Mathematical Biosciences and Engineering

MBE, 18 (4): 3180–3196.
DOI: 10.3934/mbe.2021158
Received: 23 February 2021
Accepted: 02 April 2021
Published: 06 April 2021

http://www.aimspress.com/journal/MBE

Research article

Bioinformatics analysis identified hub genes in prostate cancer tumorigenesis and metastasis

Peng Gu[1,2], Dongrong Yang[1,*], Jin Zhu[1], Minhao Zhang[2] and Xiaoliang He[2]

[1] Department of Urology, Second Affiliated Hospital of Soochow University, 1055 Sanxiang Road, Suzhou 215000, China
[2] Department of Urology, Wuxi Xishan People's Hospital, 1128 Dacheng Road, Wuxi 214000, China

* **Correspondence:** Email: doc_ydr@163.com.

Abstract: *Objective:* Prostate cancer (PCa) is the most frequent cancer found in males worldwide, and its mortality rate is increasing every year. To discover key molecular changes in PCa development and metastasis, we analyzed microarray data of localized PCa, metastatic PCa and normal prostate tissue samples from clinical specimens. *Methods:* Gene expression profiling datasets of PCa were analyzed online. The DAVID was used to perform GO functional and KEGG pathway enrichment analyses. CytoHubba in Cytoscape software was applied to identify hub genes. Survival data were downloaded from GEPIA. Gene expression data were obtained from ONCOMINE and UALCAN. *Results:* We obtained 4 sets of differentially expressed genes (DEGs), DEGs 1: a comparison of the gene expression between 4 normal prostate and 5 localized PCa samples in GSE27616; DEGs 2: a comparison of the gene expression between 6 normal prostate and 7 localized PCa samples in GSE3325; DEGs 3: a comparison of the gene expression between 5 localized PCa and 4 metastatic PCa samples in GSE27616; DEGs 4: a comparison of the gene expression between 7 localized PCa and 6 metastatic PCa samples in GSE3325. A comparison of these 4 sets of genes revealed 51 overlapped genes. GO function analysis revealed enrichment of the 51 DEGs in functions related to the proteinaceous extracellular matrix and centrosome, protein homodimerization activity and chromatin binding were the main functions of these genes, which participated in regulating cell division, mitotic nuclear division, proteinaceous extracellular matrix, cell adhesion and apoptotic process. KEGG pathway analysis indicated that these identified DEGs were mainly enriched in progesterone-mediated oocyte maturation, oocyte meiosis and cell cycle. We defined the 16 genes with the highest degree of connectivity as the hub genes in the 51 overlapped DEGs. Cox regression revealed *TOP2A*, *CCNB2*, *BUB1*, *CDK1* and *EZH2* were related to Disease-free survival (DFS). The expression levels of the 5 genes were 2.232-, 1.786-, 2.303-, 1.699-, and 1.986-fold

higher in PCa than the levels in normal tissues, respectively ($P < 0.05$). We obtained 20 hub genes from DEGs by the comparison of normal prostate tissue vs. localized cancer tissue. Among them, *KIF20A*, *CDKN3*, *PBK* and *CDCA2*, were expressed higher in PCa than in normal tissues, and were associated with the DFS of PCa patients. Meanwhile, we obtained 20 hub genes from DEGs by the comparison of localized cancer tissue vs. metastatic cancer tissue. Cox regression revealed *PLK1*, *CCNA2* and *CDC20*, were associated with both the DFS and overall survival of PCa patients. *Conclusions:* The results suggested that the functions of *KIF20A*, *CDKN3*, *PBK* and *CDCA2* may contribute to PCa development and the functions of *PLK1*, *CCNA2* and *CDC20* may contribute to PCa metastasis. Meanwhile, the functions of *TOP2A*, *CCNB2*, *BUB1*, *CDK1* and *EZH2* may contribute to both PCa development and metastasis.

Keywords: bioinformatics analysis; differentially expressed genes; prostate cancer; tumorigenesis; metastasis

1. Introduction

Prostate cancer (PCa) is the second most frequent cancer found in men worldwide (after lung cancer), with an estimate of 248,530 new cases and 34,130 deaths in the United States in 2021 [1]. And PCa is the second leading cause of cancer death in American men and survival rates are low for prostate cancers that advance to metastatic castrationresistant prostate cancer (mCRPC). Despite recent advances and a range of treatment options for mCRPC, outcomes are varied and clinicians are not able to predict response to the available therapies [2]. Therefore, the need to more accurately identify lethal PCa, in an effort to personalize medicine for those in need, has led to a large-scale push for biomarker development in the field.

Many efforts based on microarray have been done in order to select the key genes associated with PCa. Due to microarray technology, it is easier to analyze the genetic alterations underlying PCa development and progression. In addition, through bioinformatics tools it is possible to identify new biomarkers and to construct networks that could be valuable for the management of PCa patients[3]. However, to date, it has been difficult to identify key genes related and distinguished to PCa development and metastasis from microarray data. To discover key molecules active in PCa development and metastasis, we analyzed PCa data from microarray data. Our results suggested *KIF20A*, *CDKN3*, *PBK* and *CDCA2* may contribute to PCa development, *PLK1*, *CCNA2* and *CDC20* may contribute to PCa metastasis, *TOP2A*, *CCNB2*, *BUB1*, *CDK1* and *EZH2* may contribute to PCa development and metastasis.

2. Materials and methods

2.1. Online data

The gene expression profiling datasets of PCa were analyzed online (GEO; https://www.ncbi.nlm.nih.gov/geo/geo2r/). 4 normal prostate, 5 localized prostate cancer, and 4 metastatic prostate cancer samples were measured in the array GSE27616. 6 normal prostate, 7 localized prostate cancer, and 6 metastatic prostate cancer samples were measured in the array

GSE3325.

2.2. Identifying differentially expressed genes (DEGs)

To analyze the microarray data, we defined DEGs 1: a comparison of the gene expression between 4 normal prostate and 5 localized prostate cancer samples in GSE27616; DEGs 2: a comparison of the gene expression between 6 normal prostate and 7 localized prostate cancer samples in GSE3325; DEGs 3: a comparison of the gene expression between 5 localized prostate cancer and 4 metastatic prostate cancer samples in GSE27616; DEGs 4: a comparison of the gene expression between 7 localized prostate cancer and 6 metastatic prostate cancer samples in GSE3325. We acquired the overlapped genes between DEGs 1 and DEGs 2 to identify genes involved with tumorigenesis, and defined genes cluster NC represented normal prostate tissue vs. localized cancer tissue. We acquired the overlapped genes between DEGs 3 and DEGs 4 to screen genes that promote tumor metastasis, and defined genes cluster LM represented localized cancer tissue vs. metastatic cancer tissue. DEGs were screened by P value and fold change (FC), restricted by P value < 0.05 and $\log|FC| \geq 1$. FunRich software (FunRich_3.1.3) was used to identify overlapped DEGs. The upregulated and downregulated genes were measured, respectively.

2.3. Mergine data

We proposed the two methods to process the clusters NC and LM: (1) tumorigenesis and metastasis were promoted by the same genes or proteins, the overlapped genes between NC and LM were analyzed to perform Gene Ontology (GO) functional and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis; (2) tumorigenesis and metastasis were contributed by different genes, we would find key genes from clusters NC and LM individually, and NC and LM genes were individually analyzed to perform GO and KEGG pathway analysis and retrieve interacting genes.

2.4. GO functional and KEGG pathway analysis

The Database for Annotation, Visualization, and Integrated Discovery (DAVID, https://david.ncifcrf.gov/) was used to perform GO functional and KEGG pathway enrichment analyses. P value < 0.05 was set as the cut-off criterion.

2.5. Identification of hub genes

Search Tool for the Retrieval of Interacting Genes (STRING) is a biological database (https://string-db.org) for constructing protein-protein interaction (PPI) networks, providing a system-wide view of interactions between each member. DEGs was constructed using STRING database, and an interaction with a combined score > 0.4 was considered statistically significant. Then, we established PPI network using Cytoscape software (Cytoscape_v3.7.2), which visually explores biomolecular interaction networks composed of proteins, genes and other molecules. CytoHubba in Cytoscape software was applied to screen the hub genes ranked by degree. And hub genes represented genes with a large degree and were hypothesized to play a key role in the network.

2.6. Survival analysis

Disease-free survival (DFS) and overall survival (OS) were analyzed by Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn/detail/).

2.7. Gene expression data

Gene expression data were obtained from ONCOMINE website (https://www.oncomine.org/) and UALCAN (http://ualcan.path.uab.edu/analysis.html).

3. Results

3.1. DEGs screening

We identified 1881 and 2389 DEGs by analysis of normal prostate tissue vs. localized cancer tissue in GSE27616 and GSE3325, respectively. 3246 and 6331 DEGs by analysis of localized cancer tissue vs. metastatic cancer tissue in GSE27616 and GSE3325, respectively. A comparison of these 4 sets of genes revealed 51 overlapped genes, including 18 upregulated and 33 downregulated DEGs (Figure 1, Table 1). We next analyzed these genes by performing two kinds of functional analysis.

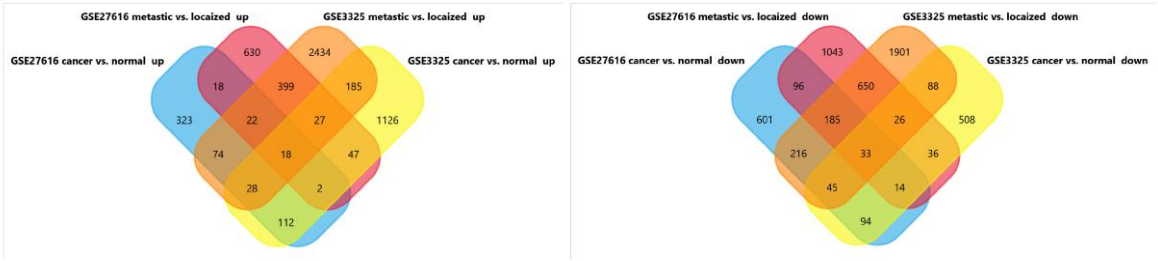

Figure 1. 1881 and 2389 differentially expressed genes (DEGs) were found by comparing normal prostate tissue and localized cancer tissue samples in GSE27616 and GSE3325, respectively. 3246 and 6331 DEGs were found by comparing localized cancer tissue and metastatic cancer tissue samples in GSE27616 and GSE3325, respectively. 18 and 33 overlapped genes were found in up- and downregulation genes.

Table 1. 51 overlapped genes in 4 sets of DEGs.

	gene symbol
upregulated DEGs	*TOP2A, MELK, WISP1, CENPF, BIRC5, BUB1, CDK1, NUSAP1, CCNB2, EZH2, DLGAP5, TROAP, KIAA0101, HMMR, CDCA3, NCAPG, CENPA, TMEM132A*
downregulated DEGs	*PGM5-AS1, SPOCK3, DST, COL4A6, ACTC1, PGR, SOX5, LDB3, GPD1L, CAV1, CLIC6, C2orf40, SLC8A1, SPON1, SERP2, KIAA1210, SPTLC3, RCAN2, ZBTB20, TENM2, ADAMTS1, CFD, LSAMP, FGFR2, DNAJB5, PCDH9, LINC00844, APCDD1, NDRG2, ZNF853, CRISPLD2, NAALADL1, VSTM4*

DEGs: differentially expressed genes.

3.2. GO and KEGG pathway analysis

In the first analysis, the 18 upregulated and 33 downregulated genes that were differentially expressed in both the comparison of normal prostate tissue and localized cancer tissue and the comparison of localized cancer tissue and metastatic cancer tissue were analyzed using the DAVID. GO functional and KEGG pathway enrichment analyses were performed. GO function analysis revealed enrichment of these DEGs in functions related to the proteinaceous extracellular matrix and centrosome. There was an enrichment of genes involved with regulating cell division, protein homodimerization activity, mitotic nuclear division, proteinaceous extracellular matrix, chromatin binding, cell adhesion and apoptotic process. KEGG pathway analysis indicated that the identified DEGs were mainly enriched in progesterone-mediated oocyte maturation, oocyte meiosis and cell cycle. The 20 most enriched classes based on GO function analysis and the 3 most enriched KEGG pathways were listed in Table 2.

Table 2. GO and KEGG pathway analysis of Differentially Expressed Genes.

	Category	Term	Count	%	P value	FDR
GO analysis	GOTERM_BP_DIRECT	GO:0051301~cell division	7	13.7254902	2.84E-04	0.074656255
	GOTERM_BP_DIRECT	GO:0000086~G2/M transition of mitotic cell cycle	5	9.803921569	4.47E-04	0.074656255
	GOTERM_BP_DIRECT	GO:0007067~mitotic nuclear division	6	11.76470588	4.58E-04	0.074656255
	GOTERM_CC_DIRECT	GO:0005578~proteinaceous extracellular matrix	6	11.76470588	5.55E-04	0.069917754
	GOTERM_BP_DIRECT	GO:0007062~sister chromatid cohesion	4	7.843137255	0.002473069	0.302332645
	GOTERM_MF_DIRECT	GO:0003682~chromatin binding	6	11.76470588	0.003359038	0.416520711
	GOTERM_CC_DIRECT	GO:0005938~cell cortex	4	7.843137255	0.003686069	0.231503818
	GOTERM_CC_DIRECT	GO:0005876~spindle microtubule	3	5.882352941	0.005511996	0.231503818
	GOTERM_BP_DIRECT	GO:0007155~cell adhesion	6	11.76470588	0.006743272	0.659491957
	GOTERM_MF_DIRECT	GO:0008201~heparin binding	4	7.843137255	0.008327156	0.460492788
	GOTERM_CC_DIRECT	GO:0000775~chromosome, centromeric region	3	5.882352941	0.009108061	0.286903911
	GOTERM_MF_DIRECT	GO:0042803~protein homodimerization activity	7	13.7254902	0.011140955	0.460492788
	GOTERM_BP_DIRECT	GO:0016055~Wnt signaling pathway	4	7.843137255	0.012879092	0.773674476
	GOTERM_BP_DIRECT	GO:0070314~G1 to G0 transition	2	3.921568627	0.013034542	0.773674476
	GOTERM_BP_DIRECT	GO:0000226~microtubule cytoskeleton organization	3	5.882352941	0.014870884	0.773674476
	GOTERM_BP_DIRECT	GO:0008283~cell proliferation	5	9.803921569	0.015179728	0.773674476
	GOTERM_BP_DIRECT	GO:0006915~apoptotic process	6	11.76470588	0.015821564	0.773674476
	GOTERM_CC_DIRECT	GO:0045120~pronucleus	2	3.921568627	0.017538638	0.399420613
	GOTERM_CC_DIRECT	GO:0031410~cytoplasmic vesicle	4	7.843137255	0.021389738	0.399420613
	GOTERM_CC_DIRECT	GO:0005813~centrosome	5	9.803921569	0.022190034	0.399420613
KEGG pathway	KEGG_PATHWAY	hsa04914:Progesterone-mediated oocyte maturation	4	7.843137255	7.97E-04	0.034252658
	KEGG_PATHWAY	hsa04114:Oocyte meiosis	4	7.843137255	0.001615269	0.034728293
	KEGG_PATHWAY	hsa04110:Cell cycle	3	5.882352941	0.029035349	0.416173338

GO: gene ontology; BP: biological process; CC: cellular component; MF: molecular function.

In the second analysis, we focused on the DEGs identified by the comparison of normal prostate and localized cancer tissues or those identified by the comparison of localized cancer and metastatic cancer tissue samples. Analysis of DEGs from the normal prostate tissue vs. localized cancer tissue comparison should reflect key genes participating in tumorigenesis or PCa development. GO function analysis of these genes found high enrichment of functions related to plasma membrane and extracellular space, heme binding and heparin binding were the main functions of these genes, which participated in oxidation-reduction process, cell adhesion, and response to drug. KEGG pathway analysis indicated enrichment of these genes in protein digestion and absorption, and TGF-beta signaling pathway (Table 3). We next analyzed the DEGs identified by the comparison of localized cancer tissue and metastatic cancer tissue samples, which should include genes related to PCa metastasis. GO function analysis revealed enrichment of these genes in functions related to the extracellular exosome, extracellular region and extracellular space, and analysis of molecular function showed enrichment in protein binding. The most relevant enriched biological processes were cell adhesion, cell division, and cell proliferation and KEGG pathway analysis indicated enrichment of these genes in pathways in cancer, HTLV-I infection, cell cycle and cell adhesion molecules (CAMs) (Table 3).

3.3. *Hub gene analysis*

We used STRING for investigating and integrating interaction between proteins. The PPI network of overlapped DEGs between NC and LM was constructed and data were exported for further analysis by Cytoscape (Figure 2a). And a total of 16 genes were identified as hub genes with degrees ≥ 14, shown in Figure 2b. Also, the top 20 hub genes with the highest degree of connectivity in clusters NC and LM were shown in Figures 2c and 2d, respectively.

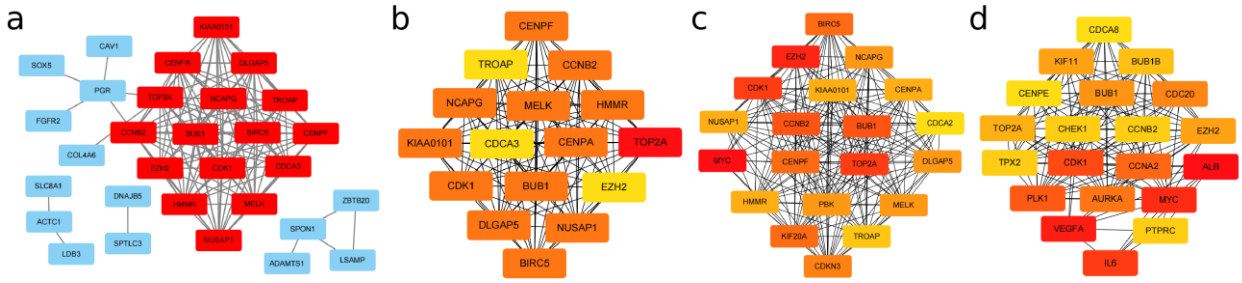

Figure 2. (a) The PPI network of overlapped DEGs between NC (normal prostate tissue vs. localized cancer tissue) and LM (localized cancer tissue vs. metastatic cancer tissue) was constructed using Cytoscape (upregulated genes are marked in red; downregulated genes are marked in blue). (b) 16 hub genes were discovered from overlapped DEGs between NC and LM. (c) 20 hub genes were discovered from NC. (d) 20 hub genes were discovered from LM.

Table3. GO and KEGG pathway analysis of differentially expressed genes in prostate cancer development and metastasis.

	Prostate cancer development Category	Term	Count	%	P value	FDR	Prostate cancer metastasis Category	Term	Count	%	P value	FDR
GO analysis	GOTERM_CC_DIRECT	GO:0005578~proteinaceous extracellular matrix	17	4.913294798	1.51E-05	0.003982498	GOTERM_CC_DIRECT	GO:0005615~extracellular space	151	11.11929308	1.24E-09	7.42E-07
	GOTERM_BP_DIRECT	GO:0042475~odontogenesis of dentin-containing tooth	7	2.023121387	3.52E-04	0.586071751	GOTERM_BP_DIRECT	GO:0051301~cell division	55	4.050073638	2.98E-08	1.30E-04
	GOTERM_CC_DIRECT	GO:0005615~extracellular space	41	11.84971098	3.99E-04	0.052470784	GOTERM_MF_DIRECT	GO:0005515~protein binding	706	51.98821797	3.68E-08	3.56E-05
	GOTERM_MF_DIRECT	GO:0020037~heme binding	10	2.89017341	5.77E-04	0.26984829	GOTERM_MF_DIRECT	GO:0008201~heparin binding	33	2.430044183	5.63E-08	3.56E-05
	GOTERM_BP_DIRECT	GO:0002053~positive regulation of mesenchymal cell proliferation	5	1.445086705	9.53E-04	0.586071751	GOTERM_CC_DIRECT	GO:0031012~extracellular matrix	48	3.53460972	5.67E-08	1.70E-05
	GOTERM_BP_DIRECT	GO:0001755~neural crest cell migration	6	1.734104046	0.001014555	0.586071751	GOTERM_CC_DIRECT	GO:0009986~cell surface	70	5.154639175	4.52E-07	9.05E-05
	GOTERM_CC_DIRECT	GO:0016323~basolateral plasma membrane	11	3.179190751	0.001050682	0.092109794	GOTERM_BP_DIRECT	GO:0007155~cell adhesion	62	4.565537555	8.88E-07	0.001389998
	GOTERM_CC_DIRECT	GO:1903561~extracellular vesicle	6	1.734104046	0.001561555	0.102671884	GOTERM_BP_DIRECT	GO:0001666~response to hypoxia	32	2.35640648	9.54E-07	0.001389998
	GOTERM_BP_DIRECT	GO:0048701~embryonic cranial skeleton morphogenesis	5	1.445086705	0.001874567	0.812156124	GOTERM_CC_DIRECT	GO:0070062~extracellular exosome	253	18.63033873	1.97E-06	2.95E-04
	GOTERM_CC_DIRECT	GO:0005886~plasma membrane	93	26.878661272	0.002175661	0.114439759	GOTERM_BP_DIRECT	GO:0002576~platelet degranulation	23	1.693667158	2.00E-06	0.002186739
	GOTERM_BP_DIRECT	GO:0007155~cell adhesion	18	5.202312139	0.002608742	0.904189938	GOTERM_CC_DIRECT	GO:0005576~extracellular region	158	11.634757	2.77E-06	3.32E-04
	GOTERM_CC_DIRECT	GO:0005911~cell-cell junction	10	2.89017341	0.002780496	0.121878392	GOTERM_BP_DIRECT	GO:0007062~sister chromatid cohesion	22	1.620029455	7.38E-06	0.006452155

Continued on next page

	Prostate cancer development Category	Term	Count	%	P value	FDR	Prostate cancer metastasis Category	Term	Count	%	P value	FDR
KEGG pathway	GOTERM_BP_DIRECT	GO:0007154~cell communication	5	1.445086705	0.004002576	1	GOTERM_BP_DIRECT	GO:0007067~mitotic nuclear division	38	2.798232695	9.69E-06	0.007062396
	GOTERM_MF_DIRECT	GO:0045545~syndecan binding	3	0.867052023	0.004158161	0.648673134	GOTERM_MF_DIRECT	GO:0005178~integrin binding	22	1.620029455	1.05E-05	0.004418724
	GOTERM_MF_DIRECT	GO:0016174~NAD(P)H oxidase activity	3	0.867052023	0.004158161	0.648673134	GOTERM_CC_DIRECT	GO:0030018~Z disc	23	1.693667158	1.51E-05	0.001511302
	GOTERM_BP_DIRECT	GO:0030574~collagen catabolic process	6	1.734104046	0.004859514	1	GOTERM_BP_DIRECT	GO:0030198~extracellular matrix organization	32	2.35640648	1.56E-05	0.0097292
	GOTERM_CC_DIRECT	GO:0005938~cell cortex	8	2.312138728	0.005103026	0.191727963	GOTERM_BP_DIRECT	GO:0008283~cell proliferation	49	3.608247423	1.82E-05	0.009935907
	GOTERM_MF_DIRECT	GO:0008201~heparin binding	9	2.601156069	0.006242515	0.730374313	GOTERM_BP_DIRECT	GO:0044344~cellular response to fibroblast growth factor stimulus	11	0.810014728	2.21E-05	0.010755347
	GOTERM_BP_DIRECT	GO:0042493~response to drug	13	3.757225434	0.006583695	1	GOTERM_MF_DIRECT	GO:0042803~protein homodimerization activity	82	6.038291605	2.34E-05	0.007414289
	GOTERM_BP_DIRECT	GO:0055114~oxidation-reduction process	20	5.780346821	0.007057009	1	GOTERM_BP_DIRECT	GO:0030336~negative regulation of cell migration	20	1.47275405	2.64E-05	0.011530048
	KEGG_PATHWAY	hsa04974:Protein digestion and absorption	7	2.023121387	0.004206668	0.803473664	KEGG_PATHWAY	hsa05200:Pathways in cancer	58	4.270986745	2.73E-06	4.52E-04
	KEGG_PATHWAY	hsa04350:TGF-beta signaling pathway	6	1.734104046	0.015350804	1	KEGG_PATHWAY	hsa05332:Graft-versus-host disease	13	0.957290133	3.65E-06	4.52E-04
							KEGG_PATHWAY	hsa04940:Type I diabetes mellitus	14	1.030927835	1.11E-05	9.19E-04
							KEGG_PATHWAY	hsa04110:Cell cycle	25	1.840942563	2.61E-05	0.001511671

Continued on next page

Prostate cancer development Category	Term	Count	%	P value	FDR	Prostate cancer metastasis Category	Term	Count	%	P value	FDR
						KEGG_PATHWAY	hsa05166:HTLV-I infection	40	2.9455081	3.05E-05	0.001511671
						KEGG_PATHWAY	hsa04114:Oocyte meiosis	23	1.693667158	3.85E-05	0.001591416
						KEGG_PATHWAY	hsa05416:Viral myocarditis	15	1.104565538	9.11E-05	0.003228693
						KEGG_PATHWAY	hsa05323:Rheumatoid arthritis	19	1.399116348	1.26E-04	0.003920294
						KEGG_PATHWAY	hsa04514:Cell adhesion molecules (CAMs)	25	1.840942563	2.43E-04	0.006698092
						KEGG_PATHWAY	hsa05144:Malaria	13	0.957290133	2.97E-04	0.007354055

GO: gene ontology; BP: biological process; CC: cellular component; MF: molecular function.

3.4. Clinical analysis

We found *TOP2A*, *CCNB2*, *BUB1*, *CDK1* and *EZH2* were appeared in Figure 2b, Figure 2c and Figure 2d at the same time. The 5 hub genes were all associated with the DFS of PCa patients (Figure 3a–e). *KIF20A*, *CDKN3*, *PBK* and *CDCA2* were only appeared in Figure 2c, and the 4 hub genes were also all associated with the DFS of PCa patients (Figure 3f–i). *ALB*, *VEGFA*, *IL6*, *PLK1*, *CCNA2*, *AURKA*, *CDC20*, *KIF11*, *BUB1B*, *CHEK1*, *PTPRC*, *TPX2*, *CENPE* and *CDCA8* were only appeared in Figure 2d, and just *ALB*, *PLK1*, *CCNA2*, *AURKA*, *CDC20*, *BUB1B*, *CHEK1*, *TPX2*, *CENPE* and *CDC*A8 were associated with the DFS of PCa patients (Figure 3j–s). Among them, we next found an association of *PLK1*, *CCNA2* and *CDC20* with overall survival of PCa patients (Figure 3t–v).

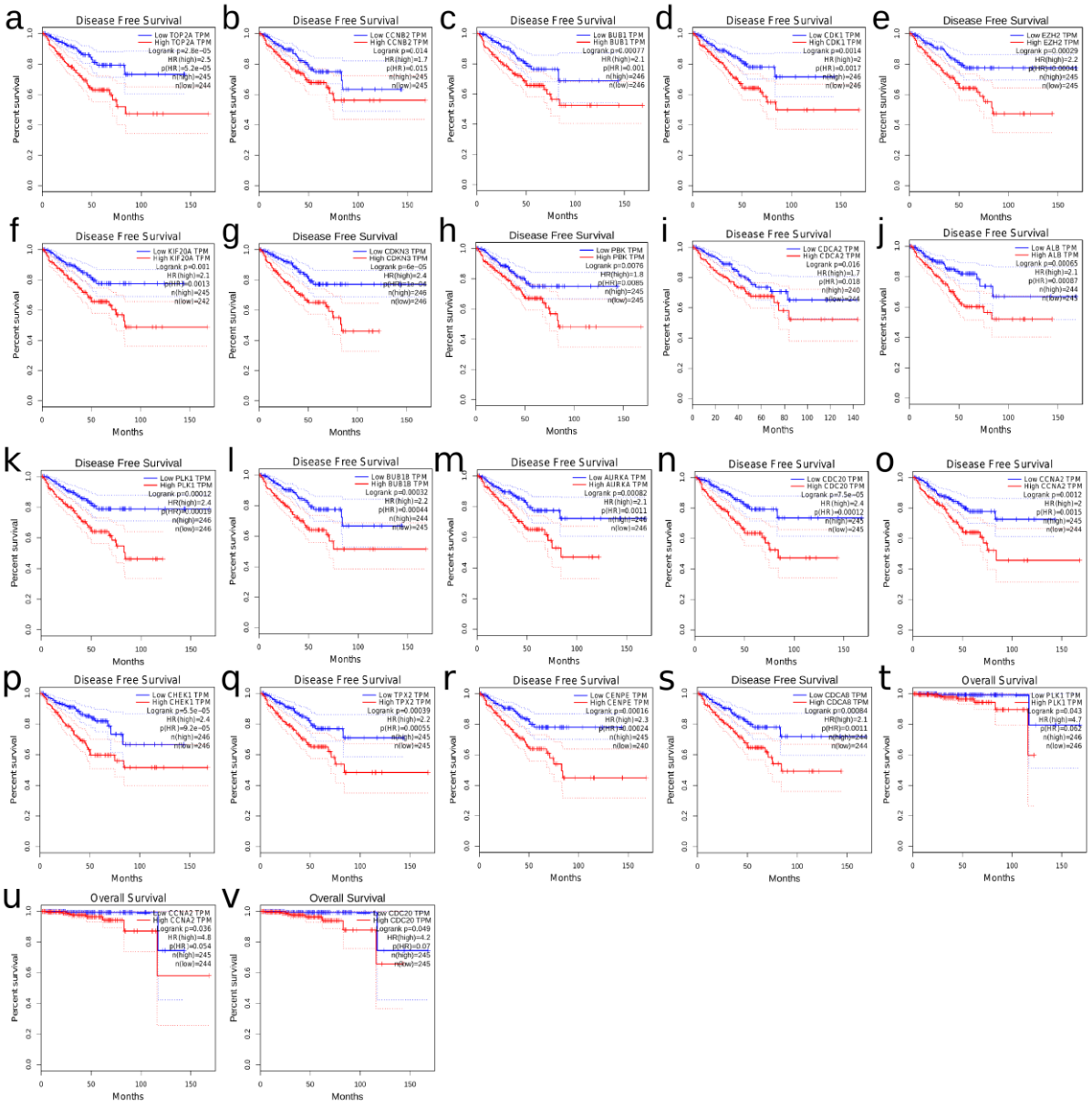

Figure 3. Disease-free survival (DFS) and Overall survival (OS) analysis for hub genes.

3.5. Gene expression in Oncomine analysis

Oncomine analysis of prostate cancer vs. normal tissue in Grasso Prostate Statistics showed that the expression levels of *TOP2A*, *CCNB2*, *BUB1*, *CDK1* and *EZH2* were 2.232-, 1.786-, 2.303-, 1.699-, and 1.986-fold higher in prostate cancer than in normal tissues (P = 8.51E-6, 6.94E-6, 1.14E-5, 0.001, 4.59E-11). The expression levels of *KIF20A*, *CDKN3*, *PBK* and *CDCA2* in cluster NC were 1.875-, 1.778-, 2.337-, and 1.889-fold higher in prostate cancer than in normal tissues (P = 5.87E-5, 8.39E-4, 1.86E-6, 7.89E-5). The expression levels of *PLK1*, *CCNA2*, *CDC20* in cluster LM were 1.326-, 1.231-, and 1.047-fold higher in prostate cancer than in normal tissues (P = 3.91E-4, 0.036, 0.253) (Figure 4, Table 4).

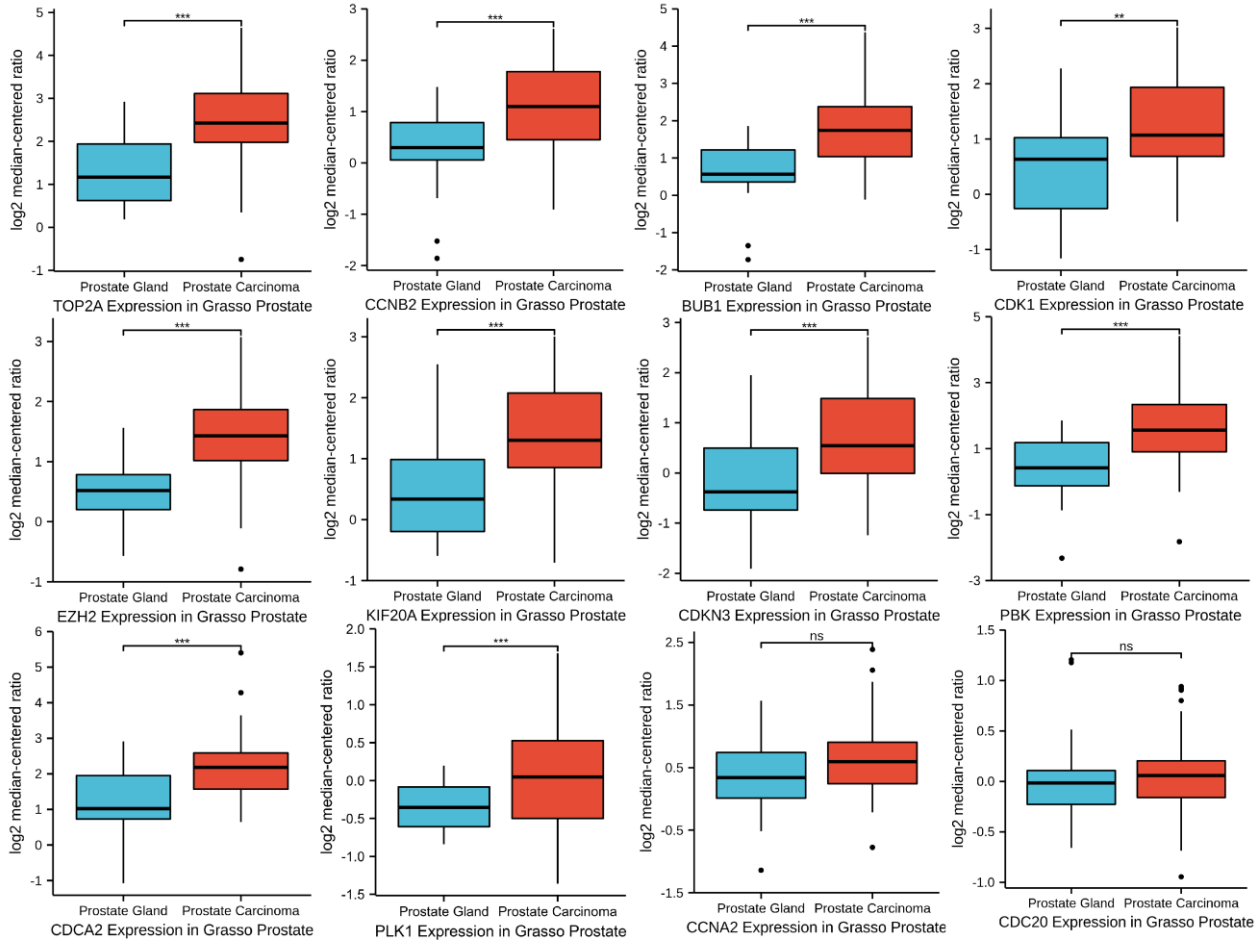

Figure 4. Oncomine analysis of prostate cancer vs. normal tissue in Grasso Prostate Statistics.

3.6. *Gene expression in UALCAN*

UALCAN is a website that helps analyze, integrate and discover cancer transcriptomic data and deep analyses of TCGA gene expression information. It allows us to offer the differential expression analyses on normal and PCa tissues. It contained 497 PCa tissues and 52 normal prostate samples. As Figure 5 demonstrated, all hub genes significantly overexpressed in PCa tissues than in normal tissues (P < 0.001).

Table 4. Gene expression in prostate cancer vs.normal tissue.

Hub Genes	Fold change (Cancer vs.normal)	P vaule
TOP2A	2.232	8.51E-6
CCNB2	1.786	6.94E-6
BUB1	2.303	1.14E-5
CDK1	1.699	0.001
EZH2	1.986	4.59E-11
KIF20A	1.875	5.87E-5
CDKN3	1.778	8.39E-4
PBK	2.337	1.86E-6
CDCA2	1.889	7.89E-5
PLK1	1.326	3.91E-4
CCNA2	1.231	0.036
CDC20	1.047	0.253

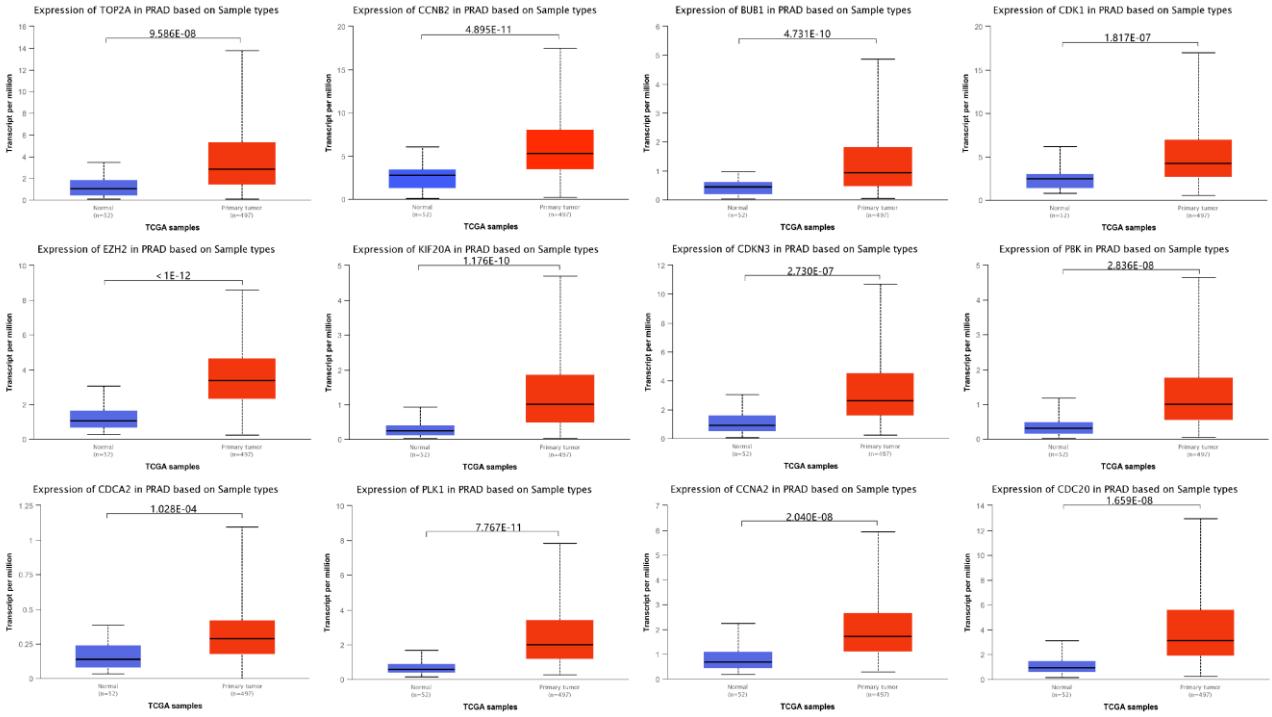

Figure 5. Expression levels of hub genes in prostate cancer and normal prostate tissues (TCGA database, $P < 0.001$).

4. Discussion

In this analysis, we defined DEGs for the NC comparison of normal prostate tissue vs. localized cancer tissue and for the LM comparison of localized cancer tissue vs. metastatic cancer tissue in GSE27616 and GSE3325, and considered the identified DEGs contributing to prostate cancer development and contributing to prostate cancer metastasis, respectively. Overlapped genes found in both NC and LM analyses affect both prostate cancer development and metastasis. GO function

analysis revealed enrichment of the 51 DEGs overlapped in clusters NC and LM in functions related to the proteinaceous extracellular matrix and centrosome, protein homodimerization activity and chromatin binding were the main functions of these genes, which participated in regulating cell division, mitotic nuclear division, proteinaceous extracellular matrix, cell adhesion and apoptotic process. KEGG pathway analysis indicated that these identified DEGs were mainly enriched in progesterone-mediated oocyte maturation, oocyte meiosis and cell cycle. GO function analysis discovered DEGs in either cluster NC or LM were both enriched in extracellular space, and participated in cell adhesion. Their molecular function and KEGG pathway were different. Heme binding and heparin binding were the main functions of DEGs in cluster NC, and protein binding was the main function of DEGs in cluster LM. KEGG pathway analysis indicated that DEGs in cluster NC were mainly enriched in protein digestion and absorption, and TGF-beta signaling pathway, DEGs in cluster LM were mainly enriched in pathways in cancer, HTLV-I infection, cell cycle and CAMs.

We found that *KIF20A*, *CDKN3*, *PBK* and *CDCA2* in cluster NC, expressed higher in prostate cancer than in normal tissues, and were associated with the DFS of prostate cancer patients, maybe associated with prostate cancer development. Although, their expressions were not all elevated in prostate cancer tissue than in normal tissues significantly, *PLK1*, *CCNA2* and *CDC20* in cluster LM, were associated with both the DFS and overall survival of prostate cancer patients, may be associated with prostate cancer metastasis. It was interesting that our analysis revealed 5 genes, *TOP2A*, *CCNB2*, *BUB1*, *CDK1* and *EZH2*, overlapped in clusters NC and LM, expressed higher in prostate cancer than in normal tissues, and were associated with the DFS of prostate cancer patients, which may be associated with both prostate cancer development and metastasis.

KIF20A, known as mitotic kinesin-like protein, is a member of the kinesin family protein which is located in the Golgi apparatus and participates in the dynamics of organelle and cell division. Studies [4] have found that KIF20A was abnormally highly expressed in prostate cancer tissues and was associated with adverse prognosis of PCa patients and knockdown of KIF20A could inhibit the proliferation and invasion of prostate cancer cell. Cyclin-dependent kinase inhibitor 3 (CDKN3, also called CDI1 or KAP) is a member of the dual specificity protein phosphatase family, plays an important role in regulating cell division. It has been reported [5] that CDKN3 was over-expressed in PCa and negatively correlated with disease free survival time of PCa. Down-regulation of CDKN3 inhibited proliferation, promoted apoptosis and invasion of PCa cells, and CDKN3 may regulate these biological progresses through associating with cell cycle and DNA replication signal pathways. PDZ-binding kinase (PBK) or T-LAK cell-originated protein kinase (TOPK), is highly homologous to the MAP kinase kinases, particularly MKK3. Physiologically, PBK/TOPK plays a positive regulatory role in proper chromosomal separation and cytokinesis through phosphorylation of various targets. It was found that PBK could be a prognostic biomarker for prostate cancer that would discriminate aggressive prostate cancer from indolent disease, and was a potential target for the therapeutic intervention of aggressive prostate cancer in men [6]. CDCA2, also named as Repo-Man (Recruits PP1 onto Mitotic chromatinat anaphase), is a nuclear protein that binds to protein phosphatase 1 γ (PP1γ). It suggested that CDCA2 was responsible for the targeting of PP1 to chromatin during anaphase, leading to the dephosphorylation of H3 and controlled cell proliferation in vitro [7].

Polo-like kinase 1 (PLK1), a serine/threonine kinase that is essential for cell cycle progression, is a novel and major regulator of FOXO1 in the late phases of the cell cycle. Studies [8] have shown that PLK1 dependent phosphorylation of FOXO1 induced its nuclear exclusion and negatively regulated FOXO1′s transcriptional activity in prostate cancer. Blocking the PLK1-dependant

phosphorylation of FOXO1 restored the pro-apoptotic function of FOXO1 in PCa. In addition, some studies [9] believed that cotargeting PLK1 and androgen receptor (AR) would be effective in the treatment of paclitaxel-resistant prostate cancer. Yang et al. [10] used weighted gene co-expression network analysis (WGCNA) to analyze hub genes associated with high-stage tumor stage and gleason scores in prostate cancer. They studied CCNA2 and found that it has a significant correlation with biochemical recurrence rate and survival rate of prostate cancer, and verified the role of CCNA2 in prostate cancer cell lines and found that CCNA2 was associated with tumor cell proliferation, invasion, metastasis, and cell cycle. Cell division cycle 20 (CDC20), is a WD40 repeat domain containing E3 ligase that associates with and activates the anaphase-promoting complex (APC). In prostate cancer, Mao et al. [11] demonstrated high expression level of CDC20 associated with higher gleason scores and predicted biochemical recurrence. Li et al. [12] presented the preliminary evidence that CDC20 was significantly involved in mediating chemoresistance to docetaxel partly through inhibiting Wnt/β- catenin signaling.

Metastatic PCa accounts for the majority of PCa specific mortality. However, the ability to distinguish primary PCa with metastatic potential has not been achieved. Studies have shown that primary human PCa with concurrent increased expression of TOP2A and EZH2 have a faster time to biochemical recurrence, were more susceptible to progress to a metastatic disease and thereby resulting in increased PCa specific death [13]. Jin et al. [14] suggested that TROAP and E2F1 were co-regulated by EZH2, TROAP was overexpressed in PCa when compared with the expression in normal tissues, and indicated that TROAP could promote PCa development and progression, at least partially, via a TWIST/c-Myc pathway. Cyclin B2 (CCNB2), a member of the cyclin protein family, has been found to be up-regulated in human cancers. Studies have suggested that detection of serum circulating CCNB2 mRNA may have potential clinical applications in screening and monitoring of metastasis and therapeutic treatments [15]. Other studies suggest that knockdown of circ_CCNB2 was shown to promote PCa radiosensitivity through autophagy repression by miR-30b-5p/KIF18A axis, developing a molecular resistance mechanism of PCa radiotherapy and a feasible strategy to increase radiosensitivity [16]. But the role of CCNB2 in prostate cancer tumorigenesis and metastasis remains unclear. Budding uninhibited by benzimidazoles 1 (BUB1) is a mitotic checkpoint serine/threonine kinase that serves a central role in aligning chromosomes and establishing the mitotic spindle checkpoint [17]. Researches have suggested that the overexpression of BUB1B in PCa cells promoted cell proliferation and migration. Furthermore, high expression levels of BUB1B in PCa were associated with poor clinicopathological features of PCa and predicted poor outcomes of patients with advanced PCa [18]. Cyclin dependent kinase 1 (CDK1, also called CDC2) as an androgen receptor Ser-81 kinase. Researches have indicated that CDK1 can stabilize ARs and that increased CDK1 activity may enhance AR responses to low levels of androgen in androgen-independent PCa [19]. In prostate cancer patients with PSA at diagnosis of $<$ 20 ng /ml , phosphorylation of AR at serine 515 by CDK1 may be an independent prognostic marker [20]. But whether CDK1 plays a role in prostate cancer metastasis is unclear.

Our methods have several limitations that could be improved in future studies. For instance, Zhang et al. [21] developed a time-course RNA-seq data-driven computational method to study the network mechanisms underlying cancer drug resistance, we also can use the method to study whether the identified hub genes are associated with the targeted therapy response and prognosis of PCa patients. Zhang et al. [22] presented a promising single-cell RNA-seq transcriptome-based multilayer network approach to elucidate the interactions between tumor cell and tumor-associated microenvironment and to identify prognostic and predictive signatures of cancer patients, we can

combine single-cell RNA-seq data with clinical gene expression data to develop a computational pipeline for identifying the prognostic and predictive signature that connects cancer cells and microenvironmental cells. Then the single-cell RNA-seq transcriptome-based multilayer network biomarker may predict survival outcome and therapeutic response of PCa patients.

5. Conclusion

The present study analyzed the gene expression profiles of localized and metastatic prostate cancer tissue using bioinformatics analysis. It was identified that DEGs, may serve roles in development and metastasis of prostate cancer. These genes may become the novel potential biomarkers and therapeutic targets for patients with prostate cancer, and distinguish metastatic potential patients with indolent prostate cancer. However, further research is required to confirm whether these genes play a developmental or metastatic role in prostate cancer, or both.

Acknowledgements

This work is supported by the Science and Technology Development Guidance Plan (Medical and Health) Project of Wuxi (CZ2020003).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, Cancer Statistics, 2021, *CA. Cancer J. Clin.*, **71** (2021), 7–33.
2. J. A. Hempelmann, C. M. Lockwood, E. Q. Konnick, M. T. Schweizer, E. S. Antonarakis, T. L. Lotan, et al., Microsatellite instability in prostate cancer by PCR or next-generation sequencing, *J. Immunother. Cancer*, **6** (2018), 29.
3. L. Foj, X. Filella, Identification of Potential miRNAs Biomarkers for High-Grade Prostate Cancer by Integrated Bioinformatics Analysis, *Pathol. Oncol. Res.*, **25** (2019), 1445–1456.
4. Z. Zhang, C. Chai, T. Shen, X. Li, J. Ji, C. Li, et al., Aberrant KIF20A expression is associated with adverse clinical outcome and promotes tumor progression in Prostate Cancer, *Dis. Markers*, (2019), 4782730.
5. C. Yu, H. Cao, X. He, P. Sun, Y. Feng, L. Chen, et al., Cyclin-dependent kinase inhibitor 3 (CDKN3) plays a critical role in prostate cancer via regulating cell cycle and DNA replication signaling, *Biomed. Pharmacother.*, **96** (2017), 1109–1118.
6. J. D. Brown-Clay, D. N. Shenoy, O. Timofeeva, B. V. Kallakury, A. K. Nandi, P. P. Banerjee, PBK/TOPK enhances aggressive phenotype in prostate cancer via beta-catenin-TCF/LEF-mediated matrix metalloproteinases production and invasion, *Oncotarget*, **6** (2015), 15594–15609.
7. Y. Zhang, Y. Cheng, Z. Zhang, Z. Bai, H. Jin, X. Guo, et al., CDCA2 inhibits apoptosis and promotes cell proliferation in Prostate Cancer and is directly regulated by HIF-1alpha pathway, *Front. Oncol.*, **10** (2020), 725.

8. L. Gheghiani, S. Shang, Z. Fu, Targeting the PLK1-FOXO1 pathway as a novel therapeutic approach for treating advanced prostate cancer, *Sci. Rep.*, **10** (2020), 12327.
9. S. B. Shin, S. U. Woo, H. Yim, Cotargeting Plk1 and androgen receptor enhances the therapeutic sensitivity of paclitaxel-resistant prostate cancer, *Ther. Adv. Med. Oncol.*, **11** (2019), 1758835919846375.
10. R. Yang, Y. Du, L. Wang, Z. Chen, X. Liu, Weighted gene co-expression network analysis identifies CCNA2 as a treatment target of prostate cancer through inhibiting cell cycle, *J. Cancer*, **11** (2020), 1203–1211.
11. Y. Mao, K. Li, L. Lu, J. Si-Tu, M. Lu, X. Gao, Overexpression of Cdc20 in clinically localized prostate cancer: Relation to high Gleason score and biochemical recurrence after laparoscopic radical prostatectomy, *Cancer Biomark.*, **16** (2016), 351–358.
12. K. Li, Y. Mao, L. Lu, C. Hu, D. Wang, J. Si-Tu, et al., Silencing of CDC20 suppresses metastatic castration-resistant prostate cancer growth and enhances chemosensitivity to docetaxel, *Int. J. Oncol.*, **49** (2016), 1679–1685.
13. D. P. Labbe, C. J. Sweeney, M. Brown, P. Galbo, S. Rosario, K. M. Wadosky, et al., TOP2A and EZH2 provide early detection of an aggressive Prostate Cancer subgroup, *Clin. Cancer Res.*, **23** (2017), 7072–7083.
14. L. Jin, Y. Zhou, G. Chen, G. Dai, K. Fu, D. Yang, et al., EZH2-TROAP pathway promotes Prostate Cancer progression via TWIST signals, *Front. Oncol.*, **10** (2021), 592239.
15. M. L. Mo, Z. Chen, J. Li, H. L. Li, Q. Sheng, H. Y. Ma, et al., Use of serum circulating CCNB2 in cancer surveillance, *Int. J. Biol. Markers*, **25** (2010), 30B1A552-6F86-4064-888D-90FEC7A5BB29.
16. F. Cai, J. Li, J. Zhang, S. Huang, Knockdown of Circ_*CCNB2* sensitizes Prostate Cancer to radiation through repressing autophagy by the miR-30b-5p/KIF18A axis, *Cancer Biother. Radiopharm.*, (2020), Online ahead of print.
17. L. J. Zhu, Y. Pan, X. Y. Chen, P. F. Hou, BUB1 promotes proliferation of liver cancer cells by activating SMAD2 phosphorylation, *Oncol. Lett.*, **19** (2020), 3506–3512.
18. X. Fu, G. Chen, Z. D. Cai, C. Wang, Z. Z. Liu, Z. Y. Lin, et al., Overexpression of *BUB1B* contributes to progression of prostate cancer and predicts poor outcome in patients with prostate cancer, *Onco. Targets Ther.*, **9** (2016), 2211–2220.
19. S. Chen, Y. Xu, X. Yuan, G. J. Bubley, S. P. Balk, Androgen receptor phosphorylation and stabilization in prostate cancer by cyclin-dependent kinase 1, *Proc. Natl. Acad. Sci. U.S.A.*, **103** (2006), 15969–15974.
20. J. M. Willder, S. J. Heng, P. McCall, C. E. Adams, C. Tannahill, G. Fyffe, et al., Androgen receptor phosphorylation at serine 515 by Cdk1 predicts biochemical relapse in prostate cancer patients, *Br. J. Cancer*, **108** (2013), 139–148.
21. J. Zhang, W. Zhu, Q. Wang, J. Gu, L. F. Huang, X. Sun, Differential regulatory network-based quantification and prioritization of key genes underlying cancer drug resistance based on time-course RNA-seq data, *PLoS. Comput. Biol.*, **15** (2019), e1007435.
22. J. Zhang, M. Guan, Q. Wang, J. Zhang, T. Zhou, X. Sun, Single-cell transcriptome-based multilayer network biomarker for predicting prognosis and therapeutic response of gliomas, *Brief. Bioinform.*, **21** (2020), 1080–1097.