

# Emerging biomarker associated with prognosis of prostate cancer identified via co-expression analysis

**fangdie ye**

Huashan Hospital Fudan University <https://orcid.org/0000-0001-8572-0493>

**Yingchun Liang**

Huashan Hospital Fudan University

**Zhang Chen**

Huashan Hospital Fudan University

**Yufei Liu**

Huashan Hospital Fudan University

**Haowen Jiang** (✉ [haowj\\_sh@fudan.edu.cn](mailto:haowj_sh@fudan.edu.cn))

Huashan Hospital, Fudan University <https://orcid.org/0000-0003-0995-8998>

---

## Primary research

**Keywords:** Foxf1, prostate cancer, weight gene co-expression network analysis, biomarker, differential gene expression analysis

**Posted Date:** March 12th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-281491/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Forkhead box protein f1 (Foxf1) is associated with oncogenesis, and maybe play a key role in prostate cancer. However, the effect of Foxf1 on the development of prostate cancer and its clinical implications were still unknown.

**Method:** The transcriptome data and clinical information were downloaded from The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO). The differential co-expression genes which were the most closely correlated with prostate cancer were identified by weighted gene co-expression network analysis (WGCNA) and differential expression analysis method. The univariate cox regressions were performed to identify the prognostic genes. Finally, Foxf1 were verified in clinical tumor samples and matched normal tissues.

**Results:** A total of 108 differential co-expression genes were identified. Foxf1 was identified that significantly associated with overall survival rate and biochemical recurrence of prostate. Then, protein-protein interaction (PPI) networks were constructed, including 21 nodes and 26 edges. The data revealed GAPVD1, MIR522, MIR937, CPS1, HOXB9, FUBP3, MIR936, SDF4, SATB2, HOXB13, MIR375, PUSL1, SATB1, MIR382, MIR429, HOXD13, PRSS50, FLNC, STRBP, KHSRP were significant associated with the regulation and function of differentially expressed Foxf1 in prostate. The experimental results confirmed the Foxf1 were downregulated in the prostate cancer tissues compared with the matched normal tissues.

**Conclusion:** we obtained a gene Foxf1 which is significantly related with the prognosis of prostate cancer, Foxf1 have significant values in predicting the patients' Overall survival and may serve as a potential prognosis biomarker and a new therapeutic target of prostate cancer.

## Introduction

Prostate Cancer is the most common malignant tumor in the male genitourinary system[1]. In the United States, the incidence rate is highest, while the disease mortality rate accounts for 9% of all malignant tumor deaths in men, second only to lung cancer[2]. According to the latest epidemiological statistics, in the United States, nearly 164,690 new cases of prostate cancer were diagnosed in 2018, and nearly 29,430 people were died because of prostate[3]. Prostate cancer lack of specific clinical symptom as the tumor grows quietly, and the failure to detect tumor early cause the patients lost the opportunity for curative surgery[4]. Today, prostate cancer was diagnosed by detecting the prostate specific antigen (PSA) and digital rectal examination and eventually confirmed by needle biopsies[5]. However, the PSA screening has low specificity although it has high sensitivity. This results in a negative needle biopsies and cause overtreatment and unnecessary injury[6, 7]. Thus, there is an urgent need for the development of efficient biomarkers.

With the development of transcriptome, bioinformatics has become more popular for identify molecular mechanisms of diseases and discover potential biomarkers[8, 9]. Weighted gene co-expression network analysis (WGCNA) can identify gene potential function and extract gene co-expression modules

according to the clinical profile[10]. This algorithm calculates the correlation coefficient value of expression level and then takes an exponentiation, which makes the genes network to meet the scale-free distribution[11]. Briefly, we could obtain interested gene modules which have high correlation with clinical profile. Furthermore, differential gene expression analysis provides methods for discovering significant differentiation in gene expression level and understanding the molecular mechanisms underlying genome regulation between cancer tissues and adjacent normal tissues[12]. Thus, the highly related genes that are found by using WGCNA and differential gene expression analysis are useful to serve as efficient biomarkers.

In this study, the gene expression data profiles of TCGA-PRAD and GSE70768 were analyzed through WGCNA and differential gene expression analysis to get hub genes. Then, we performed functional enrichment analysis and protein-protein interaction (PPI) analysis to explore the underlying mechanism.

## Method

The workflow of data preparation, analysis, and validation is shown in figure 1.

### Data preparation

The level 3 RNA sequencing data and clinical information of prostate cancer were downloaded from the TCGA(<https://portal.gdc.cancer.gov/>) up to November 9, 2020. There were 552 samples, which including 500 prostate cancers and 52 normal tissues. As suggested by the “limma” R package[13], the gene expression profiles of Pca were normalized. RNA sequence data of another 199 samples were obtained from GEO(<https://www.ncbi.nlm.nih.gov/geo/>) using R package GEOquery[14], which including 126 prostate cancers and 73 corresponding normal tissues.

Weighted correlation network analysis (WGCNA) was utilized to select candidate biomarkers and therapeutic targets[15, 16]. The gene expression data profiles of TCGA-PRAD and GSE70768 were formed which each row displays a different gene and each column displays a sample. The clinical feature table contained tumor and normal sample.

First, outlier samples by hierarchical clustering were delete, then build a scale-free network, the best soft threshold was obtained according to the value of mean connectivity and scale-independent. The adjacency matrix was created by calculating the correlation value of every two genes. To further select functional modules in the co-expression network, the association value between gene modules and the clinical modules was calculated. The gene module with high correlation with clinical profiles was identified.

### Differential Expression Analysis and Interaction with the Modules of Interest

“limma” R package which provides a method for differential expression analyses on RNA-sequence data, was used to find the differentially expressed genes (DEGs) between PRAD and corresponding normal tissues in TCGA and GSE70768. The DEGs of the TCGA-PRAD and GSE70768 were displayed as a

volcano plot by utilize the “ggplot2” R package[17]. Then, the overlapping genes of DEGs and interesting modules were identified as potential prognostic genes, which visualized as a venn diagram using the online software(<http://sangerbox.com/Tool>).

### **Functional Annotation for Genes of Interest**

To conduct Gene Ontology(GO) and Kyoto Encyclopedia of Genes and Genomes(KEGG) of selected genes, the “clusterProfiler” R package was utilized to explore the functions among the overlapping genes, with p value  $\leq 0.05$ [18]. GO annotation contains the three sub-ontologies—biological process (BP), cellular component (CC), and molecular function (MF), which can identify the biological properties of interested genes for all organisms[19].

### **Screening of Hub Gene and Construction of PPI**

In our study, the “survival” R package was utilized to determine survival curves, the clinical data of these Pca patients were came from TCGA-PRAD. Kaplan-Meier curve was used to evaluate univariate survival analysis, with P-value $\leq 0.05$ . The genes that have a significant correlation with prognosis were considered as hub genes. Then, the cytoscape (v3.8.1) was utilized to construct the PPI network of overlapping genes[20]. In a co-expression network, the relationship’s degrees were calculated by the algorithm of CytoHubba, which is a plugin in Cytoscape[21].

### **Verification of the Expression Patterns and the Prognostic Values of Hub Genes**

To affirm the reliability of the hub gene in our study, we verified the expression level of the target gene in prostate cancer and adjacent tissues, also confirm the gene expression in different nodal metastasis status by using the online tool Ualcan (<http://ualcan.path.uab.edu/cgi-bin/ualcan-res.pl>)[22] and the Oncomine (<https://www.oncomine.org/resource/login.html>) . based on the data from the TCGA and CamcAPP(<https://bioinformatics.cruk.cam.ac.uk/apps/camcAPP/>)[23], the relationship between overall survival (OS) and the hub gene were confirmed by using the survival package in R tool.

### **Immunohistochemistry (IHC) staining**

Briefly, prostate cancer tissues and adjacent normal tissues were stained via the following steps: the tissues were fixed by formalin, dehydrated and paraffin embedded. Then immunostaining of Foxf1 was carried out with rabbit polyclonal anti-Foxf1 antibody (1:100, SAB, #36858). Then two experienced pathologists assessed the staining independently.

## **Results**

### **Construction of weighted Gene Co-expression modules**

In order to identify the function modules in PRAD patients. Weighted gene co-expression analysis (WGCNA) was constructed from the TCGA-PRAD and GSE70768 with the WGCNA R package. The 11

modules were generated in TCGA-PRAD and 11 modules were generated in GSE70768 where the grey module was not assigned into any cluster. Then heatmap was plotted to show the relationship between different gene modules and the two clinical profiles. The results of the module-clinical profile relationship were displayed in figure 2 and figure 3 which can be clearly revealed that the cyan gene module had closeness with clinical profile in TCGA-PRAD, and the thistle2 gene module had the strongest relationship with clinical profile in GSE70768. The correlation coefficients between the cyan gene module and clinical profile were 0.48, and the correlation coefficients between thistle2 gene module and clinical profile was 0.64. then, the results of the Pearson correlation coefficients between cyan/thistle2 gene module and clinical profile were 0.57/0.8( $p \leq 0.01$ ). It is clearly certified that the cyan/thistle2 gene module had strongly relationship with the clinical profile.

### **Identification of hub gene**

A total of 6088 genes were differentially expressed between tumor tissues and corresponding normal tissues in the TCGA database (Figure 4A), and 563 genes were differentially expressed in GSE70768 database (Figure 4B). 1230 and 2513 co-expression genes were identified in the cyan and thistle2 modules respectively. The interactive network among this gene including 108 genes, visualized by venn diagram (figure 4C), which indicated that these genes were the hub genes. Then we further explored the relationship between this genes and prostate cancer overall survival (OS). The result revealed that Foxf1 was the target gene in prostate cancer. subsequently, we utilized the CamcAPP to verify the relationship between the expression of Foxf1 and the biochemical recurrence of prostate cancer, which had a meaningful relationship.

### **Low expression and hypermethylation of Foxf1 in prostate cancer**

We explore the expression of Foxf1 in prostate cancer and adjacent tissues by utilizing Ualcan, visualized as a box plot graph (Fig 6D). we further explored the expression level of Foxf1 in prostate cancer based on nodal metastasis status. Foxf1 was lower in cancer tissues than adjacent tissues, and lower in nodal metastasis tissues (Fig 6E). In addition, the analysis from Ualcan demonstrated that Foxf1 were significantly higher methylation in cancer tissues compared with normal tissues ( $P < 0.05$ , Fig 6F). Furthermore, the protein expressional levels of the Foxf1 gene were significantly lower in prostate cancer tissues compared with the corresponding normal tissues based on the immunohistochemistry staining (Fig 5).

### **GO and KEGG pathway analysis of the 108 Genes**

GO and KEGG pathway analysis were performed to obtain further insight into the potential function of the identified 108 genes. The analysis results were showed in figure 7. Among the most enriched GO terms were cell-substrate adhesion and focal adhesion. KEGG pathway analysis shows that hub genes are highly correlate with proteoglycans in cancer.

### **PPI network construction**

A protein-protein network among the overlapped genes was constructed by using cytoscape(v3.8.1), with 2672 nodes and 81425 edges (Fig 8A). Then we selected the particular nodes we were interested to construct the ultimate PPI by using the algorithm of Cytohubba plugin, with 21 nodes and 26 edges (Fig 8B). The data suggested GAPVD1, MIR522, MIR937, CPS1, HOXB9, FUBP3, MIR936, SDF4, SATB2, HOXB13, MIR375, PUSL1, SATB1, MIR382, MIR429, HOXD13, PRSS50, FLNC, STRBP, KHSRP were significant associated with the regulation and function of differentially expressed Foxf1 in prostate.

## Discussion

Prostate cancer is the most common malignant disease in men and it was diagnosed in 12% and accounted for 9% of male cancer deaths[24]. Nowadays, many studies endeavored to identify the hub genes which play an important role on the development of prostate cancer from the transcriptome level[25]. In this study, a total of 108 significant genes were elected by WGCNA method, which was utilized to analysis the relationship between prostate cancer transcriptome and clinical profile [26]. As displayed in functional annotation analysis by the R package “clusterProfiler”, these genes were mainly in cell-substrate adhesion and focal adhesion, which may cause local metastasis of the tumor. Then, according to the univariable cox analysis, the Foxf1 gene was carried out. Furthermore, the protein-protein interactions (PPIs) network was screened out according to MCC scores from the cytoHubba plugin in Cytoscape.

Foxf1 is a member of the forkhead box family of transcription factor and has been previous reported to be critical for multiple human cancer, such as colorectal cancer[27], gastrointestinal stromal tumor[28] breast cancer[29, 30] and non-small cell lung cancer[31]. we found that the downregulation of Foxf1 was significantly correlated with a poorer prognosis for patients with prostate cancer. Then we further evaluated the correlation between the Foxf1 expression and the pathological characteristics, we found that Foxf1 downregulation was correlated with positive nodal invasion. But the mechanism of Foxf1 dysregulation is still unclear. Previous study reported that the expression level of Foxf1 was adjusted by multiple genomic and epigenetic mechanisms. In this study, we found Foxf1 were significantly higher methylation in prostate cancer tissues compared with corresponding normal tissues, indicating that hypermethylation of Foxf1 may causes the Foxf1 gene to be lowly expressed in prostate cancer. These finding indicated that Foxf1 might be a potential prognostic biomarker in prostate cancer.

## Conclusions

In conclusion, we obtained a gene Foxf1 which is significantly related with the prognosis of prostate cancer, Foxf1 have significant values in predicting the patients' Overall survival and may serve as a potential prognosis biomarker and a new therapeutic target of prostate cancer. Further studies are necessary to confirm these results in our study.

## Abbreviations

**TCGA:** The cancer genome atlas

**GEO:** Gene Expression Omnibus

**Pca:** Prostate cancer

**WGCNA:** Weighted gene co-expression network analysis

**Foxf1:** Forkhead box protein f1

**PSA:** Prostate specific antigen

**PRAD:** Prostate adenocarcinoma

**PPI:** protein-protein interaction

**GO:** Gene Ontology

**KEGG:** Kyoto Encyclopedia of Genes and Genomes

## **Declarations**

### **Data availability statement**

All transcriptome and clinical data were available in TCGA database. All the experimental data analyzed and displayed in the present manuscript are generated for this study are included in the article material.

### **Ethics statement**

This article had been reviewed and approved by the Ethics Committee of Huashan Hospital of Fudan university.

### **Author contribution**

FY, YL, ZC contributed equally to this work. All authors contributed towards experiments and data analysis, and agreed to take responsibility for all aspects of the job.

### **Conflict of interest**

The authors declare no conflicts of interest.

### **Funding**

This study was supported by the National Natural Science Foundation of China (Grant Numbers: 81872102).

### **Acknowledgements**

We sincerely appreciate all lab members.

## Consent for publication

Not applicable.

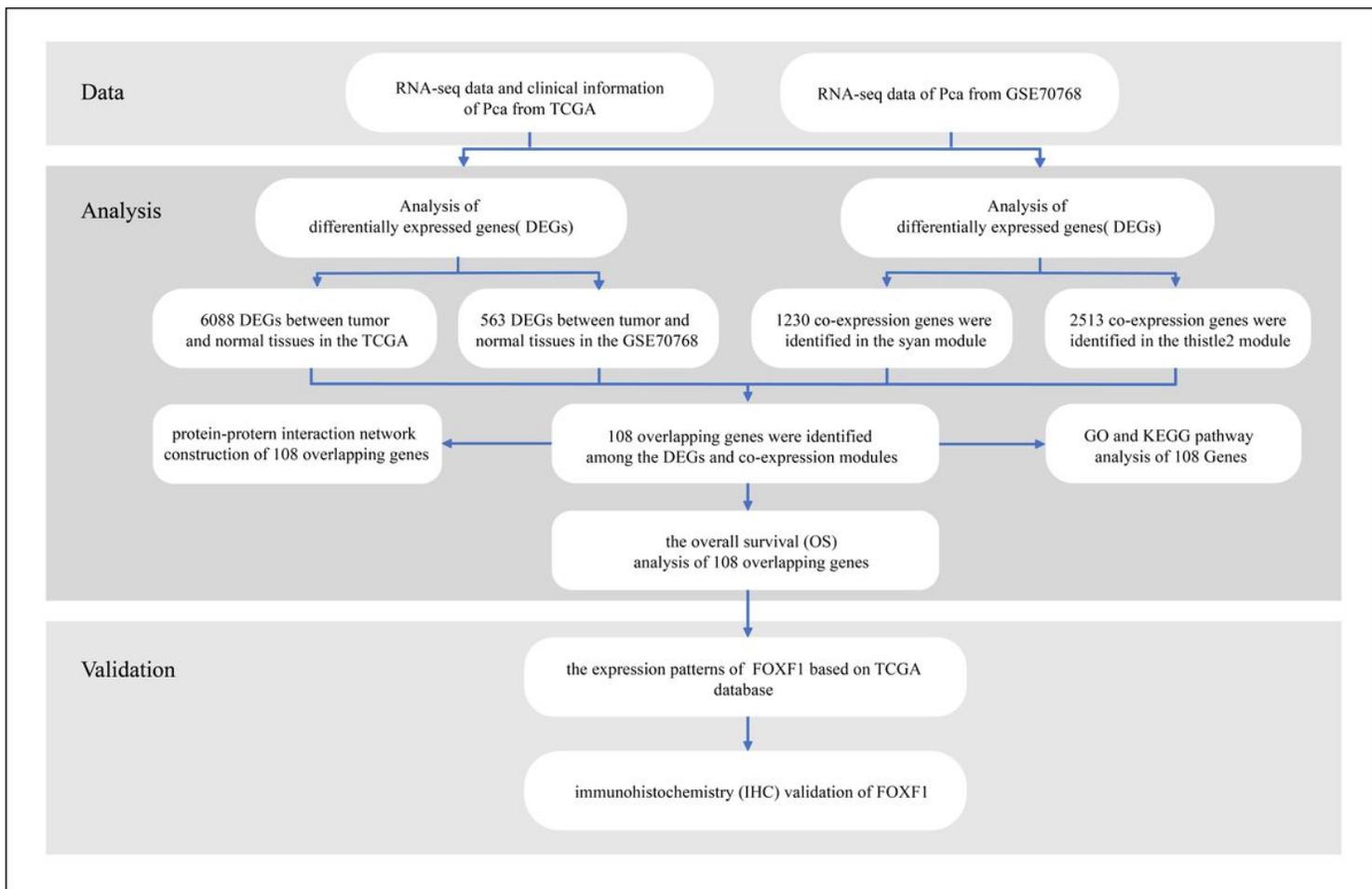
## References

1. Litwin MS, Tan HJ: **The Diagnosis and Treatment of Prostate Cancer: A Review.** *JAMA* 2017, **317**(24):2532-2542.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.** *CA Cancer J Clin* 2018, **68**(6):394-424.
3. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2019.** *CA Cancer J Clin* 2019, **69**(1):7-34.
4. Zadra G, Priolo C, Patnaik A, Loda M: **New strategies in prostate cancer: targeting lipogenic pathways and the energy sensor AMPK.** *Clin Cancer Res* 2010, **16**(13):3322-3328.
5. Nogueira L, Corradi R, Eastham JA: **Prostatic specific antigen for prostate cancer detection.** *International braz j urol : official journal of the Brazilian Society of Urology* 2009, **35**(5):521-529; discussion 530-522.
6. Day JR, Jost M, Reynolds MA, Groskopf J, Rittenhouse H: **PCA3: from basic molecular science to the clinical lab.** *Cancer Lett* 2011, **301**(1):1-6.
7. Pepe P, Fraggetta F, Galia A, Skonieczny G, Aragona F: **PCA3 score and prostate cancer diagnosis at repeated saturation biopsy. Which cut-off: 20 or 35?** *International braz j urol : official journal of the Brazilian Society of Urology* 2012, **38**(4):489-495.
8. Akalin PK: **Introduction to bioinformatics.** *Molecular nutrition & food research* 2006, **50**(7):610-619.
9. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B *et al.*: **Integrative genomic profiling of human prostate cancer.** *Cancer Cell* 2010, **18**(1):11-22.
10. Zhang B, Horvath S: **A general framework for weighted gene co-expression network analysis.** *Stat Appl Genet Mol Biol* 2005, **4**:Article17.
11. Horvath S, Dong J: **Geometric interpretation of gene coexpression network analysis.** *PLoS Comput Biol* 2008, **4**(8):e1000117.
12. Bacher U, Kohlmann A, Haferlach T: **Current status of gene expression profiling in the diagnosis and management of acute leukaemia.** *Br J Haematol* 2009, **145**(5):555-568.
13. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Res* 2015, **43**(7):e47.
14. Davis S, Meltzer PS: **GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor.** *Bioinformatics* 2007, **23**(14):1846-1847.

15. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis**. *BMC Bioinformatics* 2008, **9**:559.
16. Langfelder P, Horvath S: **Fast R Functions for Robust Correlations and Hierarchical Clustering**. *Journal of statistical software* 2012, **46**(11).
17. Gómez-Rubio V: **ggplot2 - Elegant Graphics for Data Analysis (2nd Edition)**. *Journal of statistical software* 2017, **77**(Book Review 2).
18. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological themes among gene clusters**. *OMICS* 2012, **16**(5):284-287.
19. Gene Ontology C: **The Gene Ontology (GO) project in 2006**. *Nucleic Acids Res* 2006, **34**(Database issue):D322-326.
20. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. *Genome Res* 2003, **13**(11):2498-2504.
21. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY: **cytoHubba: identifying hub objects and sub-networks from complex interactome**. *BMC Syst Biol* 2014, **8 Suppl 4**:S11.
22. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, Varambally S: **UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses**. *Neoplasia* 2017, **19**(8):649-658.
23. Dunning MJ, Vowler SL, Lalonde E, Ross-Adams H, Boutros P, Mills IG, Lynch AG, Lamb AD: **Mining Human Prostate Cancer Datasets: The "camcAPP" Shiny App**. *EBioMedicine* 2017, **17**:5-6.
24. Wedge DC, Gundem G, Mitchell T, Woodcock DJ, Martincorena I, Ghori M, Zamora J, Butler A, Whitaker H, Kote-Jarai Z *et al*: **Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets**. *Nat Genet* 2018, **50**(5):682-692.
25. Gerhauser C, Favero F, Risch T, Simon R, Feuerbach L, Assenov Y, Heckmann D, Sidiropoulos N, Waszak SM, Hubschmann D *et al*: **Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories**. *Cancer Cell* 2018, **34**(6):996-1011 e1018.
26. Chou WC, Cheng AL, Brotto M, Chuang CY: **Visual gene-network analysis reveals the cancer gene co-expression in human endometrial cancer**. *BMC genomics* 2014, **15**:300.
27. Wang S, Xiao Z, Hong Z, Jiao H, Zhu S, Zhao Y, Bi J, Qiu J, Zhang D, Yan J *et al*: **FOXF1 promotes angiogenesis and accelerates bevacizumab resistance in colorectal cancer by transcriptionally activating VEGFA**. *Cancer Lett* 2018, **439**:78-90.
28. Ran L, Chen Y, Sher J, Wong EWP, Murphy D, Zhang JQ, Li D, Deniz K, Sirota I, Cao Z *et al*: **FOXF1 Defines the Core-Regulatory Circuitry in Gastrointestinal Stromal Tumor**. *Cancer Discov* 2018, **8**(2):234-251.
29. Katoh M, Igarashi M, Fukuda H, Nakagama H, Katoh M: **Cancer genetics and genomics of human FOX family genes**. *Cancer Lett* 2013, **328**(2):198-206.

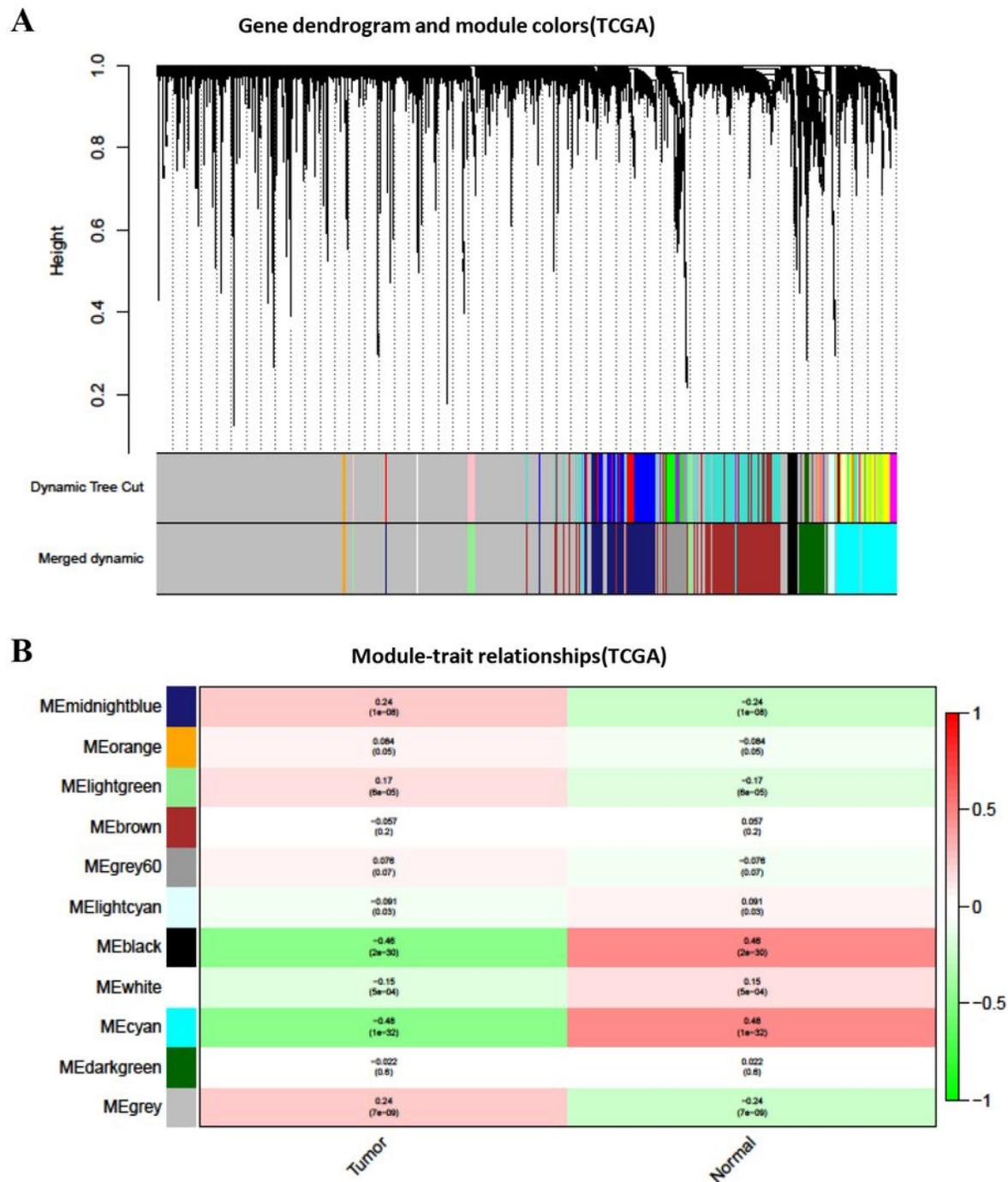
30. Lo PK, Lee JS, Liang X, Han L, Mori T, Fackler MJ, Sadik H, Argani P, Pandita TK, Sukumar S: **Epigenetic inactivation of the potential tumor suppressor gene FOXF1 in breast cancer.** *Cancer Res* 2010, **70**(14):6047-6058.
31. Zhao J, Xue X, Fu W, Dai L, Jiang Z, Zhong S, Deng B, Yin J: **Epigenetic activation of FOXF1 confers cancer stem cell properties to cisplatinresistant nonsmall cell lung cancer.** *Int J Oncol* 2020, **56**(5):1083-1092.

## Figures



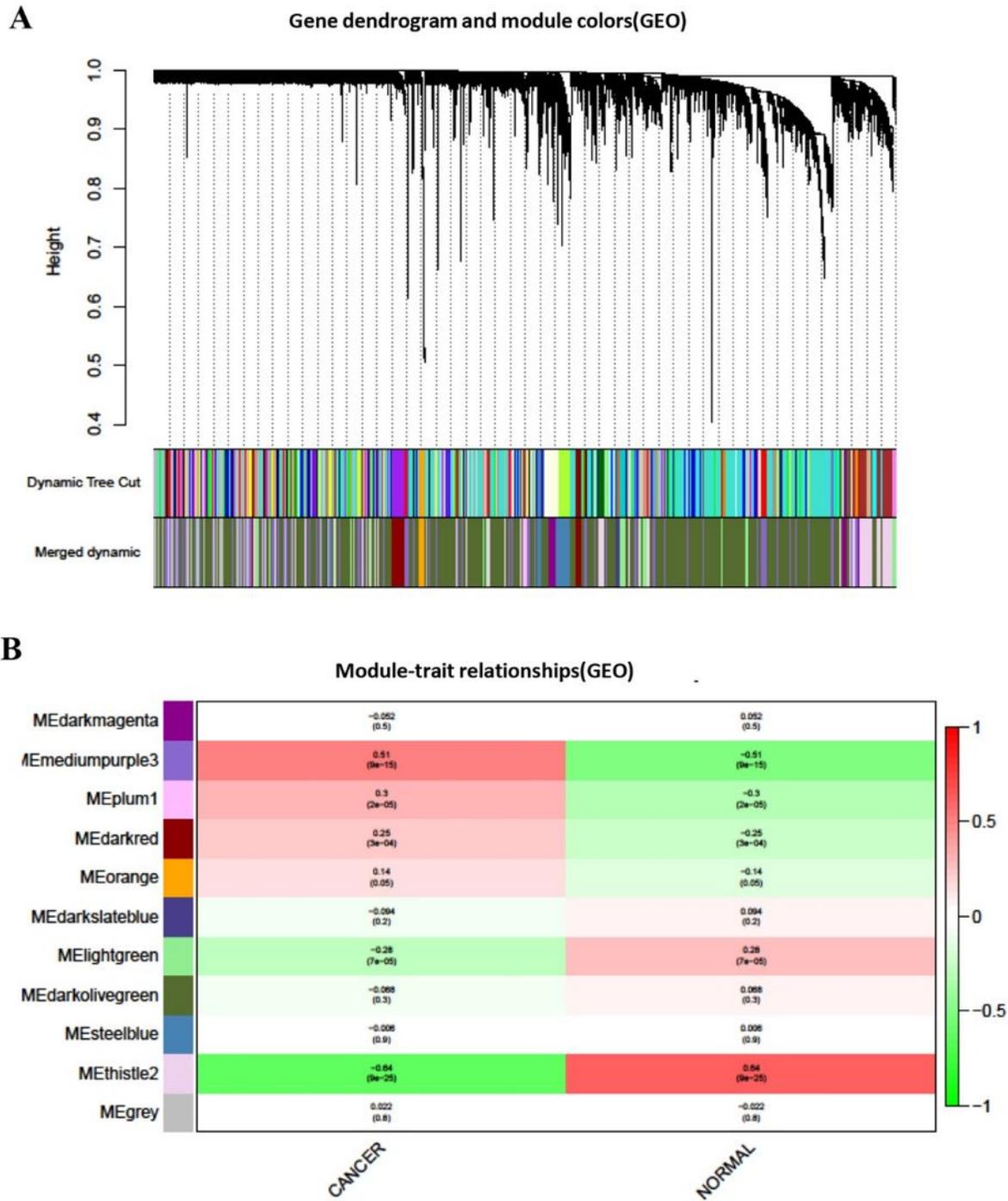
**Figure 1**

Workflow of this study



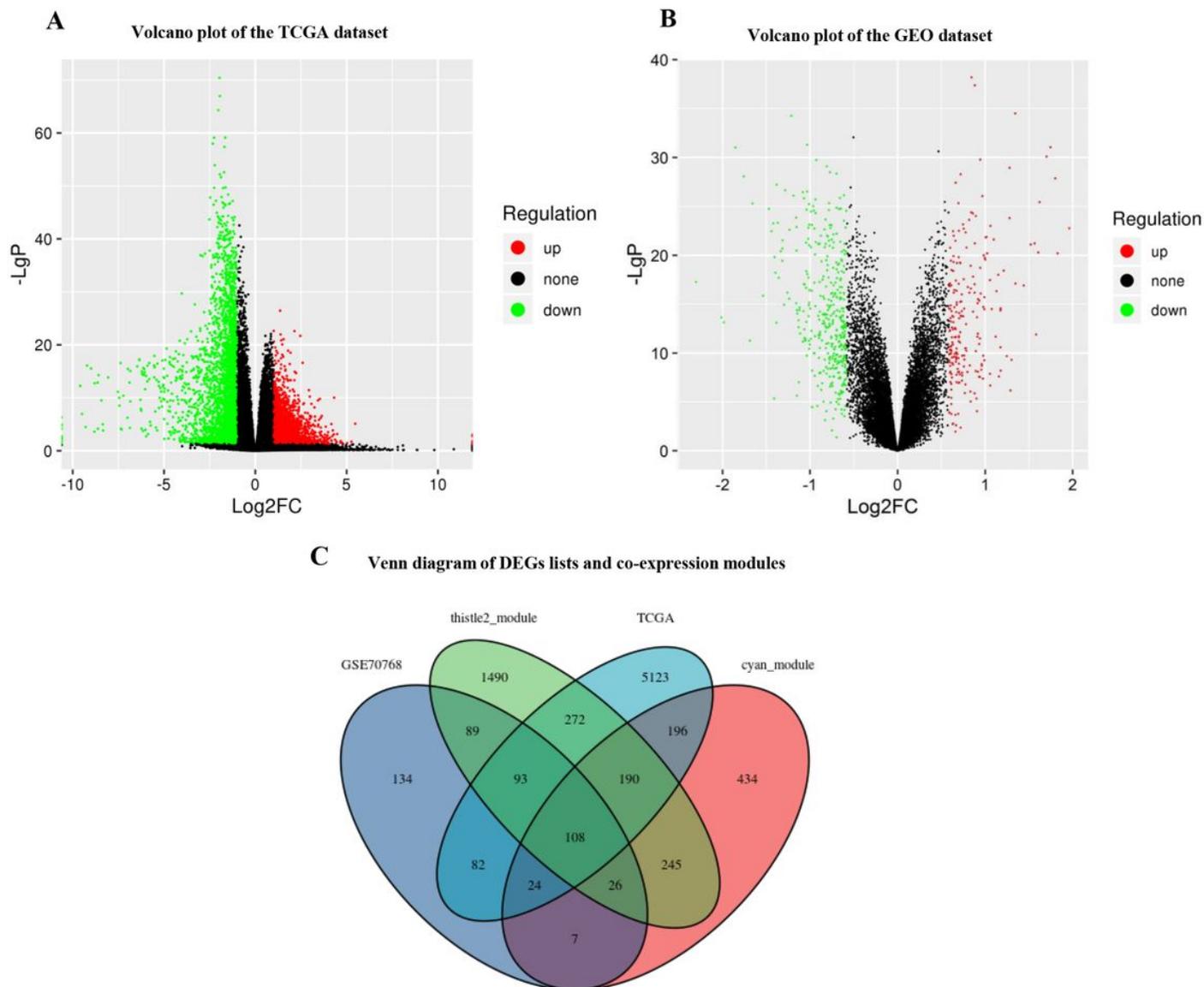
**Figure 2**

Identification of modules associated with the clinical profiles in the TCGA-PRAD dataset. (A) The Cluster dendrogram of co-expression network modules were created by a hierarchical clustering of genes based on the 1-TOM matrix. Each module was assigned different colors. (B) Module-trait relationships. Each row corresponds to a color module and column corresponds to a clinical profile (cancer and normal). Each cell contains the corresponding correlation and P-value.



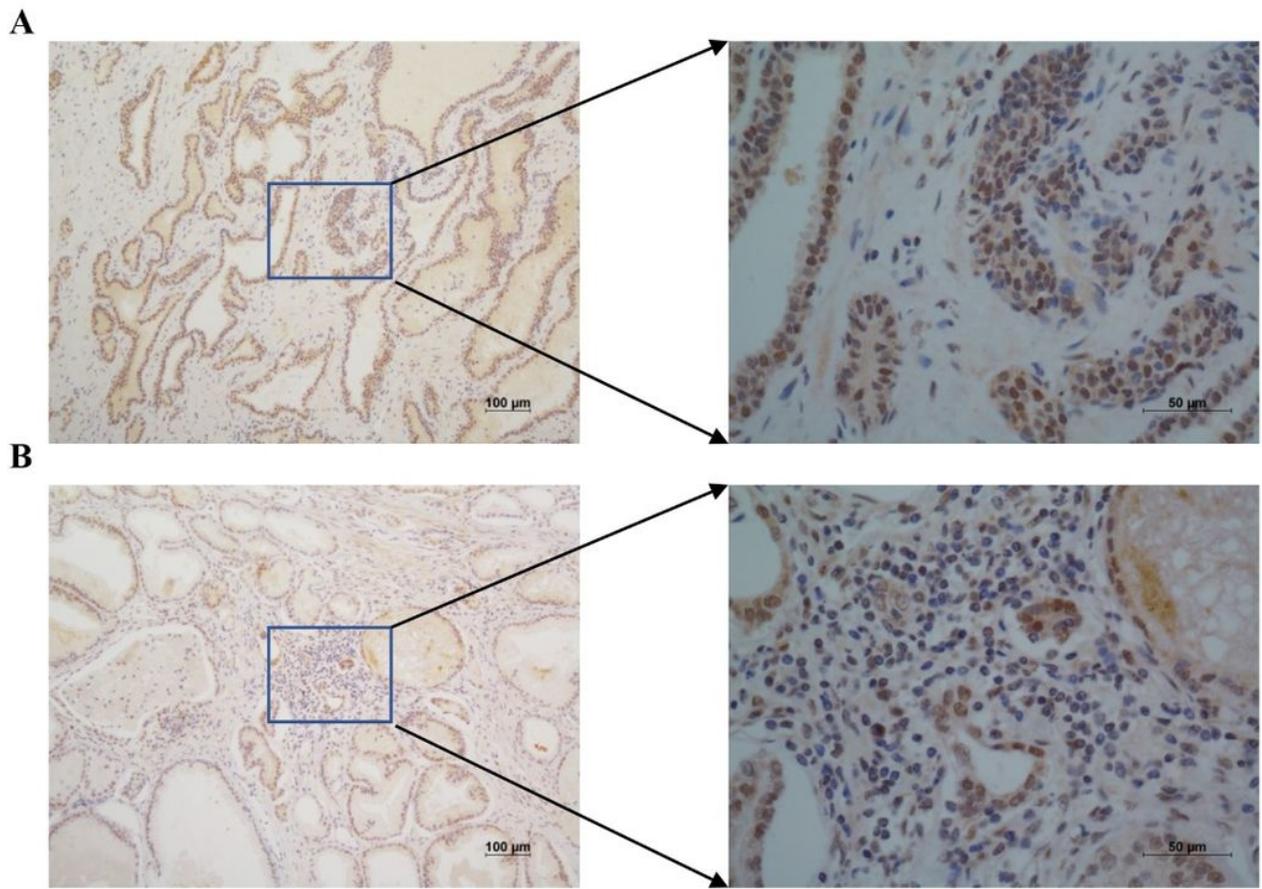
**Figure 3**

Identification of modules associated with clinical profiles in the GSE70768 dataset. (A) The Cluster dendrogram of co-expression network modules was created by a hierarchical clustering of genes based on the 1-TOM matrix. Each module was assigned different colors. (B) Module-trait relationships. Each row corresponds to a color module and each column correlates to clinical profiles (cancer and normal). Each cell contains the corresponding correlation and P-value.



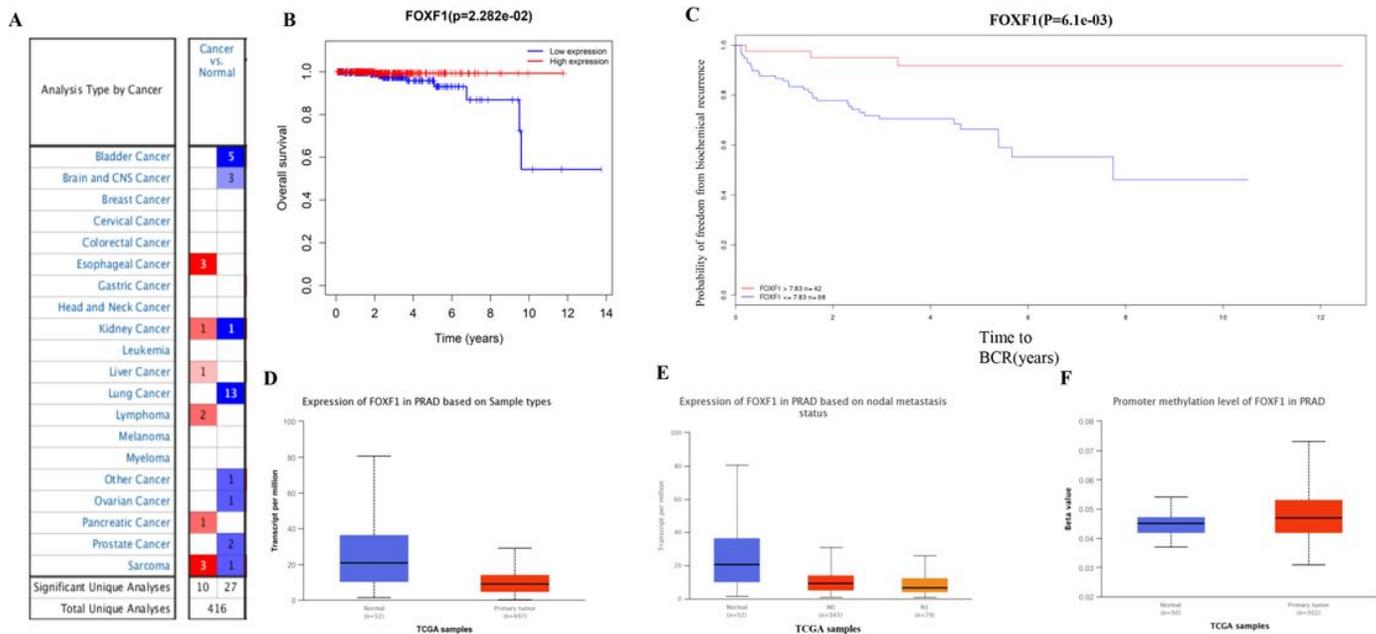
**Figure 4**

Identification of differentially expressed genes (DEGs) among the TCGA-PRAD and GSE70768 datasets of prostate cancer (A) Volcano plot of DEGs in the TCGA-PRAD dataset. (B) Volcano plot of DEGs in the GSE70768 dataset. (C) The Venn diagram of genes among DEG lists and co-expression module. In total, 108 overlapping genes in the intersection of DEG lists and two co-expression modules.



**Figure 5**

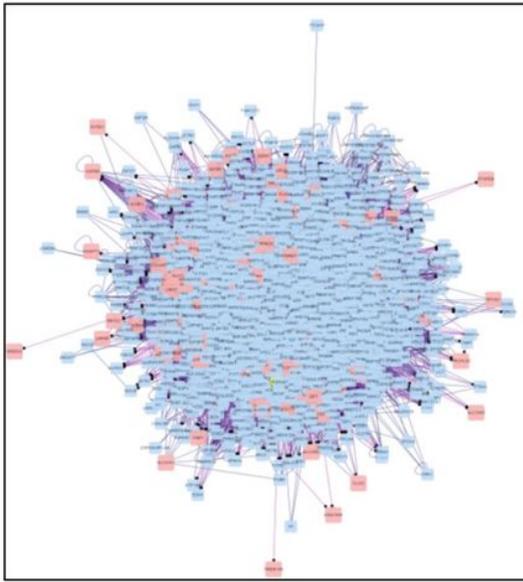
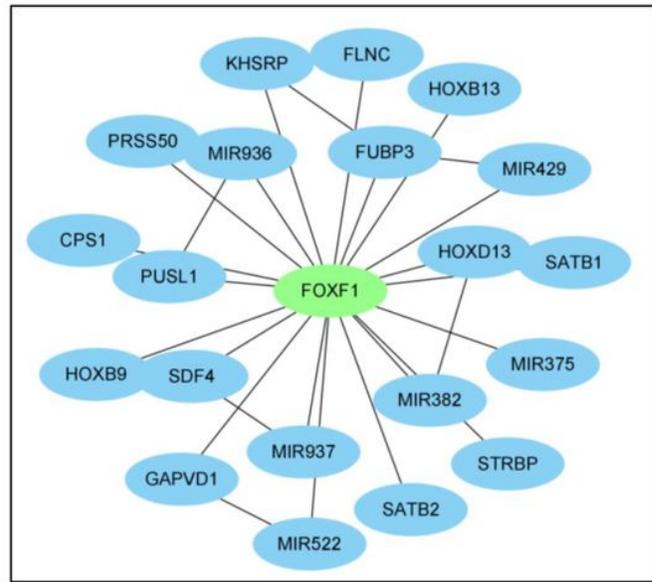
Differential expression level of FOXF1 in prostate cancer tissue(A) and Corresponding adjacent tissue(B).



**Figure 6**

This figure describes the correlation between CCNA2 and clinical features and some cytological features. (A) Expression of Foxf1 in different cancer types based on the OncoPrint platform. (B) Kaplan-Meier curves for the OS of patients in the high expression and low expression of Foxf1. (C) biochemical recurrence rate of prostate cancer of patients in the high expression and low expression of Foxf1. (D) expression of Foxf1 in sample of normal and cancer in TCGA-PRAD. (E) expression of Foxf1 in TCGA-PRAD based on nodal metastasis. (F) promoter methylation level of Foxf1 in TCGA-PRAD.



**A****B**

## Figure 8

Construction of the PPI networks. (A) PPI network of the genes between DEG lists and two co-expression modules. (B) a PP1 network associated with FOFX1 by using algorithm of Cytohubba plugin.