ACTG2</i>	22	down	module 4

UBE2C, Ubiquitin Conjugating Enzyme E2 C; *CCNB1*, Cyclin B1; *TOP2A*, DNA Topoisomerase II Alpha; *TPX2*, TPX2 Microtubule Nucleation Factor; *CENPM*, Centromere Protein M; *F5*, Coagulation Factor V; *APOE*, Apolipoprotein E; *NPY*, Neuropeptide Y; *TRIM36*, Tripartite Motif Containing 36; *MYH11*, Myosin Heavy Chain 11; *FLNA*, Filamin A; *ACTA2*, Actin Alpha 2, Smooth Muscle; *MYL9*, Myosin Light Chain 9; *TAGLN*, Transgelin; *ACTG2*, Actin Gamma 2, Smooth Muscle.

tripartite motif-containing 36 (*TRIM36*). The significantly downregulated genes were myosin heavy chain 11 (*MYH11*), filamin A (*FLNA*), actin alpha 2, smooth muscle (*ACTA2*), myosin light chain 9 (*MYL9*), transgelin (*TAGLN*), and actin gamma 2, smooth muscle (*ACTG2*) (Table 4).

3.4 Cross-Validation of Hub Genes in the Multi-Database

To conduct deeper exploration of the potential hub genes, validation of these findings was carried out using the Gene Expression Profiling Interactive Analysis database. The results obtained from this analysis are illustrated in Fig. 3. Notably, the hub genes demonstrated distinct significance in both the tumor and normal groups.

To confirm the hub gene expression among various cancers, the Oncomine database was employed to evaluate the expression of hub genes in tumor and normal tissues. By employing the criteria of $p < 0.01$ and $|\log_2\text{FC}| > 1.5$, a total of 396, 470, 461, 426, 459, 470, 452, 399, 419, and 443 unique analyses for *UBE2C*, *CCNB1*, *TOP2A*, *TPX2*, *CENPM*, *F5*, *APOE*, *NPY*, and *TRIM36*, respectively, were performed in Fig. 4.

Some of the downregulated hub genes were enriched in *ACTG2*, *MYH11*, *MYL9*, *FLNA*, *TAGLN*, and *ACTA2*, which were validated by the Genotype-Tissue Expression database (Fig. 5). Genotype-Tissue Expression (GTEx)

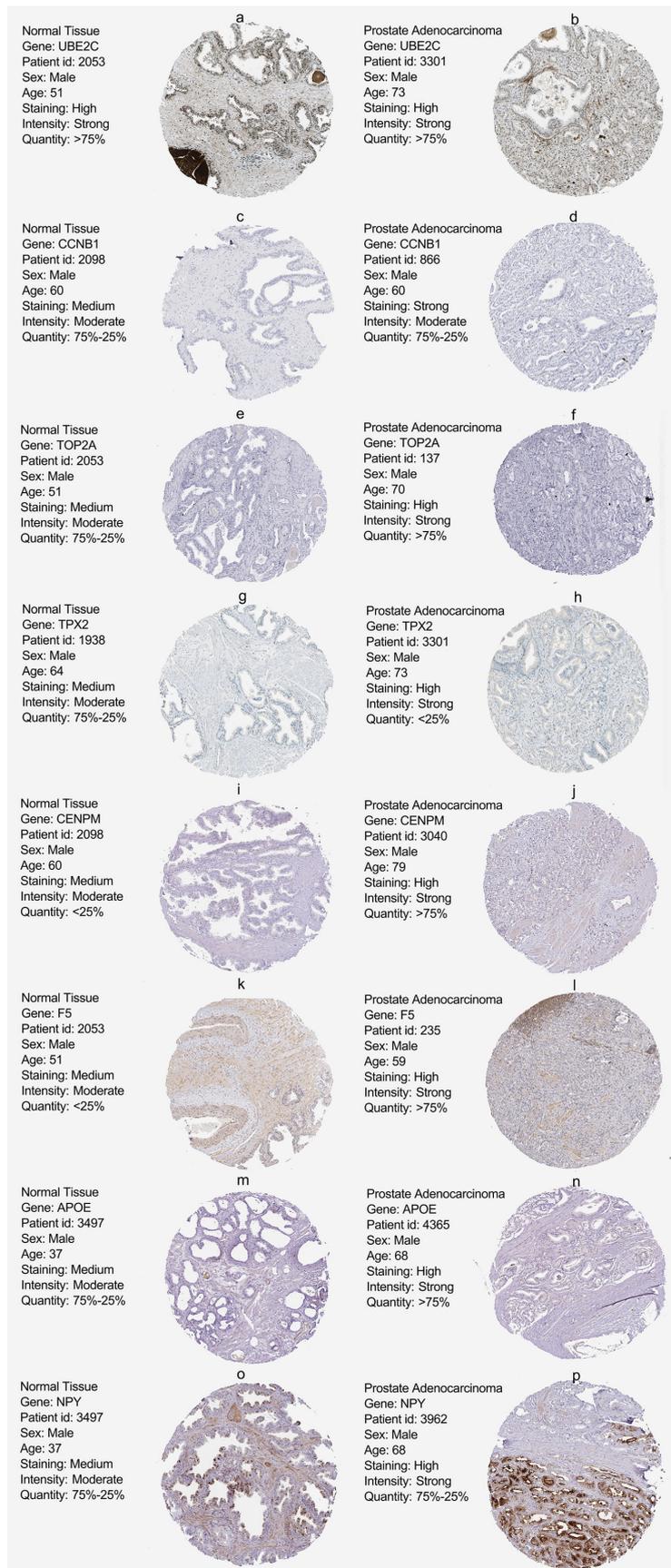


Fig. 6. Validation of the hub genes was performed using the Human Protein Atlas database.

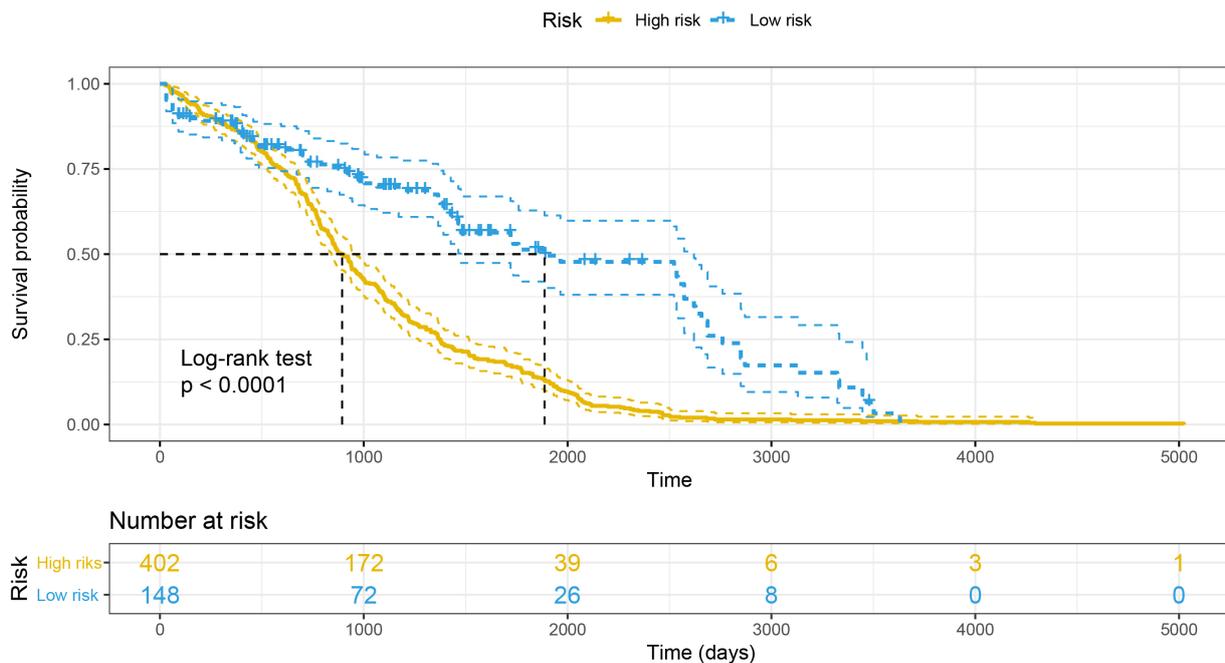


Fig. 7. Kaplan–Meier plots for biomarkers in the high- and low risk groups.

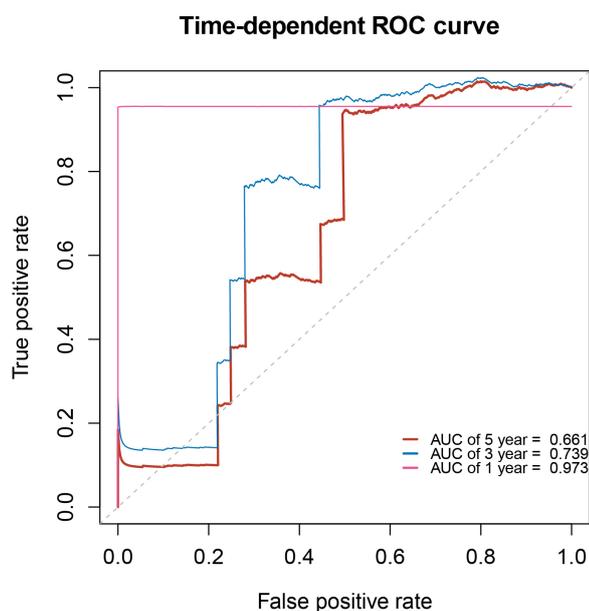


Fig. 8. Time-dependent (receiver operating characteristic, ROC) biomarker curves of 1-, 3- and 5-year overall survival.

Program established a data resource and tissue bank to study the relationship between genetic variants and gene expression in multiple human tissues.

The Human Protein Atlas database was utilized for the validation of hub genes (Fig. 6). However, there was no pathology results for TRIM36 in this database.

3.5 Construction of a Prognostic Model to Predict the OS of PCa Patients

Given the crucial role of hub genes in PCa, we analyzed the role of hub genes in predicting OS in PCa using univariate Cox regression analysis. Our findings revealed that nine genes exhibited a significant association with OS in patients with PCa (Table 5).

Finally, a stepwise multivariate Cox proportional hazards model was constructed (Table 6), and four genes were selected to build the following risk model:

$$\begin{aligned} \text{risk score} = & (-0.5663 * \text{expression level of TOP2A}) \\ & + (-1.2489 * \text{expression level of UBE2C}) \\ & + (-1.4976 * \text{expression level of MYL9}) \\ & + (-0.9500 * \text{expression level of FLNA}) \end{aligned}$$

Risk scores were computed for every patient in the training group, and subsequently, the patients were categorized into high- and low-risk groups. Fig. 7 illustrates the Kaplan–Meier OS curves, which clearly demonstrated that patients with high-risk scores exhibited markedly distinct survival outcomes compared to those with low-risk scores. The AUC values of the four-gene biomarker prognostic model at 1, 3, and 5 years were 0.973, 0.793, and 0.66, respectively (Fig. 8).

3.6 Validation of the Independent of Prognostic Model with Clinical Information

To assess the autonomous predictive significance of a prognostic model composed of four genes, univariate and multivariate Cox regression analyses were implemented us-

Table 5. Prognostic genes related to (overall survival) OS based on univariate Cox analysis.

Gene	Coef	(Hazard Ratio) HR	(Confidence Interval) CI	p-value	Regulation
<i>TPX2</i>	0.5047	1.656432	1.023110–2.681788	0.0040816	up
<i>CENPM</i>	0.2255	1.252922	0.7103220–2.210004	0.4361504	up
<i>NPY</i>	-0.1172	0.889405	0.7681821–1.029757	0.1169458	up
<i>APOE</i>	0.0829	1.086374	0.663267–1.779384	0.7420967	up
<i>TOP2A</i>	0.3193	1.376153	0.903891–2.095159	0.1365386	up
<i>CCNB1</i>	0.422	1.524966	0.7778998–2.989488	0.2191977	up
<i>TRIM36</i>	0.2853	1.330103	0.5813955–3.042977	0.4993109	up
<i>UBE2C</i>	0.4762	1.610022	1.052690–2.462424	0.0080329	up
<i>F5</i>	0.1178	1.125063	0.8306933–1.523748	0.4464131	up
<i>MYL9</i>	-0.386	0.679401	0.4523929–1.020319	0.0624607	down
<i>ACTA2</i>	-0.3975	0.671972	0.425578–1.061018	0.0880427	down
<i>MYH11</i>	-0.301	0.740083	0.5300268–1.033387	0.0772040	down
<i>TAGLN</i>	-0.3541	0.701817	0.4527059–1.088007	0.0034455	down
<i>ACTG2</i>	-0.3208	0.725560	0.5270065–0.9989209	0.0492314	down
<i>FLNA</i>	-0.3015	0.739676	0.4698100–1.164557	0.1928742	down

TPX2, TPX2 Microtubule Nucleation Factor; *CENPM*, Centromere Protein M; *NPY*, Neuropeptide Y; *APOE*, Apolipoprotein E; *TOP2A*, DNA Topoisomerase II Alpha; *CCNB1*, Cyclin B1; *TRIM36*, Tripartite Motif Containing 36; *UBE2C*, Ubiquitin Conjugating Enzyme E2 C; *F5*, Coagulation Factor V; *MYL9*, Myosin Light Chain 9; *ACTG2*, Actin Gamma 2, Smooth Muscle; *MYH11*, Myosin Heavy Chain 11; *TAGLN*, Transgelin; *ACTA2*, Actin Alpha 2, Smooth Muscle; *FLNA*, Filamin A.

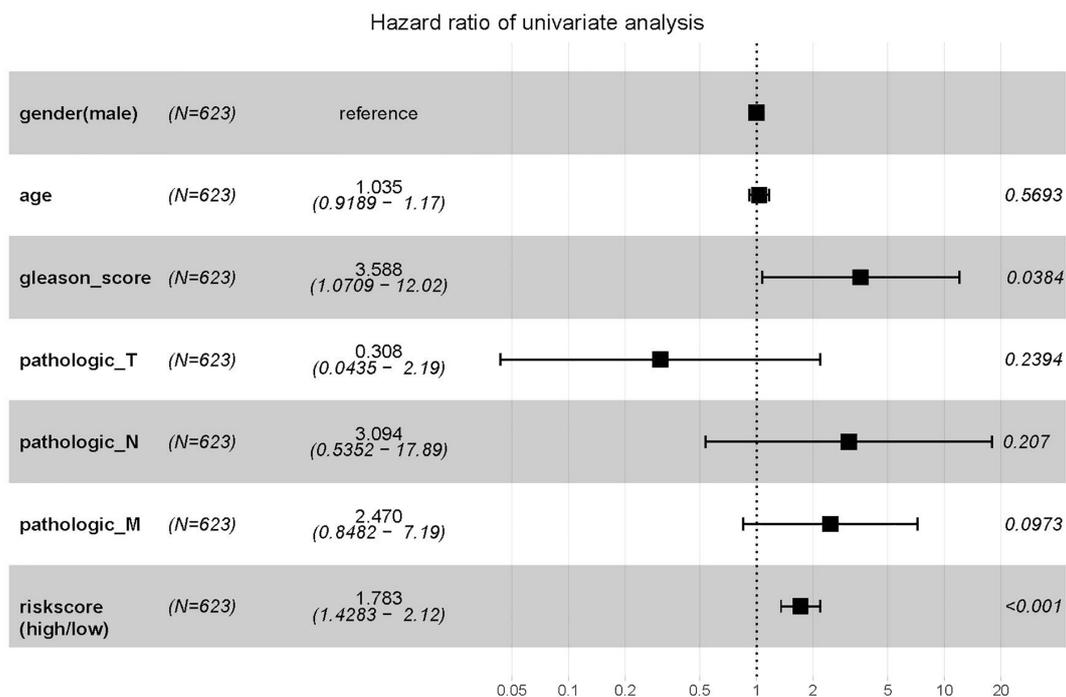


Fig. 9. Forest plot of risk scores and clinical factors based on univariate Cox regression analysis.

ing both TCGA prostate adenocarcinoma (PRAD) cohort and GSE 16560 cohort. Univariate Cox regression analysis demonstrated that the prognostic model incorporating clinical information such as GS and pathologic stage exhibited some prognostic value (Fig. 9). The statistical significance of age and GS was nearly achieved, which prompted us to include age, GS, and the prognostic model in the multivariate

Cox regression analysis (Fig. 10). The results of the multivariate Cox regression analysis showed that the prognostic model was independently associated with OS.

In addition, the prognostic model's predictive value was evaluated using the GSE16560 dataset. The dataset consisted of 280 patients, who were divided into a high-risk group (n = 190) and low-risk group (n = 90) based

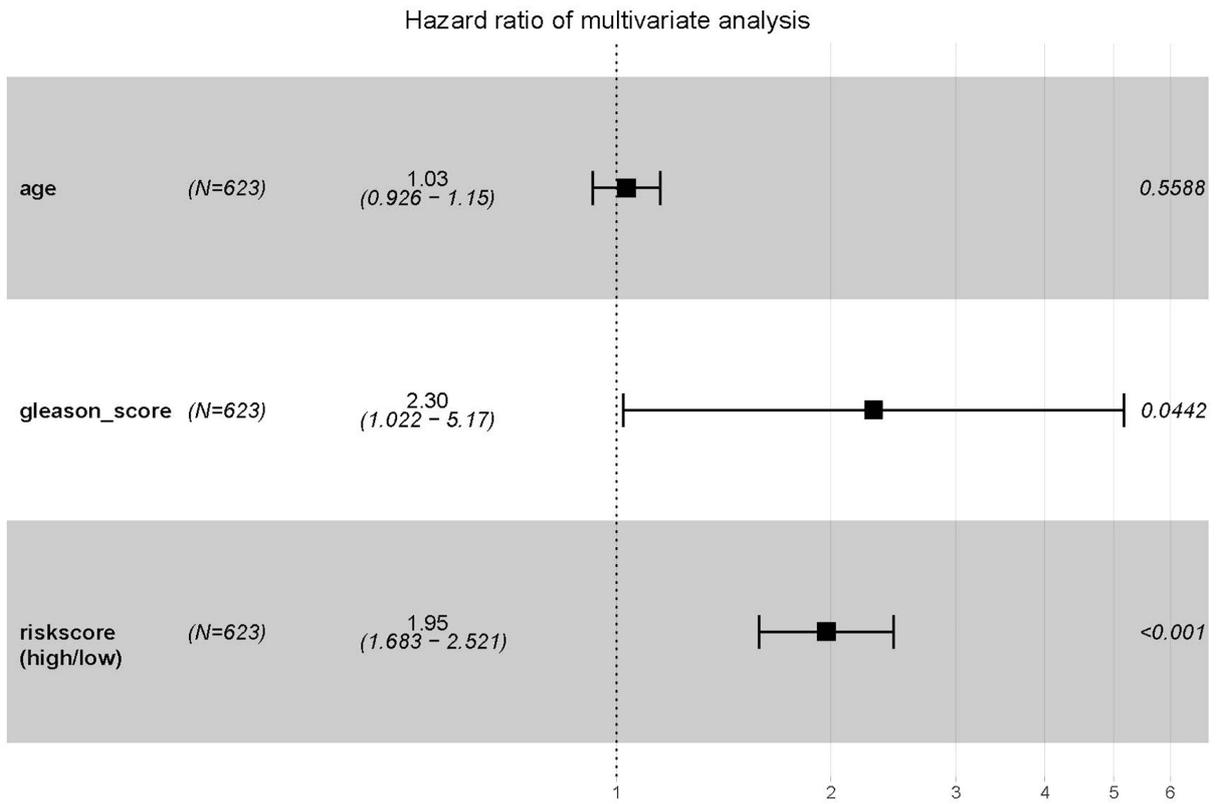


Fig. 10. Forest plot of risk scores and clinical factors based on multivariate Cox regression analysis.

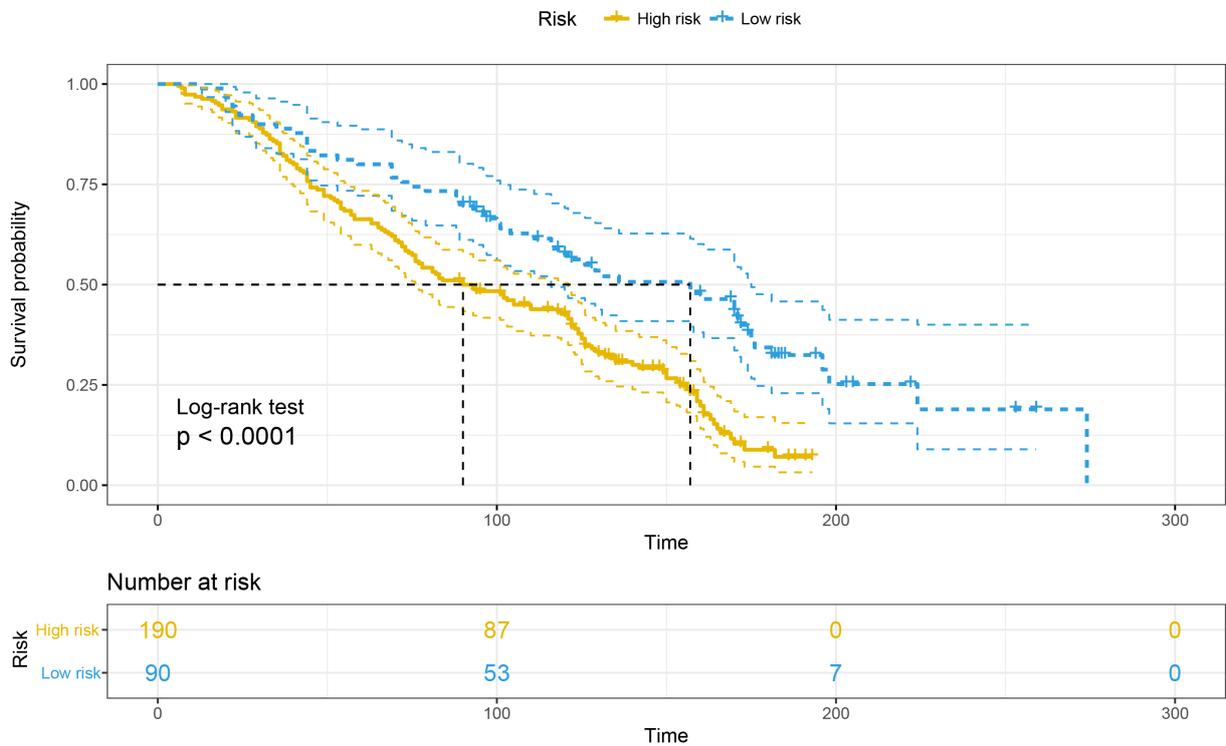


Fig. 11. Kaplan-Meier plots for biomarkers in the GSE16560 dataset.

Table 6. OS-related prognostic genes based on the multivariate Cox proportional hazards model.

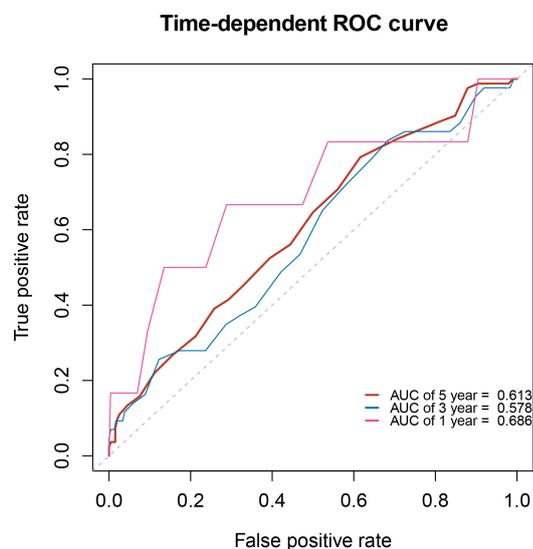
Gene	Coef	(Hazard Ratio) HR	(Confidence Interval) CI	p-value
<i>TOP2A</i>	-0.5663	0.567603	0.547639–0.867957	0.00276814
<i>UBE2C</i>	-1.2489	0.28680	0.181526–0.318121	0.00694469
<i>MYL9</i>	-1.4976	0.22372	0.201772–0.499484	0.00732035
<i>FLNA</i>	-0.9500	0.38651	0.203403–0.479835	0.00419485

TPX2, TPX2 Microtubule Nucleation Factor; *UBE2C*, Ubiquitin Conjugating Enzyme E2 C; *MYL9*, Myosin Light Chain 9; *FLNA*, Filamin A.

Table 7. Expression of candidate tumor suppressor genes in different PCa stages (H: upregulated; NO: not significant).

Hub gene	Gleason Score (GS)				
	GS = 6	GS = 7	GS = 8	GS = 9	GS = 10
<i>UBE2C</i>	NO	NO	H	H	H
<i>CCNB1</i>	H	H	H	H	H
<i>TOP2A</i>	NO	NO	H	H	H
<i>TPX2</i>	NO	NO	H	H	H
<i>CENPM</i>	H	H	H	H	H
<i>F5</i>	NO	H	H	H	H
<i>APOE</i>	H	H	H	H	H
<i>NPY</i>	H	H	H	H	NO
<i>TRIM36</i>	H	H	H	H	H

UBE2C, Ubiquitin Conjugating Enzyme E2 C; *CCNB1*, Cyclin B1; *TOP2A*, DNA Topoisomerase II Alpha; *TPX2*, TPX2 Microtubule Nucleation Factor; *CENPM*, Centromere Protein M; *F5*, Coagulation Factor V; *APOE*, Apolipoprotein E; *NPY*, neuropeptide Y; *TRIM36*, Tripartite Motif Containing 36.

**Fig. 12. Time-dependent (ROC) curve of biomarkers in the GSE16560 dataset.**

on the optimal cut-off value (Fig. 11). Fig. 12 illustrates the time-dependent ROC analysis results for the prognostic

model's survival prediction, which showed AUC values of 0.69, 0.58, and 0.61 at 1, 3, and 5 years, respectively.

3.7 Integrated Analyze for Drug Targets

All of the hub genes were uploaded to the cBioPortal to make pan-cancer analysis by TCGA PRAD dataset. The results showed that most of the hub genes had significant mutations (**Supplementary Fig. 7**). Furthermore, network analysis showed that the hub genes of *NYP*, *TOP2A*, and *TPX2* could also serve as drug targets (Fig. 13).

In the DrugBank database, the drugs amsacrine, biclutamide, dexrazoxane, doxorubicin, daunorubicin, enzalutamide, epirubicin, fleroxacin, mitoxantrone, teniposide, and valrubicin were PCa target drugs (**Supplementary Material 3**). The drug targets calculated by systemsDock showed that *NY*, *TOP2A*, and *TPX2* had significant docking scores with drug targets (Fig. 14).

3.8 Gleason Score (GS) System for PCa

PCa is usually classified by the GS system. The analysis of PCa tissue is done through the GS grading system, which characterizes the tissue based on its microscopic growth pattern. This evaluation method, known as the GS system, encompasses various levels of stratification. These include GS ≤ 6 , 3+4, 4+3, 8, 4+5, 5+4, and 10, corresponding with Gleason Grading Group 1, 2, 3, 4, and 5, respectively [46]. Consequently, GS ranges from 4 (2+2) to 10 (5+5). A favorable prognosis can be expected for individuals with a low GS score (≤ 6), as it indicates no risk of lymphatic metastasis. Conversely, individuals with a high GS score (> 8) are more likely to experience distal metastasis. Table 7 shows the significant deregulation of *UBE2C*, *CCNB1*, *TOP2A*, *TPX2*, *CENPM*, *F5*, *APOE*, *NPY*, and *TRIM36* in PCa based on the GS. Significant gene expression of *UBE2C*, *CCNB1*, *TOP2A*, *TPX2*, *CENPM*, and *APOE* was observed in the tumor samples (Fig. 15). There was no expression of *NPY*, which had a GS of 10. A nomogram with clinical information indicated that the GS score showed moderate value in PCa (**Supplementary Fig. 8**).

3.9 PSA-Related Pathway

PCa exhibited PSA dependence on *AR* signaling [47]. The relationship between PSA and the *AR* pathway is shown in **Supplementary Fig. 9**. PSA is a significant

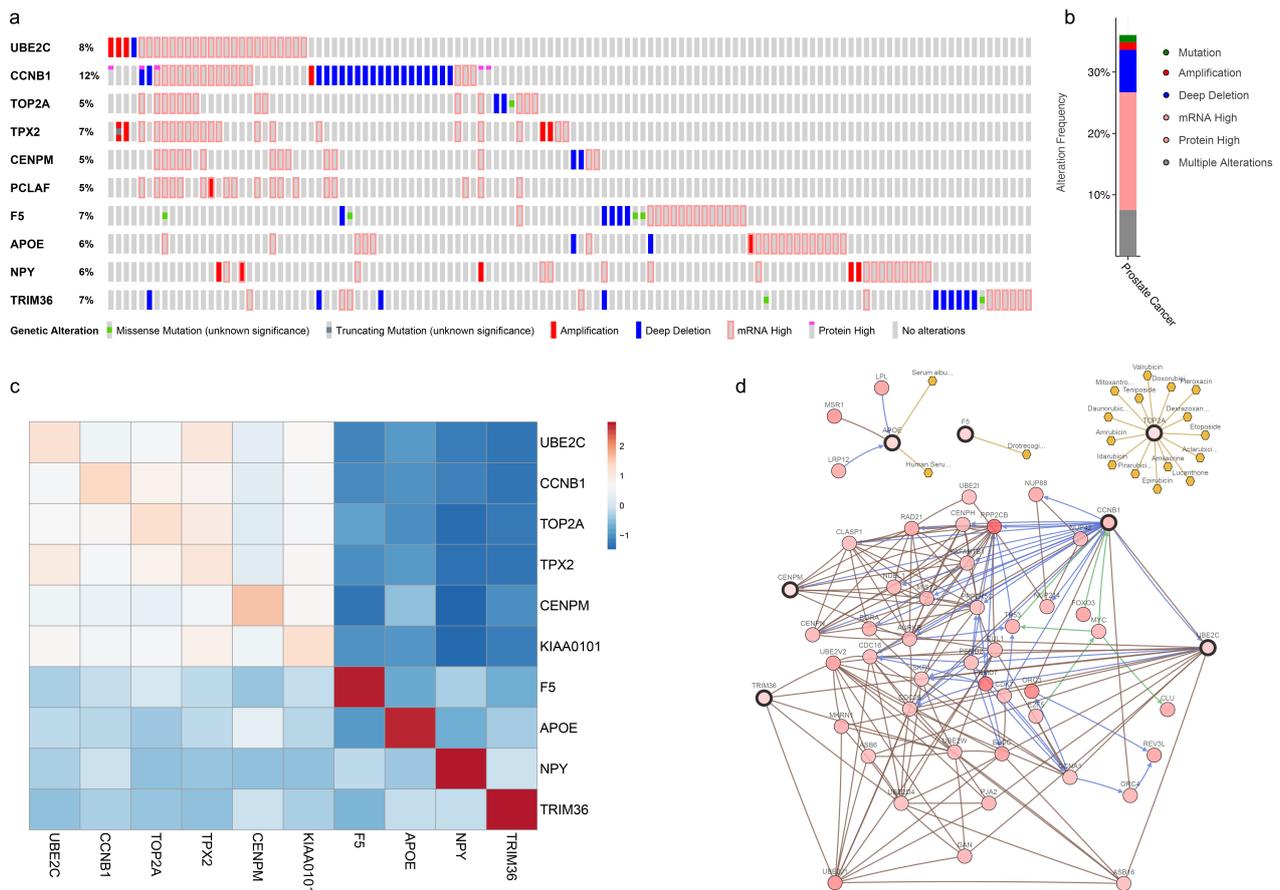


Fig. 13. Integration analysis for drug target. (a) Expression of hub genes in TCGA pan-cancer dataset. (b) Alteration frequency of hub genes in PCa. (c) Clustering of multivariate data. (d) Network of hub genes and drug targets.

biomarker that is currently used for PCa screening and diagnosis [48]. *AR* splice variants may contribute to the progression of PCa. Several therapeutic drugs, including bicalutamide and enzalutamide, specifically target *AR* signaling in the treatment of PCa [49].

4. Discussion

PCa is a prevalent and highly malignant tumor that is known for its complex molecular heterogeneity, making it a worldwide health concern. Therefore, uncovering its biological process may provide insights into the diagnosis and treatment for this disease. In this study, we used GSE6919 and GSE30174 as training datasets, and GSE16560 as the validation dataset. Univariate Cox regression analysis showed that 10 hub genes were related to the survival of patients with PCa. The four-gene risk model was established by multivariate Cox regression analysis.

The *AR* plays a crucial role in the development and progression of PCa. From the therapeutic perspective, abiraterone with prednisolone combined with androgen deprivation therapy (ADT) should be considered a new standard treatment for patients with high-risk non-metastatic PCa. In the metastatic setting, enzalutamide and abiraterone should

not be combined for those starting long-term ADT. Clinically important improvements in survival from the addition of abiraterone to ADT are maintained for longer than 7 years [50]. By conducting genome-wide mapping of AR-binding sites, researchers identified AR-binding regions that regulate the expression of *UBE2C*, a gene targeted by *AR* [51]. *UBE2C* has been shown to specifically regulate *AR* splice variant 7 (*AR-V7*) through the *UBE2C* promoter [52]. The expression of *UBE2C* is tightly controlled by the cell cycle, which makes it a highly relevant target for *AR* regulation, even under conditions of androgen independence. Chromatin immunoprecipitation sequencing analysis using an antibody that recognizes the N-terminal section of the *AR* [53] led to the identification of AR-binding sites. Previous studies have also confirmed a significant correlation between *UBE2C* and *AR-V7/AR3* [54]. *CCNB1* is crucial for controlling the cell cycle at the G2/M (mitosis) transition [55,56]. The upregulation of *CCNB1* can occur through Akt phosphorylation when there is overexpression of *Jagged1* and *AR* in PCa [57]. Li and colleagues [58] found androgen and the *AR* can increase the transactivation of *CCNB1* in LNCaP cells. Moreover, the co-overexpression of *Jagged1* and *AR* in PCa leads to the

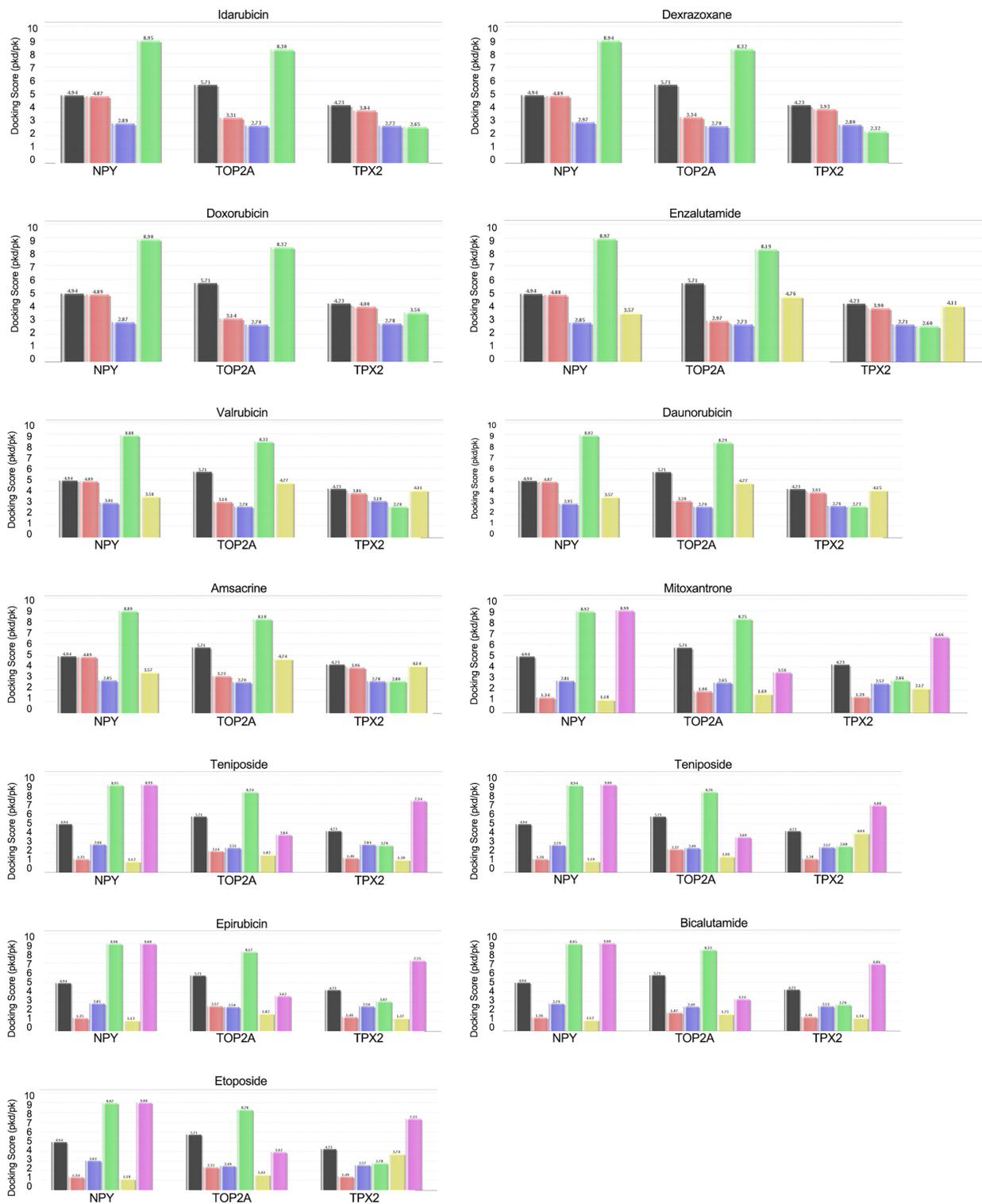


Fig. 14. Molecule docking prediction for drug target genes (*NPY*, *TOP2A*, and *TPX2*).

high expression of *CCNB1* [52]. The *AR* signaling pathway interacts with several other cell signaling pathways in PCa [59]. Forkhead box M1 serves as a common central transcriptional regulator that may play an important role in cell cycle-related targets such as *CCNB1* [60]. Upregu-

lated *CCNB1* may affect proliferation and contribute to tumorigenesis [61–64]. Compounds of small-molecule drugs that can inhibit PCa growth and block *CCNB1*-related pathways [58,60]. *TOP2A* has the function of controlling DNA topological structure, cell cycle progression, tumor devel-

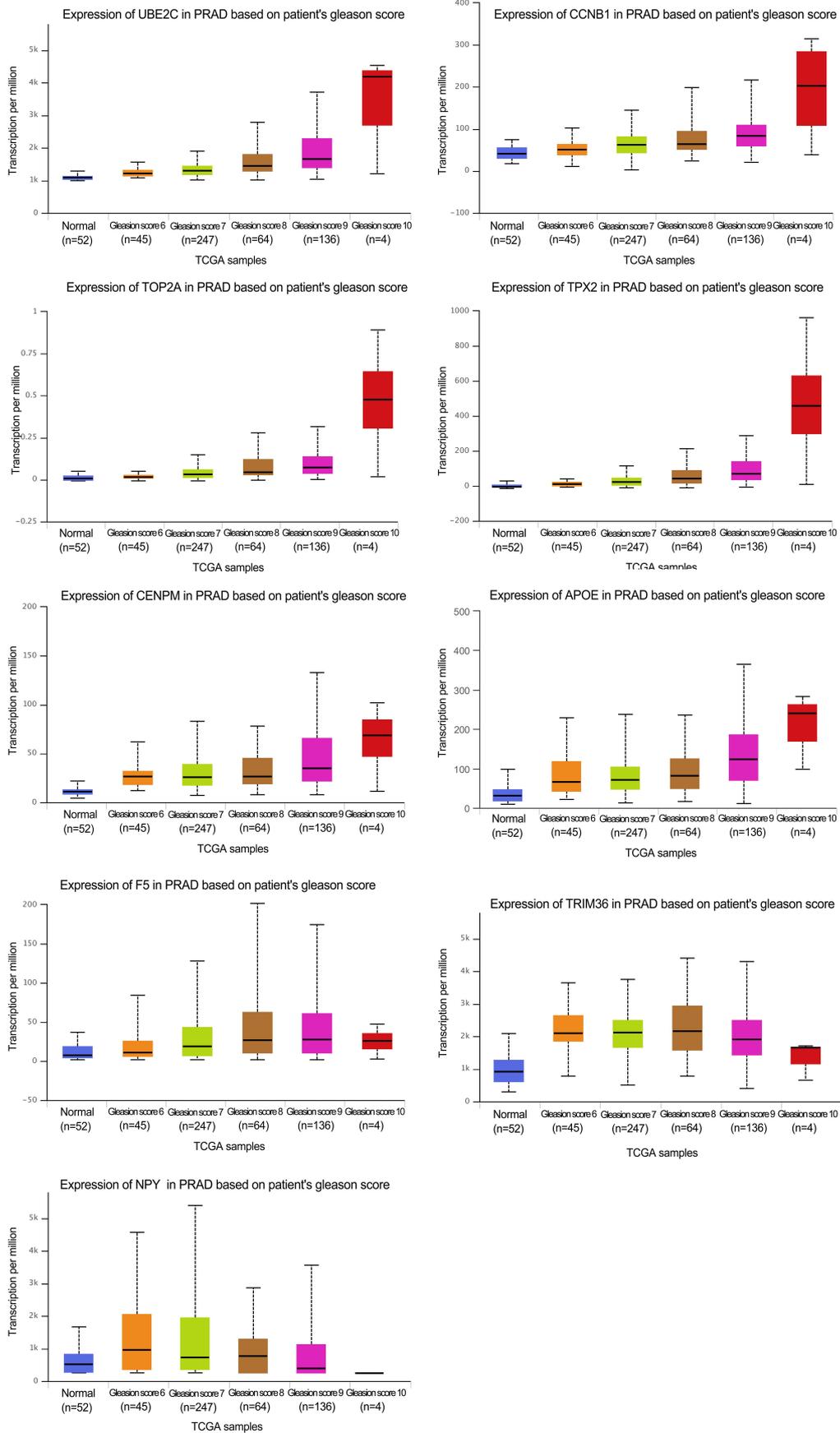


Fig. 15. Expression of hub genes according to the Gleason Score system.

opment [65]. Numerous studies have shown that aberrantly expressed *TOP2A* is associated with tumor progression in PCa [66]. Expression of the *TOP2A* protein is associated with increased GSs and elevated levels of preoperative PSA [67]. Modulation of DNA topological states and replication mainly account for the expression of *TOP2A* [68–70]. During mitosis, *TPX2* performs a crucial function in the process of chromosome segregation [71]. In mitosis, *TPX2* activates Aurora kinase A and directs its activity towards the mitotic spindle, thus playing a significant role [72]. Analysis using microarray techniques revealed a positive correlation between *TPX2* and the GS. The overexpression of *TPX2* is associated with the aggressiveness of PCa [73,74]. Inhibiting *TPX2* leads to the inhibition of PCa cell growth, increased apoptosis, and a reduction in tumorigenesis. Numerous studies have confirmed the potential therapeutic value of targeting *TPX2* in the treatment of PCa [73,74]. *CENPM* is a complex protein that plays a pivotal role in the assembly of kinetochore proteins, the progression of mitosis, and the segregation of chromosomes. *CENPM* is reportedly a signature PCa-related gene [75]. *F5* is a central regulator of hemostasis. It serves as a critical cofactor linked to cancer progression [76]. *APOE* is a crucial protein involved in regulating cholesterol levels. It also has the potential to inhibit cell proliferation, modulate immune regulation, and regulate cell growth and differentiation [77]. *APOE* has a potential role in PCa progression [78]. *NPY* has been implicated in the regulation of tumor advancement, including neuroendocrine tumors, as well as breast cancer and PCa [79]. The presence of reduced *NPY* expression levels is significantly linked to a more aggressive clinical phenotype in PCa. In TCGA pan-cancer cohort, PCa exhibits elevated *NPY* expression [80,81]. *TRIM36*, a member of the B-box family of zinc-finger proteins, plays a crucial role in cell cycle progression and cell growth attenuation [82]. Numerous studies have established a significant association between *TRIM36* and the GS, as well as its upregulation in the majority of PCa. Moreover, *TRIM36* delays the progression of the PCa cell cycle and prevents excessive cell proliferation. Intriguingly, restoring *TRIM36* expression during anti-androgen therapy has been shown to enhance the effectiveness of the drug [83].

5. Conclusions

In this study, a genome-wide analysis approach was utilized to indicate hub genes and drug targets in PCa. The findings hold promise in terms of offering a valuable collection of biomarkers for further exploration into the underlying molecular mechanisms of PCa. The identified biomarkers were found to be significantly associated with the OS of patients by employing multiple databases and multivariate analysis for validation. The AUC curve was used to verify the classification of these biomarkers and risk model. All of the biomarkers and pathways were based on a mathematical algorithm and bioinformatics tools. These potentially

prognostic biomarkers may be used to predict the molecular mechanisms and drug targets associated with PCa. Further experimental validation and clinical studies on these biomarkers should be conducted.

Availability of Data and Materials

All datasets (GSE6919, GSE30174 and GSE16560) in this article can be obtained through the GEO database at <https://www.ncbi.nlm.nih.gov>.

Author Contributions

YT and XG conceived the study; ZT and PW retrieved the literature; YT and ZT collected the data; YT wrote the R code; YT and ZT polished the manuscript; YT and XG wrote the paper. All the authors read and approve the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work. All authors contributed to editorial changes in the manuscript.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Not applicable.

Funding

This work was funded by Chongqing Language and Writing Research Project (yyk21213), Chongqing Natural Science Foundation General Project (cstc2021jcyj-msxmX0485), Humanities & social sciences of the Ministry of Education of the People's Republic of China (19YJA860022) and Basic science and frontier project of Chongqing Municipal Science and Technology Commission (cstc2016jcyjA0582).

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at <https://doi.org/10.31083/j.fbl2812333>.

References

- [1] Sekhoacha M, Riet K, Motloung P, Gumenu L, Adegoke A, Mashele S. Prostate Cancer Review: Genetics, Diagnosis, Treatment Options, and Alternative Approaches. *Molecules*. 2022; 27: 5730.
- [2] Gómez-Aparicio MA, López-Campos F, Pelari-Mici L, Buchser D, Pastor J, Maldonado X, *et al.* Bone health and therapeutic agents in advanced prostate cancer. *Frontiers in Bioscience-Landmark*. 2022; 27: 34.
- [3] Saxby H, Mikropoulos C, Boussios S. An Update on the Prog-

- nostic and Predictive Serum Biomarkers in Metastatic Prostate Cancer. *Diagnostics*. 2020; 10: 549.
- [4] Oh JJ, Shivakumar M, Miller J, Verma S, Lee H, Hong SK, *et al.* An exome-wide rare variant analysis of Korean men identifies three novel genes predisposing to prostate cancer. *Scientific Reports*. 2019; 9: 17173.
- [5] Aran V, Victorino AP, Thuler LC, Ferreira CG. Colorectal Cancer: Epidemiology, Disease Mechanisms and Interventions to Reduce Onset and Mortality. *Clinical Colorectal Cancer*. 2016; 15: 195–203.
- [6] Shah S, Rachmat R, Enyoma S, Ghose A, Revythis A, Bousios S. BRCA Mutations in Prostate Cancer: Assessment, Implications and Treatment Considerations. *International Journal of Molecular Sciences*. 2021; 22: 12628.
- [7] Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nature Clinical Practice. Oncology*. 2008; 5: 588–599.
- [8] Nannini M, Pantaleo MA, Maleddu A, Astolfi A, Formica S, Biasco G. Gene expression profiling in colorectal cancer using microarray technologies: results and perspectives. *Cancer Treatment Reviews*. 2009; 35: 201–209.
- [9] Bustin SA, Dorudi S. Gene expression profiling for molecular staging and prognosis prediction in colorectal cancer. *Expert Review of Molecular Diagnostics*. 2004; 4: 599–607.
- [10] Alkhateeb A, Rezaeian I, Singireddy S, Cavallo-Medved D, Porter LA, Rueda L. Transcriptomics Signature from Next-Generation Sequencing Data Reveals New Transcriptomic Biomarkers Related to Prostate Cancer. *Cancer Informatics*. 2019; 18: 1176935119835522.
- [11] Hamzeh O, Alkhateeb A, Zheng JZ, Kandalam S, Leung C, Atikukke G, *et al.* A Hierarchical Machine Learning Model to Discover Gleason Grade-Specific Biomarkers in Prostate Cancer. *Diagnostics*. 2019; 9: 219.
- [12] Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE. Neural Message Passing for Quantum Chemistry. *Proceedings of the 34th International Conference on Machine Learning. PMLR*. 2017; 70: 1263–1272.
- [13] Fout A, Byrd J, Shariat B, Ben-Hur A. Ben-Hur. Protein interface prediction using graph convolutional networks. *NIPS*. 2017; 6530–6539.
- [14] Zitnik M, Leskovec J. Predicting multicellular function through multi-layer tissue networks. *Bioinformatics*. 2017; 33: i190–i198.
- [15] Wale N, Watson I A, Karypis G. Comparison of descriptor spaces for chemical compound retrieval and classification. *Knowledge and Information Systems*. 2008; 14: 347–375.
- [16] Borgwardt KM, Ong CS, Schönauer S, Vishwanathan SVN, Smola AJ, Kriegel HP. Protein function prediction via graph kernels. *Bioinformatics*. 2005; 21: i47–i56.
- [17] Sawant S S, Prabukumar M. A review on graph-based semi-supervised learning methods for hyperspectral image classification. *Egyptian Journal of Remote Sensing and Space Science*. 2018; 23: 243–248.
- [18] Shirui Pan, Jia Wu, Xingquan Zhu, Guodong Long, Chengqi Zhang. Task Sensitive Feature Exploration and Learning for Multitask Graph Classification. *IEEE Transactions on Cybernetics*. 2017; 47: 744–758.
- [19] Fu Y, Yang R, Zhang L. Association prediction of CircRNAs and diseases using multi-homogeneous graphs and variational graph auto-encoder. *Computers in Biology and Medicine*. 2016; 151: 106289.
- [20] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*. 2015; 43: e47.
- [21] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)*. 1995; 57: 289–300.
- [22] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*. 2000; 25: 25–29.
- [23] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*. 2000; 28: 27–30.
- [24] Dennis G, Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology*. 2003; 4: P3.
- [25] Theocharidis A, van Dongen S, Enright AJ, Freeman TC. Network visualization and analysis of gene expression data using BioLayout Express(3D). *Nature Protocols*. 2009; 4: 1535–1550.
- [26] Milenković T, Przulj N. Uncovering biological network function via graphlet degree signatures. *Cancer Informatics*. 2008; 6: 257–273.
- [27] Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Research*. 2017; 45: W98–W102.
- [28] Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, *et al.* Proteomics. Tissue-based map of the human proteome. *Science*. 2015; 347: 1260419.
- [29] Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, *et al.* A subcellular map of the human proteome. *Science*. 2017; 356: eaal3321.
- [30] Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhorji G, *et al.* A pathology atlas of the human cancer transcriptome. *Science*. 2017; 357: eaan2507.
- [31] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*. 2012; 2: 401–404.
- [32] Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling*. 2013; 6: pii.
- [33] Carithers LJ, Moore HM. The Genotype-Tissue Expression (GTEx) Project. *Biopreservation and Biobanking*. 2015; 13: 307–308.
- [34] Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BVSK, *et al.* UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia*. 2017; 19: 649–658.
- [35] Selamat SA, Chung BS, Girard L, Zhang W, Zhang Y, Campan M, *et al.* Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. *Genome Research*. 2012; 22: 1197–1211.
- [36] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*. 2018; 46: D1074–D1082.
- [37] Hsin KY, Matsuoka Y, Asai Y, Kamiyoshi K, Watanabe T, Kawaoka Y, *et al.* systemsDock: a web server for network pharmacology-based prediction and analysis. *Nucleic Acids Research*. 2016; 44: W507–W513.
- [38] Hsin KY, Ghosh S, Kitano H. Combining machine learning systems and multiple docking simulation packages to improve docking prediction reliability for network pharmacology. *PLoS ONE*. 2013; 8: e83922.
- [39] GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015; 348: 648–660.
- [40] Reck M, Kriner M, Ulm K, Hessler S, Eberle S. Statistical methods to identify predictive factors. In Crowley J, Ankerst D (ed) *Handbook of statistics in clinical oncology* (pp. 335–345). 2nd edn. Chapman & Hall/CRC: Boca Raton. 2006.

- [41] Zhang X, Fan X, Li X, Wang Y, Zhang Y, Li Y, *et al.* Abnormal TACC3 Expression is an Independent Prognostic Biomarker in Lung Carcinoma. *Frontiers in Bioscience-Landmark*. 2022; 27: 252.
- [42] Terry M. Therneau, Patricia M. Grambsch. *Modeling Survival Data: Extending the Cox Model* (pp. 39–77). Springer: New York. 2000.
- [43] Martin S, Norbert H, Guido S, Willi S. Prognostic factor studies. In Crowley J, Ankerst D(ed). *Handbook of statistics in clinical oncology* (pp. 289–333). 2nd edn. Chapman & Hall/CRC: Boca Raton. 2006.
- [44] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102: 15545–15550.
- [45] Zhu Y, Zhang F, Zhang S, Yi M. Predicting latent lncRNA and cancer metastatic event associations via variational graph auto-encoder. *Methods*. 2016; 211: 1–9.
- [46] Epstein JI, Zelefsky MJ, Sjoberg DD, Nelson JB, Egevad L, Magi-Galluzzi C, *et al.* A Contemporary Prostate Cancer Grading System: A Validated Alternative to the Gleason Score. *European Urology*. 2016; 69: 428–435.
- [47] Shukla S, Cyrt J, Murphy DA, Walczak EG, Ran L, Agrawal P, *et al.* Aberrant Activation of a Gastrointestinal Transcriptional Circuit in Prostate Cancer Mediates Castration Resistance. *Cancer Cell*. 2017; 32: 792–806.e7.
- [48] Pérez-Ibave DC, Burciaga-Flores CH, Elizondo-Riojas MÁ. Prostate-specific antigen (PSA) as a possible biomarker in non-prostatic cancer: A review. *Cancer Epidemiology*. 2018; 54: 48–55.
- [49] Liang C, Wang S, Qin C, Bao M, Cheng G, Liu B, *et al.* TRIM36, a novel androgen-responsive gene, enhances anti-androgen efficacy against prostate cancer by inhibiting MAPK/ERK signaling pathways. *Cell Death & Disease*. 2018; 9: 155.
- [50] Attard G, Murphy L, Clarke NW, Cross W, Jones RJ, Parker CC, *et al.* Abiraterone acetate and prednisolone with or without enzalutamide for high-risk non-metastatic prostate cancer: a meta-analysis of primary results from two randomised controlled phase 3 trials of the STAMPEDE platform protocol. *Lancet*. 2022; 399: 447–460.
- [51] Wu D, Zhang C, Shen Y, Nephew KP, Wang Q. Androgen receptor-driven chromatin looping in prostate cancer. *Trends in Endocrinology and Metabolism*. 2011; 22: 474–480.
- [52] Bastos DA, Antonarakis ES. AR-V7 and treatment selection in advanced prostate cancer: are we there yet? *Precision Cancer Medicine*. 2018; 1: 13.
- [53] Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983; 148: 839–843.
- [54] Hu R, Lu C, Mostaghel EA, Yegnasubramanian S, Gurel M, Tannahill C, *et al.* Distinct transcriptional programs mediated by the ligand-dependent full-length androgen receptor and its splice variants in castration-resistant prostate cancer. *Cancer Research*. 2012; 72: 3457–3462.
- [55] Brown NR, Lowe ED, Petri E, Skamnaki V, Antrobus R, Johnson LN. Cyclin B and cyclin A confer different substrate recognition properties on CDK2. *Cell Cycle*. 2007; 6: 1350–1359.
- [56] Petri ET, Errico A, Escobedo L, Hunt T, Basavappa R. The Crystal Structure of Human Cyclin B. *Cell Cycle*. 2007; 6: 1342–1349.
- [57] Yu Y, Zhang Y, Guan W, Huang T, Kang J, Sheng X, *et al.* Androgen receptor promotes the oncogenic function of overexpressed Jagged1 in prostate cancer by enhancing cyclin B1 expression via Akt phosphorylation. *Molecular Cancer Research*. 2014; 12: 830–842.
- [58] Li Y, Zhang DY, Ren Q, Ye F, Zhao X, Daniels G, *et al.* Regulation of a novel androgen receptor target gene, the cyclin B1 gene, through androgen-dependent E2F family member switching. *Molecular and Cellular Biology*. 2012; 32: 2454–2466.
- [59] Léotoing L, Manin M, Monté D, Baron S, Communal Y, Lours C, *et al.* Crosstalk between androgen receptor and epidermal growth factor receptor-signalling pathways: a molecular switch for epithelial cell differentiation. *Journal of Molecular Endocrinology*. 2007; 39: 151–162.
- [60] Shiao SL, Chu GCY, Chung LWK. Regulation of prostate cancer progression by the tumor microenvironment. *Cancer Letters*. 2016; 380: 340–348.
- [61] Zhao P, Zhang P, Hu W, Wang H, Yu G, Wang Z, *et al.* Upregulation of cyclin B1 plays potential roles in the invasiveness of pituitary adenomas. *Journal of Clinical Neuroscience*. 2017; 43: 267–273.
- [62] Marconett CN, Morgenstern TJ, San Roman AK, Sundar SN, Singhal AK, Firestone GL. BZL101, a phytochemical extract from the *Scutellaria barbata* plant, disrupts proliferation of human breast and prostate cancer cells through distinct mechanisms dependent on the cancer cell phenotype. *Cancer Biology & Therapy*. 2010; 10: 397–405.
- [63] Pilkinton M, Sandoval R, Song J, Ness SA, Colamonici OR. Mip/LIN-9 regulates the expression of B-Myb and the induction of cyclin A, cyclin B, and CDK1. *The Journal of Biological Chemistry*. 2007; 282: 168–175.
- [64] Huang F, Xu X, Xin G, Zhang B, Jiang Q, Zhang C. Cartwheel disassembly regulated by CDK1-cyclin B kinase allows human centriole disengagement and licensing. *The Journal of Biological Chemistry*. 2022; 298: 102658.
- [65] Pei YF, Yin XM, Liu XQ. TOP2A induces malignant character of pancreatic cancer through activating β -catenin signaling pathway. *Biochimica Et Biophysica Acta. Molecular Basis of Disease*. 2018; 1864: 197–207.
- [66] Schaefer-Klein JL, Murphy SJ, Johnson SH, Vasmatzis G, Kovtun IV. Topoisomerase 2 Alpha Cooperates with Androgen Receptor to Contribute to Prostate Cancer Progression. *PLoS ONE*. 2015; 10: e0142327.
- [67] de Resende MF, Vieira S, Chinen LTD, Chiappelli F, da Fonseca FP, Guimarães GC, *et al.* Prognostication of prostate cancer based on TOP2A protein and gene assessment: TOP2A in prostate cancer. *Journal of Translational Medicine*. 2013; 11: 36.
- [68] Sønderstrup IMH, Nygård SB, Poulsen TS, Linnemann D, Stenvang J, Nielsen HJ, *et al.* Topoisomerase-1 and -2A gene copy numbers are elevated in mismatch repair-proficient colorectal cancers. *Molecular Oncology*. 2015; 9: 1207–1217.
- [69] Deng S, Yan T, Nikolova T, Fuhrmann D, Nemecek A, Gödtel-Armbrust U, *et al.* The catalytic topoisomerase II inhibitor dexrazoxane induces DNA breaks, ATF3 and the DNA damage response in cancer cells. *British Journal of Pharmacology*. 2015; 172: 2246–2257.
- [70] Chen T, Sun Y, Ji P, Kopetz S, Zhang W. Topoisomerase II α in chromosome instability and personalized cancer therapy. *Oncogene*. 2015; 34: 4019–4031.
- [71] Pan HW, Su HH, Hsu CW, Huang GJ, Wu TTL. Targeted TPX2 increases chromosome missegregation and suppresses tumor cell growth in human prostate cancer. *OncoTargets and Therapy*. 2017; 10: 3531–3543.
- [72] Zou J, Huang RY, Jiang FN, Chen DX, Wang C, Han ZD, *et al.* Overexpression of TPX2 is associated with progression and prognosis of prostate cancer. *Oncology Letters*. 2018; 16: 2823–2832.
- [73] Isayeva T, Moore LD, Chanda D, Chen D, Ponnazhagan S. Tumoristatic effects of endostatin in prostate cancer is dependent on androgen receptor status. *The Prostate*. 2009; 69: 1055–1066.
- [74] Wang L, Tang H, Thayanyithy V, Subramanian S, Oberg AL,

- Cunningham JM, *et al.* Gene networks and microRNAs implicated in aggressive prostate cancer. *Cancer Research*. 2009; 69: 9490–9497.
- [75] Cuzick J, Swanson GP, Fisher G, Brothman AR, Berney DM, Reid JE, *et al.* Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. *The Lancet. Oncology*. 2011; 12: 245–255.
- [76] Tinholt M, Garred Ø, Borgen E, Beraki E, Schlichting E, Kristensen V, *et al.* Subtype-specific clinical and prognostic relevance of tumor-expressed F5 and regulatory F5 variants in breast cancer: the CoCaV study. *Journal of Thrombosis and Haemostasis*. 2018; 16: 1347–1356.
- [77] Ifere GO, Desmond R, Demark-Wahnefried W, Nagy TR. Apolipoprotein E gene polymorphism influences aggressive behavior in prostate cancer cells by deregulating cholesterol homeostasis. *International Journal of Oncology*. 2013; 43: 1002–1010.
- [78] Yencilek F, Yilmaz SG, Yildirim A, Gormus U, Altinkilic EM, Dalan AB, *et al.* Apolipoprotein E Genotypes in Patients with Prostate Cancer. *Anticancer Research*. 2016; 36: 707–711.
- [79] Ruscica M, Dozio E, Motta M, Magni P. Role of neuropeptide Y and its receptors in the progression of endocrine-related cancer. *Peptides*. 2007; 28: 426–434.
- [80] Alshalalfa M, Nguyen PL, Beltran H, Chen WS, Davicioni E, Zhao SG, *et al.* Transcriptomic and Clinical Characterization of Neuropeptide Y Expression in Localized and Metastatic Prostate Cancer: Identification of Novel Prostate Cancer Subtype with Clinical Implications. *European Urology Oncology*. 2019; 2: 405–412.
- [81] Singh N, Kumble Bhat V, Tiwari A, Kodaganur SG, Tontanahal SJ, Sarda A, *et al.* A homozygous mutation in TRIM36 causes autosomal recessive anencephaly in an Indian family. *Human Molecular Genetics*. 2017; 26: 1104–1114.
- [82] Balint I, Müller A, Nagy A, Kovacs G. Cloning and characterisation of the RBCC728/TRIM36 zinc-binding protein from the tumor suppressor gene region at chromosome 5q22.3. *Gene*. 2004; 332: 45–50.
- [83] Miyajima N, Maruyama S, Nonomura K, Hatakeyama S. TRIM36 interacts with the kinetochore protein CENP-H and delays cell cycle progression. *Biochemical and Biophysical Research Communications*. 2009; 381: 383–387.