Identification of differentially expressed genes, biological pathways and prognostic signature in bladder cancer

Fucai Tang
The Eighth Affiliated Hospital of Sun Yat-Sen University
Xiayan Qian
The First School of Clinical Medicine, Southern Medical University
Zeguang Lu
The Second Clinical College of Guangzhou Medical University
Yongchang Lai
The Eighth Affiliated Hospital of Sun Yat-Sen University
Zhibiao Li
Third Affiliated Hospital of Guangzhou Medical College
Chengwu He
The Eighth Affiliated Hospital of Sun Yat-Sen University
Zhaohui He (hechh9@mail.sysu.edu.cn)
The Eighth Affiliated Hospital of Sun Yat-Sen University https://orcid.org/0000-0003-2289-9603

Research article

Keywords: bladder cancer, TCGA, GEO, differentially expressed genes, prognostic signature

Posted Date: March 31st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-362542/v1

Abstract

Background

Bladder cancer (BC) is one of the most common malignant cancer of urinary system in the worldwide. The purpose of the present study was to analysis differentially expressed genes (DEGs), biological pathways and prognostic significance BC by bioinformatics analysis.

Methods

The gene expression dataset GSE7476 and the mRNA Seq sequencing data were downloaded respectively from GEO and TCGA. A total of 220 DEGs were obtained in BC. GO analysis and KEGG pathway analysis were performed for up- and down-regulated DEGs. Then, a protein-protein interaction (PPI) networks and module were constructed by Cytoscape software. Survival analysis of hub genes was performed.

Results

The result of GO analysis revealed that the up-regulated DEGs were enriched mainly in sister chromatid segregation, while the down-regulated DEGs were enriched mainly in muscle contraction. The result of KEGG pathway analysis showed that up-regulated DEGs were enriched mainly in cell cycle, while down-regulated DEGs enriched in IL-17 signaling pathway. 41 hub gene and 3 crucial modules were identified in the PPI network. 15 genes significantly associated with patient prognosis in BC were obtained by Kaplan-Meier analysis.

Conclusions

In summary, the present study identified hub genes, crucial pathways and provide possible the molecular targets and prognostic biomarkers for targeted therapy and prognostic assessment of BC.

Introduction

As a malignant tumor, bladder cancer (BC) threatens seriously to human health. Worldwide, BC is the ninth among malignant tumors (1) while it is the most common urologic malignancy in China (2). The risk factors for BC include Tobacco smoking and occupational exposure to carcinogens (3). The majority of BC consists of urothelial carcinoma as the predominant histologic type (4). Clinically, the confirmed diagnosis of BC is mainly cystoscopy (5). And surgical treatment is the first treatment choice. With the development of surgical operation, the treatment of BC enters the epoch of precision medicine. Investigation of the etiological factors and therapeutic targets of disease at the molecular level improves the therapeutic effect of diseases (5). However, the pathogenesis of BC and the mechanism of its occurrence, development and recurrence are still unclear. Therefore, it is of great value to investigate deeply the molecular mechanism of apoptosis,

proliferation, metastasis and invasion of BC cells for improving the prevention, diagnosis and treatment of BC.

In recent years, biomarkers of bladder cancer have become a research focus. In previous studies, somatic copy number changes are found in multiple regions, including amplification of PPARG and E2F3, and deletion of CDKN2A and RB1 (6) (7). The Cancer Genome Atlas (TCGA) project reported that 69% of BC have identified potential therapeutic targets, 42% have targets in the phosphatidylinositol-3-oh kinase/AKT/mTOR pathway, and 45% have targets in the RTK/MAPK pathway (8). Currently, no effective definitely molecular or gene-targeted drugs have been approved for the treatment of this disease. Therefore, excavation of the related genes and biomarkers involved in pathogenesis and molecular mechanism of BC is of great vital in the diagnosis and treatment of patients.

With the development of bioinformatics and molecular biology, gene chips and high-throughput sequencing have become important means for obtaining information about cancer gene expression profiles in a large-scale and highly efficient manner (9). The gene expression omnibus (GEO) database is a comprehensive public gene expression data resource that contains a variety of tumor gene expression profile datasets (10). The Cancer Genome Atlas (TCGA) is utilized to identify genetic mutations implicated in cancer by using genome sequencing and bioinformatics (11). In the present study, we obtained the significantly differentially expressed genes(DEGs) by downloading BC chip and genome sequencing data from GEO and TCGA databases and combining gene data with R software. GO and KEGG pathway analysis were conducted to investigate the functional enrichment of significantly DEGs. In addition, by building a PPI network of significantly differential genes, we identified the key genes and crucial modules. Prognosis correlation analysis of the key genes were conducted by combining with the prognostic information of TCGA database. These genes validated by using the BC tissue samples might provide possible the molecular markers and new therapeutic targets for prognostic assessment and targeted therapy of BC.

Materials And Methods

Microarray data

The gene expression dataset GSE7476 of BC and normal bladder tissues were downloaded from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database. The GSE7476 dataset included 12 normal bladder samples and 43 BC samples. GPL570 Affymetrix Human Genome U133 Plus 2.0 Array platform was used for detection. The mRNA Seq sequencing data and clinical follow-up data of the BC and para-bladder cancer tissues samples were downloaded from the Cancer Genome Atlas (TCGA, https://cancergenome.nih.gov/), including a total of 19 para-bladder cancer tissues and 414 BC tissues.

Data preprocessing and screening of DEGs

We downloaded the original matrix file, and used limma package for the expression value background correction and the normalized preprocessing of the expression spectrum data. The unpaired *t* test provided by limma package was used to calculate the differential expression analysis of BC and para-bladder cancer tissues in each gene expression set, and we obtained the corresponding P-value of all the genes. Using

Benjamin & Hochberg (12) method for multiple check and correction, we got the corrected P-value, namely adjust P value. The adjust P value < 0.05 and |log2FC| > 2 were set as the thresholds for identifying DEGs.

We used edgeR package (Version: 3.4, http://www.bioconductor.org/packages/release/bioc/html/edgeR.html) for differential expression analysis of the mRNA Seq sequencing data of BC and para-bladder cancer tissues which from TCGA data. The false discovery rate (FDR) < 0.05 and |log2FC| > 2 were set as the thresholds for identifying DEGs. The differentially expressed up- and down-regulated genes from the above two data sets were intersected to obtain the final DEGs.

Gene Ontology (GO) analysis of DEGs

GO analysis is an internationally standardized Gene functional classification system, which provides a set of dynamically updated standard vocabularies to comprehensively describe the properties of genes and Gene products in organisms. GO analysis can make simple annotation of genes, namely, classification annotation of Biological Process (BP), Cellular Component (CC) and Molecular Function (MF). We conducted GO analysis on the up- and down-*DEGs* by R clusterProfiler package(Bioconductor version: Release (3.8); http://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html) (13–15). The p.adjust < 0.05 and the number of parameter enriched genes ≥ 2 were set as the statistically significant difference.

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of DEGs

KEGG pathway is a public database related to signaling pathways. It can use hypergeometric test to find the significant signal pathway of DEGs compared with the whole genome as the background, so as to obtain the major functional and metabolic pathways involved in DEGs. We conducted KEGG Pathway analysis on the up- and down-*DEGs* by R clusterProfiler package(Bioconductor version: Release (3.8); http://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html)(13–15). The p.adjust < 0.05 and the number of parameter enriched genes ≥ 2 were set as the statistically significant difference.

PPI network construction and the screening and functional analysis of sub network modules

STRING database was used to analyze the differential proteins of significantly DEGs, and constructed a network of Protein-Protein Interaction (PPI). Based on the general histological principles of the functional cell system, PPI is able to find functional associations among proteins on a genome-wide scale. We selected Required Confidence (combined score) > 0.4 as the threshold of protein-protein interaction, constructed the network graph for the obtained PPI by using Cytoscape (Version 3.5.1) software, and analyzed the Degree of nodes in the network by using network topological properties. The higher the degree of the node was, the more important it was in the network. In present study, Molecular Complex Detection (MCODE) (16) plugin in cytoscape was utilized to conduct the sub network module analysis in PPI network. The following parameters were Set: Node score cutoff ≥ 2; degree cutoff ≥ 2; max depth = 100; and K-core ≥ 2. And conducted KEGG pathway analysis on the module genes.

Survival analysis

Prognostic information for BC were downloaded from the TCGA database, including total survival time and vital status. The survival package of R was utilized to investigative the prognostic value in BC patients. All samples were divided into either the high-expression group and the low-expression group according to the median in terms of the expression of each hub-gene. The survival curves were plotted for the two groups by the Kaplan-Meier method, and perform log-rank to analyze the prognostic correlation. $P < 0.05$ was considered as statistically significant difference.

Results

Data preprocessing and screening of DEGs

The gene expression dataset (GSE7476) was screened and downloaded from the GEO database. The platform file GPL570-55999 was used for the probe and gene matching, and a total of 20484 genes were obtained. A total of 359 DEGs were identified in BC from GSE7476 dataset by using the limma package of R, including 62 up-regulated genes and 297 down-regulated genes. And, a total of 2,106 DEGs were identified in TCGA via edgeR package, including 1,293 up-regulated genes and 813 down-regulated genes. Finally, a total of 220 DEGs (Table S1) were obtained by intersecting the up- and down-regulated DEGs in the two data sets (Fig. 1), among which 28 DEGs were up-regulated and 192 DEGs were down-regulated.

GO analysis of DEGs

The clusterProfiler package of R was used to annotate 28 up-regulated significantly DEGs and 192 down-regulated DEGs. We obtained a total of 172 BP categories by GO analysis of up-regulated DEGs, including sister chromatid segregation, mitotic sister chromatid segregation, mitotic nuclear division and other regulatory functions; 35 CC categories, including spindle, cyclin B1-CDK1 complex, anaphase-promoting complex and other cell components; 4 MF categories, including histone kinase activity, DNA binding, bendinga, cyclin-dependent protein serine/threonine kinase activity and cyclin-dependent protein kinase activity molecular functions (Fig. 2A). A total of 124 BP categories were obtained by GO analysis of down-regulated DEGs, including muscle contraction, muscle system process, regulation of muscle contraction and other regulatory functions; 35 CC categories, including contractile fiber part,contractile fiber□myofibril and other cell components; 14 MF categories, including glycosaminoglycan binding□structural constituent of muscle□heparin binding and other molecular functions (Fig. 2B).

KEGG pathway enrichment analysis results of DEGs

Through R clusterProfiler package, 28 up-regulated DEGs and 192 down-regulated DEGs were analyzed for KEGG pathway signal pathway enrichment. The up-regulated DEGs were mainly enriched in 6 signaling pathways, including Cell cycle, Oocyte meiosis, Cellular senescence and so on. (Fig. 3A), while the down-regulated DEGs involved in 15 signaling pathways, mainly including IL-17 signaling pathway, vascular smooth muscle contraction,AGE-RAGE signaling pathway in diabetic complications and son on (Fig. 3B)

PPI network construction and the establishment of sub network module

We analyzed the DEGs by STRING database, obtained PPI network, and introduced it into Cytoscape software for analysis. There are 178 nodes and 1383 relationship pairs in PPI network (Fig. 4A). gene with node degree ≥ 10 were defined as the hub genes. A total of 41 hub genes were obtained (Table 1). Three important module of PPI network (score > 5) were identified via Cytoscape MCODE (Fig. 4B-D). KEGG pathway analysis was carried out for the node genes of the three modules. The node genes of module 1 were enriched with a total of 4 signaling pathways, which were mainly enriched in Cell cycle, Oocyte meiosis, Cellular senescence and p53 signaling pathway (Table 2); The node genes of module 2 were enriched with a total of 44 signaling pathway, which were enriched in IL-17 signaling pathway, C-type lectin receptor signaling pathway and Viral protein Human T-cell leukemia virus 1 infection and so no (Table 2); The node genes of module 3 were enriched with 2 signaling pathways, which were enriched in Vascular smooth muscle contraction and Tight junction (Table 2).

Table 1
The hub genes in PPI network

Gene		Degree	Gene		Degree
IL6	DOWN	27	MYL9	DOWN	15
JUN	DOWN	26	MYH11	DOWN	15
FOS	DOWN	24	TRIP13	UP	15
UBE2C	UP	23	KIF4A	UP	15
CDK1	UP	22	KIF20A	UP	15
CDC20	UP	21	DLGAP5	UP	15
PTGS2	DOWN	20	BUB1B	UP	15
EGR1	DOWN	20	FLNC	DOWN	14
CCNB1	UP	20	DUSP1	DOWN	14
TPM1	DOWN	19	NR4A2	DOWN	13
ATF3	DOWN	18	GINS1	UP	13
CCL2	DOWN	17	CTGF	DOWN	13
FOXM1	UP	17	CYR61	DOWN	13
TOP2A	UP	17	TAGLN	DOWN	13
FOSB	DOWN	17	KLF4	DOWN	12
UHRF1	UP	16	LMOD1	DOWN	12
MELK	UP	16	MYOCD	DOWN	11
ACTC1	DOWN	16	EGR2	DOWN	11
TPX2	UP	16	CNN1	DOWN	10
ZWINT	UP	16	ACTG2	DOWN	10
FLNA	DOWN	15			

Table 2
KEGG pathway analysis of the top three sub network module

Module		Name	Count	p.adjust	Genes
Module 1	hsa04110	Cell cycle	4	2.31E-05	BUB1B/CDK1/CCNB1/CDC20
	hsa04114	Oocyte meiosis	3	0.00082404	CDK1/CCNB1/CDC20
	hsa04218	Cellular senescence	3	0.00106494	CDK1/CCNB1/FOXM1
Module 2	hsa04657	IL-17 signaling pathway	5	8.95E-06	FOS/JUN/CCL2/IL6/PTGS2
	hsa04625	C-type lectin receptor signaling pathway	5	8.95E-06	EGR3/EGR2/JUN/IL6/PTGS2
	hsa05166	Human T-cell leukemia virus 1 infection	6	8.95E-06	ZFP36/EGR2/FOS/JUN/EGR1/IL6
Module 3	hsa04270	Vascular smooth muscle contraction	3	0.00066101	MYH11/ACTG2/MYL9
	hsa04530	Tight junction	2	0.03249374	MYH11/MYL9

	gene
up	TOX3 HMGB3 CCNB1 KIF20AGINS1 BUB1B DLGAP5 TOP2AINA OXM1 TRIP13 ZWINT CA2 KIF4A CDK1 HIST1H2BD MELK TPX2 CDC20 UHRF1 PTPRR C4orf48 HIST1H1C UBE2C MAGEA6 MMP1 FCRLB PM20D1
down	

Supplementary Table 1. DEGs in up- and down-regulated differential expression

Module	Name	Count	p.adjust	Genes
				MYOC SYNM MYH11 ACTC1 SCARA5 LRRC3B SYNPO2 ASB5 FLNC PDZRN4 CASQ2 MYOCD HSPB6 ATP1A2 HLF CNN1 C2orf40 OGN ADH1B P2RX1 DES LMOD1 PRUNE2 FHL1 SORBS1 ACTG2 C16orf89 PLP1 PTGS1 TCF21 PCP4 NEGR1 GPR133 PLN FAM107A C7 CPED1 KCNMB1 MYL9 REEP1 NR4A3 SDPR FOSB PCOLCE2 FAM129A SCN7A C1QTNF7 EGR3 PRELP CRYAB ZBTB16 CFD RBPMS2 LPPR4 RAB9B MAMDC2 AOX1 PDK4 TMOD1 MSRB3 ITIH5 FGL2 IL6 ASB2 TPM1 AOC3 DPT FXYD6 EGR1 PTGIS ADAMTS1 CSRP1 BMP5 NEXN SBSPON BCHE MAP1B ASPA TAGLN PRDM6 FLNA TCEAL2 CH25H CFL2 CCL2 RCAN2 HRDL1 GEM PRIMA1 PDLIM3 CILP TMEM100 C11orf96 NR3C2 ATF3 PPP1R14A MFAP4
Supplementary Table 1. DEGs in up- and down-regulated differential expression				

Module	Name	Count	p.adjust	Genes
				SPARCL1PALLD ACKR1 DUSP1 P2RY14 GPRASP1 FOS RASL12 MRVI1 FOXF1 FXYD1 PLCB4 MRGPRF RBM24 CTGF NDNF RAB23 CYR61 SMOC2 SH3BGR ZFP36 SOX17 NR4A2 RGS2 PAMR1 FERMT2 NBEA SLIT2 CSRNP1 RAI2 ZFPM2 LRRN4CL DDR2 ZCCHC24 CCDC80 LYVE1 COX7A1 JAM2 AKAP12 HSD17B6 MYOT BAG2 FILIP1L LRCH2 FBXL7 WFDC1 TCEAL7 SELE ITGA5 EBF1 SYNC MFAP5 IER3 SRPX KLF4 FBLN5 TMPRSS11E APOLD1 GHR AQP1 SPON1 MYADM GPR124 DPYSL3 PTGS2 COL21A1 PTGDS PTX3 PPAP2B SLC24A3 CPA3 EGR2 PRICKLE2 TGFB1I1 OLFML1 CYBRD1 DUSP5 CAP2 PRAC1 CLIC4 TPSB2 KLHL13 INMT DNAJB4 TSLP CRISPLD2 JUN TGFBR3 PKIG COLEC12 ITM2A TSHZ3 SETBP1 JAZF1 C8orf4

Supplementary Table 1. DEGs in up- and down-regulated differential expression

Survival analysis

The survival time and status of patients were downloaded from the TCGA database, and the survival analysis of hub genes was carried out by combining the differential expression values of genes.15 hub genes related to prognosis were obtained. They are JUN(P = 0.034), CDC20(P = 0.032), CCNB1(P = 0.00064), TPM1(P = 0.0033), ACTC1(P = 0.0079), MYH11(P = 0.021), BUB1B(P = 0.042), DLGAP5(P = 0.042), FLNC(P = 0.00012), CTGF(P = 0.017), CYR61(P = 0.033), TAGLN(P = 0.018), LMOD1 (P = 0.025), EGR2(P = 0.0079) and CNN1(P = 0.014). The survival curve (Fig. 5) indicated that the survival rate of high expression of all hub genes was poorer than that of low expression.

Discussion

BC is the most common urologic malignancy (2) with the feature of high recurrence and high mortality in China (17). Bladder urothelial carcinoma is the predominant histologic type (4). Currently, resection of bladder tumors is the most common surgical operation for noninvasive bladder cancer, but the recurrence rate is still high (18). Therefore, the investigations of BC using bioinformatics may contribute to explore the hub genes and important signaling pathways, reveal the development mechanism, and provide possible molecular markers and potential therapeutic targets for the clinical diagnosis and treatment of BC.

In the present study, we downloaded a dataset GSE7476 from GEO and the sequencing data of BC from TCGA. Comparison of gene data between the BC tissues and para-cancer tissues were performed by R software. After the intersection, 220 DEGs were obtained, including 28 up-regulated genes and 192 down-regulated genes. The up-regulated DEGs were mainly enriched in the BP, mainly including sister chromatid segregation, mitotic sister chromatid segregation and mitotic nuclear division; in the CC, it was mainly concentrated in spindle, cyclin B1-CDK1 complex and anaphase-promoting complex; in terms of MF, it is mainly concentrated in the histone kinase activity, DNA binding bending and cyclin-dependent protein serine/threonine kinase activity. The down-regulated DEGs were also mainly enriched in the BP, including muscle contraction, muscle system process, regulation of muscle contraction and so on; in the CC, the main components were such as the contractile fiber part, contractile fiber, myofibril; in terms of MF, it is mainly concentrated in glycosaminoglycan binding,structural constituent of muscle and heparin binding. The KEGG pathway analysis showed that the up-regulated DEGs were co-enriched in 6 signaling pathways, which might play an important role in the cell cycle signaling pathway. Cell cycle includes cell replication and cell division, and it is closely related to cell growth, anabolism, and proliferation. The abnormal cell proliferation and DNA replication are important causes of cancer. In the past, many studies have reported that cell cycle plays an important role in the occurrence and development of tumors (19). In the occurrence process of tumor, it is common that the cell grow out of control, basically manifested as cell cycle supernormal operation out of control (20). Therefore, the changes of cell cycle affect in all probability the normal proliferation and apoptosis of cells, and in-depth study of cell cycle is conducive to revealing the specific process of the occurrence and development of tumors, sequentially helping to find key therapeutic targets. As shown by KEGG pathway analysis, the down-regulated DEGs showed that these genes were involved in 15 signaling pathways, and were most likely to play a significant role in Interleukin 17(IL-17) pathway. IL-17 is a pro-inflammatory cytokine secreted primarily by activated CD4 (+) T helper cells, CD8 (+) T cells, and

macrophages (21). It's main biological functions include inducing tumor necrosis factor, stimulating vascular endothelial cell migration and angiogenesis (22, 23). More and more reports indicate that IL-17 plays an important role in the occurrence and development progression of tumors, and its expression is increased in prostate cancer (24), colorectal cancer (25), breast cancer (26), ovarian cancer (27), hepatocellular carcinoma (28), non-small cell lung cancer (29) and other tumors. It has been reported that IL-17 could promote the formation of blood vessels (30), shield the body from immune surveillance of cancer cells (31) and promote the occurrence and development progression of tumors. However, it has also been reported that IL-17 could inhibit tumor progression by enhancing the body's immune resistance to tumor, thereby preventing tumor growth and metastasis (32, 33). Therefore, the study of these signaling pathways may help clarify the carcinogenic mechanism of BC.

In the present study, PPI protein network was constructed to screen out 41 hub gene and three functional modules. These modules are mainly concentrated in signaling pathways such as Cell cycle, IL-17 signaling pathway and Vascular smooth muscle contraction. Then the prognostic survival of hub genes was analyed by TCGA clinical follow-up data, and 15 hub genes significantly related to prognosis were found. These results suggested that these hub genes might play an important role in the development and prognosis of BC.

Encoded by JUN, c-Jun is an oncogenic transcription factor. Wang SS et al. report that JUN is noted in the high-risk groups of BC patients with higher expression, while the lower JUN expression groups have the longer overall survival (34). Li Z et al. demonstrate that c-Jun is significantly overespressed in aggressive bladder cancer tissues and cell lines, and c-Jun is related with anti-proliferative factor (APF)-mediated inhibition for bladder cancer cell growth (35). These reported suggest that JUN may inhibit the growth of BC cell and lower expression of JUN may indicate the better BC prognosis. ACTC1 is reported to be down-regulated in BC (36), which is consistent with the result of our RT-PCT experiment. The previous research manifest that the expression of ACTC1 gradually increases according to WHO grade in glioma and ACTC1-positive groups had a shorter survival time and poorer prognosis (37). However, our research found that the expression of ATCT1 is closely related to the prognosis of BC, and the low expression of ATCT1 indicates a good prognosis. ATCT1 may be a biomarker for prognosis evaluation of BC patients. Of course, more research is also needed to verify this. In regards to connective tissue growth factor (CTGF), Chen et al. make a conclude that urothelial carcinoma cells show increasing apoptosis using extrinsic pathways through upregulation of CTGF (38). Previous researches reveal that CTGF can promoted the progression of tumor cell and significantly correlated with poor clinical prognosis in breast cancer(39).Yuen *et al* (40) found that TAZ-AXL-CTGF overexpression are associated with colon cancer progression. In addition, CTGF and CYR61 both belong to CCN (CYR61, CTGF, NOV) family of protein. Nevertheless, Jiang et al. show that the members of CCN family play contradictory roles in progression of breast cancer with CYR61 appearing to be an aggressive stimulator and CTGF acting as a cancer suppressor (41). Our results show that CTGF and CYR61 are related to the prognosis of BC patients, and down-regulated in BC. These molecules are pivotal targets affecting the prognosis of BC although the role of these molecules in BC remains obscure. As a member of the calponin family of actin-binding proteins, Transgelin (TAGLN) participates in cell motility and migration (42). As is well-known, cell migration is the crucial process of cancer metastasis. Wu et al. certified that TAGLN overexpression promote the tumor progression in lung adenocarcinoma (43). Calponin h1(CNN1) is

certified to have the function of cytoskeleton-stabilizing and transformation-inhibiting by combining with actin and related proteins (44). Therefore, CNN1 is possible to inhibit the tumor cell growth in the process of cancer development. As a matter of fact, it is reported that downregulation of CNN1 is necessary for cell transformation, and low of CNN1 expression is crucial for its metastasis in ovarian high-grade serous carcinoma (44). Liu Y identify that TAGLN and CNN1 are connected to both overall survival and disease-free survival in BC and act as potential prognostic molecular markers (45). Our researched also showed that TAGLN and CNN1 expression was down-regulated in BC and closely related to the prognosis of patients with BC. The hub gene EGR2 (early growth response 2 protein) which as a cancer suppressor is down-regulated in gastric cancer, and the oncogenic effect of miR-20a is significantly attenuated via up-regulated EGR2 expression (46). Low expression of EGR2 is associated with cancer metastasis and shorter survival in gastric cancer patients (47). Unoki et al. demonstrats that EGR2 can induce the apoptosis of cancer cell lines by activating the expression of BNIP3L and BAK (48). In our study, low expression of EGR2 was found to be associated with a good prognosis in patients with BC. However, there are few studies on EGR2 in BC, and more studies are needed. Myosin heavy chain 11 (MYH11) belongs to myosin heavy chain family, and the mutation of MYH11 was related to is related to the development of intestinal cancer (49, 50). It has been also reported to be a biomarker for BC. Hu et al. reported that Myh11 expression was significantly lower in tumor samples and the high expression of MYH11 was related to worse OS of BC patients, which was consistent with our study (51). Filamin C (FLNC) is one member of the filamin protein family, which was associated with heart disease (52). The methylation of FLNC has reported in several cancers, such as prostate cancer (53) and gastric cancer (54). These reports showed that the downregulation of FLNC can induce the expression of matrix metallopeptidase, which can regulate the degradation of extracellular matrix and cancer metastasis.

Similarly studies show that FLNC with low expression in BC was involved in tumor migration and invasion (55). In addition, this study showed that FLNC is lowly expression in BC. And it was significantly associated with prognosis of BC patients.

Based on our results and the above literature, we believe that these hub genes may be potential prognostic markers and therapeutic targets for bladder cancer. There, it is necessary and meaningful for further study to investigate the specific mechanism of these genes in BC.

Conclusion

In summary, the present research screened out DEGs, hub genes and prognostic signature in bladder cancer by bioinformatics analysis. Furthermore, this study may provide the potential prognostic biomarkers and possible new therapeutic targets for bladder cancer. Future experimental study should be conducted to clarify the pathogenic mechanism of bladder cancer.

Declarations

Contributions: (I) Conception and design: Z He, F Tang, X Qian and Z Lu; (II) Administrative support: Z He, Y Lai; (III) Provision of study materials or patients: C He; (IV) Collection and assembly of data: J Wang; (V)

Data analysis and interpretation: Z Li, H Lei ; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Acknowledgements

Funding: The present study was supported by National Key Researchand Development Program of China (ontract No. 2018YFA0902801), National Natural Science Foundation of China (contract No. 81803576), Guangdong Basic and Applied Basic Research Foundation (ontract No. 2020A1515010152), and the research start-up fee for the Eighth Affiliated Hospital, Sun Yat-sen University (grant no. zdbykyqdf005).

Footnote

Conflicts of Interests: All authors have completed the ICMJE uniform disclosure form. All authors declare that they have no conflict of interest to state.

Availability of data and materials: The datasets of gene expression profiles for bladder cancer (GSE7476) are available in the GEO on the NCBI website (http://www.ncbi.nlm.nih.gov/geo).The mRNA Seq sequencing data and clinical follow-up data of the BC and para-bladder cancer tissues samples were downloaded from the TCGA (https://cancergenome.nih.gov/).

Ethics approval and consent to participate: In the present study, the data were downloaded from the GEO and TCGA databases; therefore, no studies with human participants or animals were performed by any of the authors. Thus, no ethical approval or patient consent was required.

References

1. D. M. Parkin, F. Bray, J. Ferlay, P. Pisani, Global cancer statistics, 2002. *CA: a cancer journal for clinicians* 2005;55**:**74-108.
2. Li Kaiwen, Lin Tianxin, Chinese Bladder Cancer Consortium et al. Current status of diagnosis and treatment of bladder cancer in China - Analyses of Chinese Bladder Cancer Consortium database. Asian J Urol 2015;2:63-69.
3. Cumberbatch Marcus George Kwesi, Jubber Ibrahim,Black Peter C et al. Epidemiology of Bladder Cancer: A Systematic Review and Contemporary Update of Risk Factors in 2018.Eur Urol 2018; 74:784-795.
4. Wong Martin C S, Fung Franklin D H, Leung Colette et al. The global epidemiology of bladder cancer: a joinpoint regression analysis of its incidence and mortality trends and projection.Sci Rep 2018; 8:1129.
5. Su Hongwei, Jiang Haitao,Tao Tao et al. Hope and challenge: Precision medicine in bladder cancer.Cancer Med 2019;8:1806-1816.
6. Forbes Simon A, Bindal Nidhi, Bamford Sally et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 2011; 39:D945-50.
7. P. J. Goebell, M. A. Knowles. Bladder cancer or bladder cancers? Genetically distinct malignant conditions of the urothelium. *Urologic oncology* 2010; **28**:409-428.

8. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma.Nature 2014;507:315-22.
9. Lu Xuesong, Zhang Xuegong.The effect of GeneChip gene definitions on the microarray study of cancers. Bioessays 2006;28:739-46.
10. Jiang Peng, Liu X Shirley. Big data mining yields novel insights on cancer. NatGenet 2015;47:103-4.
11. Zhu Qiandong, Sun Yunpeng, Zhou Qingqing et al. Identification of key genes and pathways by bioinformatics analysis with TCGA RNA sequencing data in hepatocellular carcinoma. Mol Clin Oncol 2018; 9:597-606.
12. T. J. Hardcastle. Generalized empirical Bayesian methods for discovery of differential data in high-throughput biology. Bioinformatics (Oxford, England) 2016; 32:195-202.
13. Yu Guangchuang, He Qing-Yu. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst 2016;12:477-9.
14. Yu Guangchuang, Wang Li-Gen, Han Yanyan et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284-7.
15. Yu Guangchuang, Wang Li-Gen, Yan Guang-Rong et al. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics 2015; 31:608-9.
16. Bandettini W Patricia, Kellman Peter, Mancini Christine et al. MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study. J Cardiovasc Magn Reson 2012;14:83.
17. Dai Qi-Shan, He Hui-Chan, Cai Chao et al. Multicenter case-control study of the relationship between smoking and bladder cancer in China. Zhonghua Yi Xue Za Zhi 2011;91:2407-10.
18. Gao Xin, Chen Yinyi, Chen Mei et al. Identification of key candidate genes and biological pathways in bladder cancer.PeerJ 2018; 6:e6036.
19. da Silva Glenda N, Evangelista Adriane F, Magalhães Danielle A et al. Expression of genes related to apoptosis, cell cycle and signaling pathways are independent of TP53 status in urinary bladder cancer cells. Mol Biol Rep 2011; 38:4159-70.
20. Zhou B B, Elledge S J. The DNA damage response: putting checkpoints in perspective. Nature 2000;408:433-9.
21. Rouvier E, Luciani M F, Mattéi M G et al. CTLA-8, cloned from an activated T cell, bearing AU-rich messenger RNA instability sequences, and homologous to a herpesvirus saimiri gene. J Immunol 1993;150:5445-56.
22. Numasaki Muneo, Fukushi Jun-ichi, Ono Mayumi et al. Interleukin-17 promotes angiogenesis and tumor growth. Blood 2003;101:2620-7.
23. Tartour E, Fossiez F, Joyeux I et al. Interleukin 17, a T-cell-derived cytokine, promotes tumorigenicity of human cervical tumors in nude mice.Cancer Res 1999; 59:3698-704.
24. Steiner Georg E, Newman Martin E, Paikl Doris et al. Expression and function of pro-inflammatory interleukin IL-17 and IL-17 receptor in normal, benign hyperplastic, and malignant prostate. Prostate 2003;56:171-82.

25. Le Gouvello S, Bastuji-Garin S, Aloulou N et al. High prevalence of Foxp3 and IL17 in MMR-proficient colorectal carcinomas. Gut 2008;57:772-9.
26. Zhu XingWu, Mulcahy Lori A, Mohammed Rabab A A et al. IL-17 expression by breast-cancer-associated macrophages: IL-17 promotes invasiveness of breast cancer cell lines. Breast Cancer Res 2008;10: R95.
27. Miyahara Yoshihiro, Odunsi Kunle, Chen Wenhao et al. Generation and regulation of human CD4+ IL-17-producing T cells in ovarian cancer. Proc Natl Acad Sci USA 2008; 105:15505-10.
28. Zhang Jing-Ping, Yan Jing, Xu Jing et al. Increased intratumoral IL-17-producing cells correlate with poor survival in hepatocellular carcinoma patients. J Hepatol 2009;50:980-9.
29. Numasaki Muneo, Watanabe Mika, Suzuki Takashi et al. IL-17 enhances the net angiogenic activity and in vivo growth of human non-small cell lung cancer in SCID mice through promoting CXCR-2-dependent angiogenesis. J Immunol 2005;175:6177-89.
30. Chang Seon Hee, Mirabolfathinejad Seyedeh Golsar, Katta Harshadadevi et al. T helper 17 cells play a critical pathogenic role in lung cancer. Proc Natl Acad Sci USA2014;111:5664-9.
31. Wu Dang, Wu Pin, Huang Qi et al. Interleukin-17: a promoter in colorectal cancer progression. Clin Dev Immunol 2013; 2013:436307.
32. Benchetrit Fabrice, Ciree Arnaud, Vives Virginie et al. Interleukin-17 inhibits tumor cell growth by means of a T-cell-dependent mechanism. Blood 2002; 99: 2114-21.
33. Kryczek Ilona, Banerjee Mousumi, Cheng Pui et al. Phenotype, distribution, generation, and functional and clinical relevance of Th17 cells in the human tumor environments. Blood 2009;114:1141-9.
34. 34. Wang Shi-Shuo, Chen Gang, Li Sheng-Hua et al. Identification and validation of an individualized autophagy-clinical prognostic index in bladder cancer patients. Onco Targets Ther 2019;12:3695-3712.
35. 35. Li Zeliang, Zhu Yuyan, Yu Meng et al. c-Jun is involved in interstitial cystitis antiproliferative factor (APF)-induced growth inhibition of human bladder cancer T24 cells. Urol Oncol 2013; 31:228-33.
36. 36. Zaravinos Apostolos, Lambrou George I, Boulalas Ioannis et al. Identification of common differentially expressed genes in urinary bladder cancer. PLoS ONE 2011; 6:e18135.
37. 37. Ohtaki Shunya, Wanibuchi Masahiko, Kataoka-Sasaki Yuko et al. ACTC1 as an invasion and prognosis marker in glioma. J Neurosurg 2017;126:467-475.
38. 38. Chen Shao-Kuan, Chung Chih-Ang, Cheng Yu-Che et al. Toll-like receptor 6 and connective tissue growth factor are significantly upregulated in mitomycin-C-treated urothelial carcinoma cells under hydrostatic pressure stimulation. Genet Test Mol Biomarkers 2014;18:410-6.
39. 39. Zhu Xiaoping, Zhong Jing, Zhao Zhen et al. Epithelial derived CTGF promotes breast tumor progression via inducing EMT and collagen I fibers deposition. Oncotarget 2015;6:25320-38.
40. 40. Yuen Hiu-Fung, McCrudden Cian M, Huang Yu-Han et al. TAZ expression as a prognostic indicator in colorectal cancer. PLoS ONE 2013; 8: e54211.
41. 41. Yuen Hiu-Fung, McCrudden Cian M, Huang Yu-Han et al. TAZ expression as a prognostic indicator in colorectal cancer. PLoS ONE 2013; 8: e54211.
42. 42. Ning Xin, Deng Yaoliang, Identification of key pathways and genes influencing prognosis in bladder urothelial carcinoma.Onco Targets Ther 2017; 10:1673-1686.

43. 43. Wu Xiaohong, Dong Liangliang, Zhang Ruifeng et al. Transgelin overexpression in lung adenocarcinoma is associated with tumor progression. Int J Mol Med 2014;34:585-91.
44. 44. Wang Kai-Hung, Chu Sung-Chao, Chu Tang-Yuan. Loss of calponin h1 confers anoikis resistance and tumor progression in the development of high-grade serous carcinoma originating from the fallopian tube epithelium. Oncotarget 2017;8:61133-61145.
45. 45. Liu *et al*., CALD1, CNN1, and TAGLN identified as potential prognostic molecular markers of bladder cancer by bioinformatics analysis. *Medicine***98**, e13847 (2019).
46. 46. Liu Yun, Wu Xia, Wang Guokun et al. CALD1, CNN1, and TAGLN identified as potential prognostic molecular markers of bladder cancer by bioinformatics analysis. Medicine (Baltimore) 2019; 98:e13847.
47. 47. Chen Mi, Fan Li, Zhang Si-Min et al. LINC01939 inhibits the metastasis of gastric cancer by acting as a molecular sponge of miR-17-5p to regulate EGR2 expression. Cell Death Dis 2019;10:70.
48. 48. Unoki Motoko, Nakamura Yusuke, EGR2 induces apoptosis in various cancer cell lines by direct transactivation of BNIP3L and BAK. Oncogene 2003;22: 2172-85.
49. Alhopuro P, Phichith D, Tuupanen S, et al. Unregulated smooth-muscle myosin in human intestinal neoplasia. Proc Natl Acad Sci USA 2008;105:5513-5518.
50. Ning X, Deng Y. Identification of key pathways and genes influencing prognosis in bladder urothelial carcinoma. Onco Targets Ther 2017; 10:1673-1686.
51. Hu J, Zhou L, Song Z, et al. The identification of new biomarkers for bladder cancer: A study based on TCGA and GEO datasets. J Cell Physiol 2019;10.1002.
52. Ortiz-Genga MF, Cuenca S, Dal Ferro M, et al. Truncating FLNC Mutations Are Associated With High-Risk Dilated and Arrhythmogenic Cardiomyopathies. J Am Coll Cardiol 2016;68:2440-2451.
53. Vanaja Donkena Krishna, Ehrich Mathias, Van den Boom Dirk et al. Hypermethylation of genes for diagnosis and risk stratification of prostate cancer. Cancer Invest 2009;27:549-60.
54. Kaneda A, Kaminishi M, Yanagihara K, Sugimura T, Ushijima T. Identification of silencing of nine genes in human gastric cancers. Cancer Res 2002;62:6645-6650.
55. Li W, Liu J, Zou D, et al. Exploration of bladder cancer molecular mechanisms based on miRNA-mRNA regulatory network. Oncol Rep 2017;37:1461-1468.

Figures

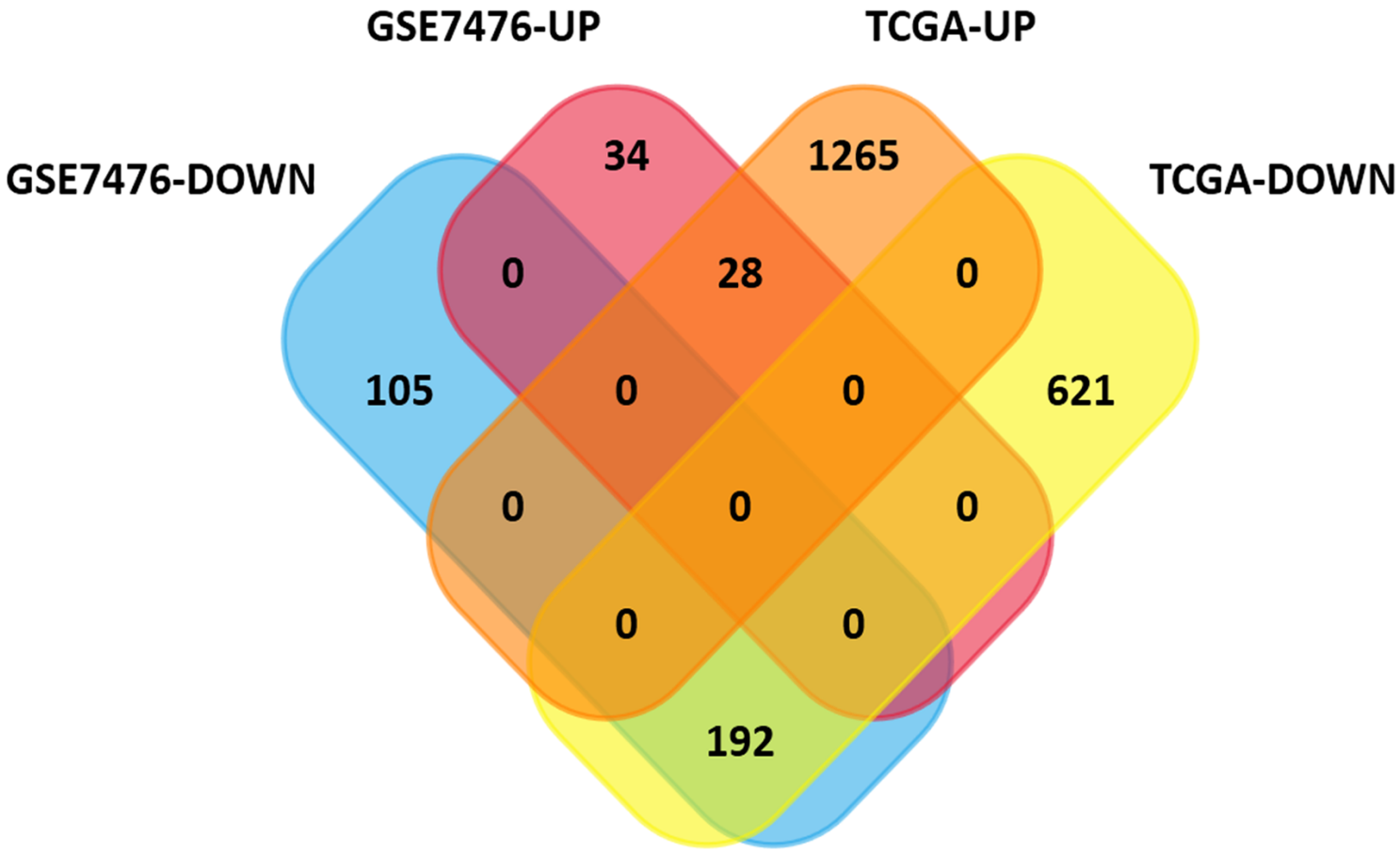

Figure 1

Venn diagram shows the number of up- and down-regulated DEGs in the two datasets, and cross regions represent for the intersection DEGs in different datasets. 28 up-regulated genes and 192 down-regulated genes were identified in GSE7476 and TCGA databases.

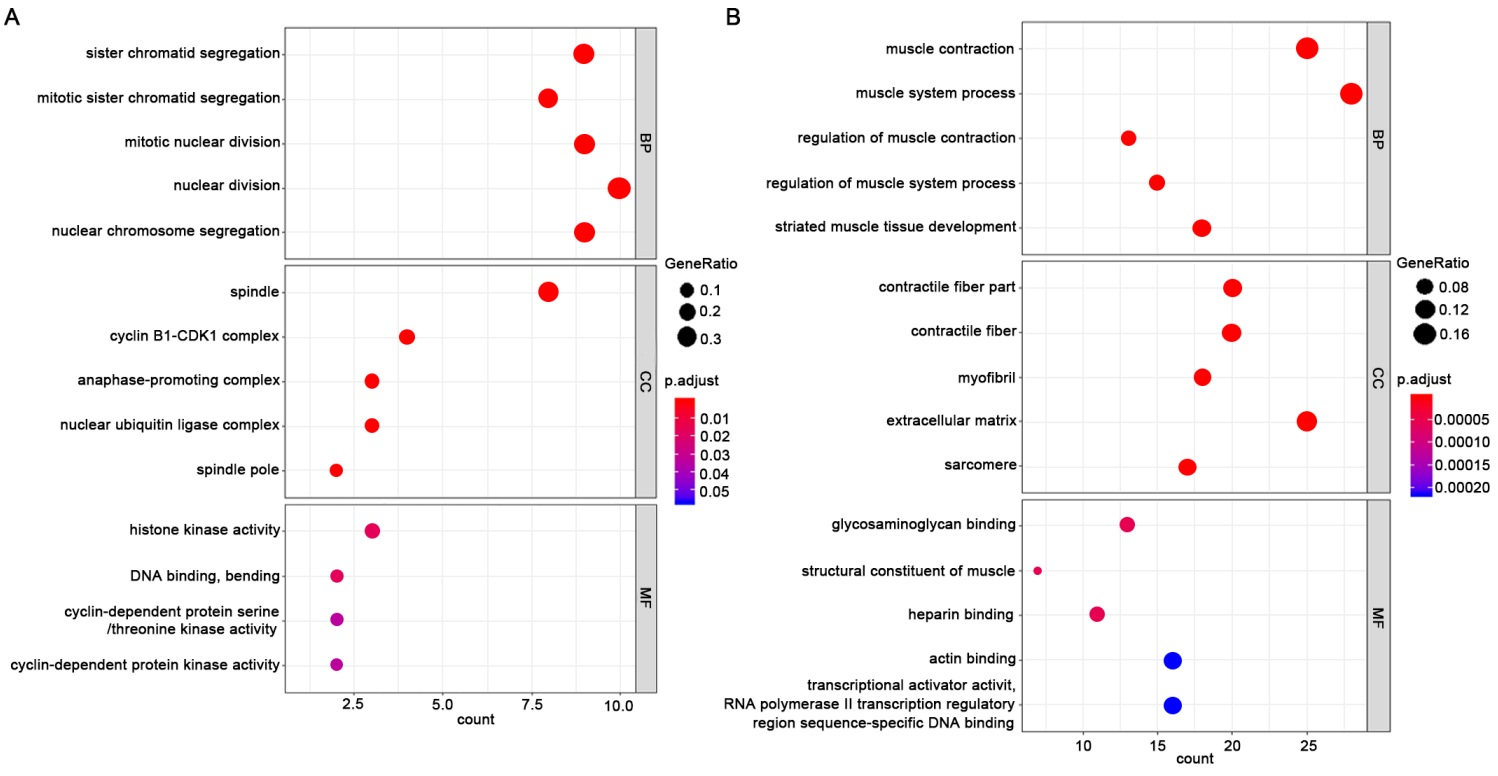

Figure 2

The top 5 enrichment of GO terms for up- and down-regulated DEGs. (A) The top 5 enrichment of GO terms for up-regulated DEGs; (B) The top 5 enrichment of GO terms for down-regulated DEGs; The X-axis indicates the number of genes, and the top 5 GO terms shown on the Y-axis; The color change between blue and red denote the significance of the correlation between low and high, and the different size of a dot represents related the gene ratio. DEG, differentially expressed genes. GO, Gene Ontology; BP, biological process; CC, cell components; MF, molecular function.

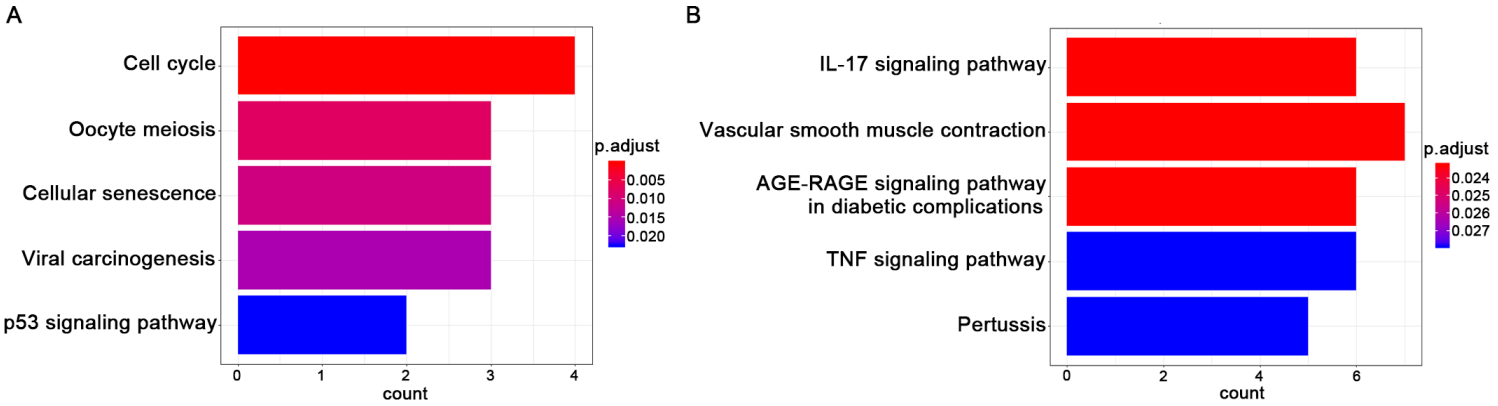

Figure 3

The top 5 KEGG pathway analysis results of up- and down-regulated DEGs. (A) The top 5 KEGG pathway of up-regulated DEGs; (B) The top 5 KEGG pathway of down-regulated DEGs. The X-axis represents the number of genes, and the top 5 KEGG pathway shown on the Y-axis; The color change between blue and red denote the significance of the correlation between low and high. KEGG, Kyoto Encyclopedia of Genes and Genomes; DEG, differentially expressed genes.

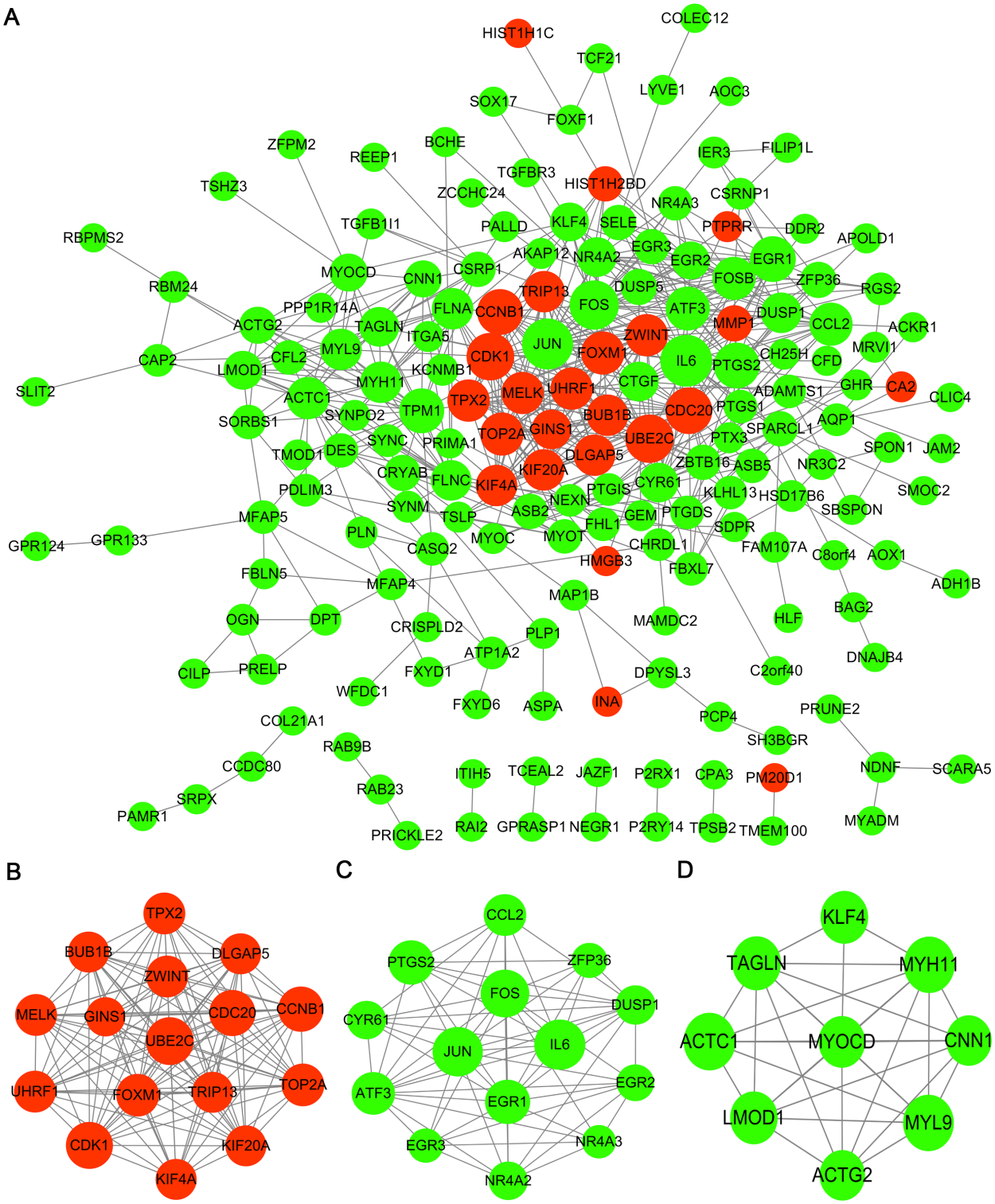

Figure 4

PPI network of DEGs and sub module of PPI network. (A) PPI network of significantly DEGs, (B) important sub network module 1 of PPI network, (C) important sub network module 2 of PPI network and (D) important sub network module 3 of PPI network. Red nodes represent up-regulated DEGs, while green nodes represent down-regulated DEGs. The increased nodes interactions indicate the greater biological significance. PPI, protein-protein interaction; DEG, differentially expressed genes.

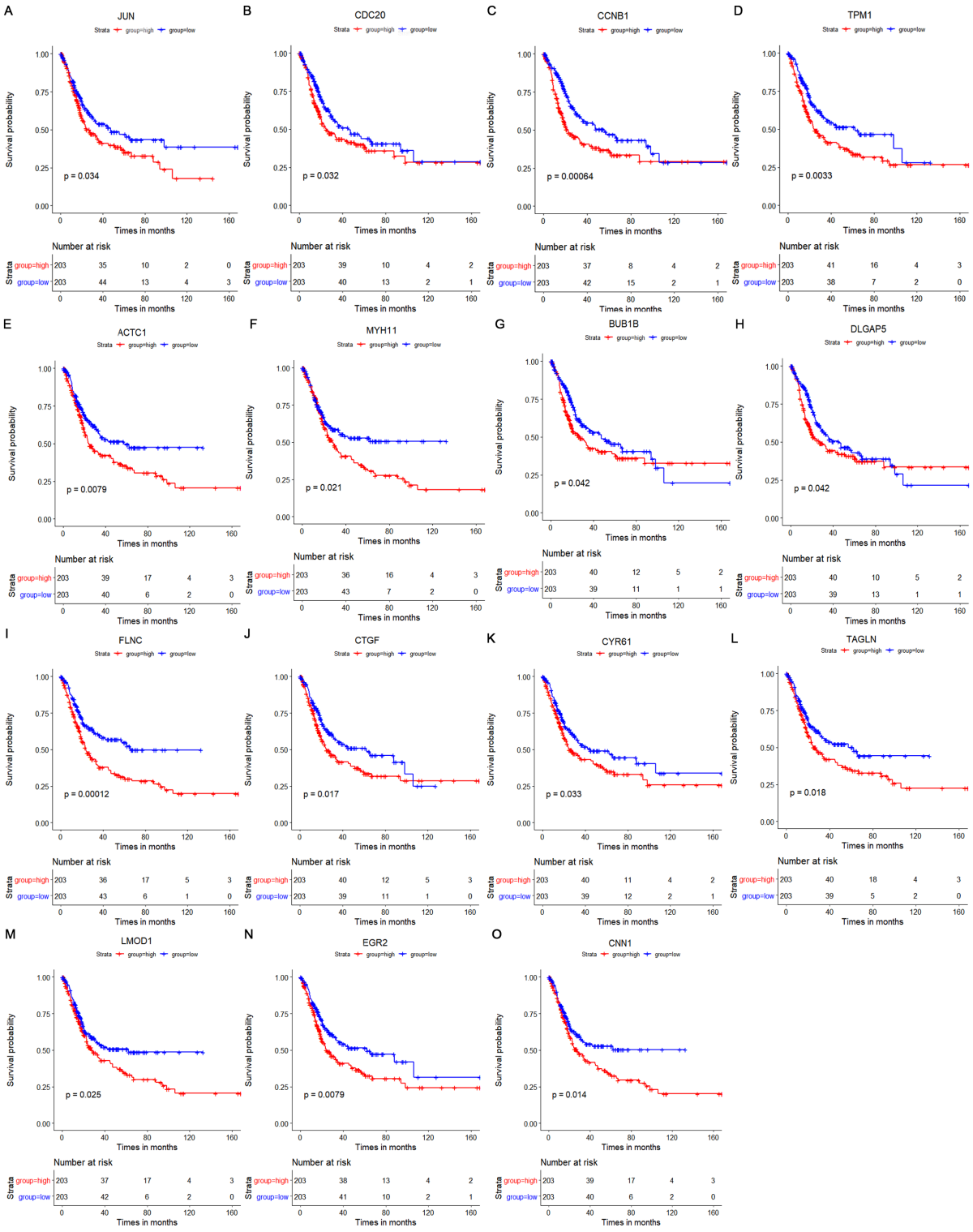

Figure 5

Prognostic relationship of hub genes in BC. (A) JUN; (B) CDC20; (C) CCNB1; (D) TPM1; (E) ACTC1; (F) MYH11; (G) BUB1B; (H) DLGAP5; (I) FLNC; (J) CTGF; (K) CYR61; (L) TAGLN; (M) LMOD1; (N) EGR2; (O) CNN1. The red line is the high expression group, while the blue line is the low expression group. The X-axis represents the total survival time (months), and while the Y-axis represents survival probability.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.docx