Received: 21 November 2018 | Revised: 27 February 2019 | Accepted: 6 March 2019

DOI: 10.1002/jcp.28546

ORIGINAL RESEARCH ARTICLE

WILEY Journal of Cellular Physiology

Four novel biomarkers for bladder cancer identified by weighted gene coexpression network analysis

Xin Yan[1,2] | Zi-Xin Guo[1] | Xiao-Ping Liu[1] | Yu-Jia Feng[1] | Ying-Jie Zhao[3] | Tong-Zu Liu[1] | Sheng Li[1,2,4]

[1]Department of Urology, Zhongnan Hospital of Wuhan University, Wuhan, China

[2]Department of Biological Repositories, Zhongnan Hospital of Wuhan University, Wuhan, China

[3]Department of Urology, Zunyi Medical University, Zunyi, China

[4]Human Genetics Resource Preservation Center of Hubei Province, Wuhan, China

Correspondence
Sheng Li, Department of Biological Repositories, Zhongnan Hospital of Wuhan University, 169 Donghu Road, 430071 Wuhan, China.
Email: lisheng-znyy@whu.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 81802541; Zhongnan Hospital of Wuhan University Science, Technology and Innovation Seed Fund, Grant/Award Number: znpy2017050; 351 Talent Project of Wuhan University (Luojia Young Scholars: SL)

Abstract

Bladder cancer (BC) is one of the most malignancies in terms of incidence and recurrence worldwide. The aim of this study is to find out novel and prognostic biomarkers for patients with BC. First, we identified 258 differentially expressed genes by using GSE19915 from Gene Expression Omnibus database. Second, a total of 33 modules were identified by constructing a coexpression network by using weighted gene coexpression network analysis and yellow module was regarded as the key module. Furthermore, by constructing protein–protein interaction networks, we preliminarily picked out 13 genes. Among them, four hub genes (CCNB1, KIF4A, TPX2, and TRIP13) were eventually identified by using five different methods (survival analysis, one-way analysis of variance, the Spearman correlation analysis, receiver operating characteristic curve, and expression value comparison), which were significantly correlated with the prognosis of BC. The validation of transcriptional and translational levels made sense (based on Oncomine and The Human Protein Atlas database). Moreover, functional enrichment analysis suggested that all the hub genes played crucial roles in chromosome segregation, sister chromatid segregation, nuclear chromosome segregation, mitotic nuclear division, nuclear division, and organelle fission during cell mitosis. In addition, three of the hub genes (KIF4A, TPX2, and TRIP13) might be potential targets of cancer drugs according to the results of the genetical alteration. In conclusion, this study indicates that four hub genes have great predictive value for the prognosis of BC, and may contribute to the exploration of the further and more in-depth research of BC.

KEYWORDS
bladder cancer, coexpression, histologic grade, hug genes, WGCNA

1 | INTRODUCTION

Bladder cancer (BC) is the most common malignancy of the urinary system (Ebrahimi et al., 2019). There are 437,422 new cases annually in the world according to a recent research (Ebrahimi et al., 2019). Although in the country with health facilities well developed, BC causes a lot of troubles (Siegel, Miller, & Jemal, 2018). For example, in the USA, BC is the 13th most common cause of deaths among all cancers (Siegel et al., 2018). For diagnosis, cystoscopy and biopsy are still the gold standard (Emerson & Cheng, 2005). Unfortunately, the average age for BC at diagnosis is 65 years, which means that most BCs are diagnosed at an advanced stage (Shadpour, Emami, & Haghdani, 2016). Many patients often lost the optimal chance for most effective treatment. As for the prognosis of BC, the situation was not optimistic. Five-year survival rate for BC patients was reported as low as 50–70% (Hussein et al., 2016). Moreover, it has

been shown that about 30–70% of BC tumors will relapse (Hussein et al., 2016). Because of the poor prognosis and difficulties in early diagnosis, we aim to develop novel and specific prognostic markers for patients with BC.

Weighted gene coexpression network analysis (WGCNA) has been widely used in studying biological networks and analyzing potential gene modules associated with large gene expression data (Langfelder & Horvath, 2008). It can be applied to most high-dimensional data sets especially for genomic applications (Ivliev, Pa, & Sergeeva, 2010). By using WGCNA, one can classify genes into different gene modules based on their connectivity with others. So that we can further explore the relationships between gene modules and clinical features to find out candidate biomarkers for cancers (Clarke et al., 2013). With the aim of identifying biomarkers or therapeutic targets for BC, a coexpression network was constructed in this study by using TCGA-BLCA data. Key module related to histologic grade was selected. We also detected differentially expressed genes (DEGs) on the basis of GSE19195. After finishing all the steps, 13 genes were picked out preliminarily. We used five different methods to identify hub genes among them. Finally, four hub genes were picked out and we validated them using two data sets (GSE13507 and GSE31684) and six databases (Oncomine, HPA, GEO, TCGA, GEPIA, and STRING). Based on CBioPortal, we also explored the hub genes' correlation with other famous genes and drugs aiming at finding new targets for anticancer drugs.

2 | MATERIALS AND METHODS

2.1 | Identification of BC microarray studies

Microarray data set of BC (TCGA-BLCA), which was downloaded from The Cancer Genome Atlas (TCGA) database (https://genome-cancer.ucsc.edu/), was used to construct coexpression network. Data set GSE19915 (Lindgren et al., 2010) was downloaded from Gene Expression Omnibus (GEO) database to screen DEGs (http://www.ncbi.nlm.nih.gov/geo/). Also, data GSE13507 (Kim et al., 2010; Lee et al., 2010) and GSE31684 (Riester et al., 2012, 2014) were used to further verify our results. Data set GSE19915 was performed on three independent platforms, GPL3883 (Swegene Human 27K RAP UniGene188 array), GPL4723 (SWEGENE_BAC_32K_Full), and GPL5186 (SWEGENE H_v3.0.1 35K). When performed on GPL3883, this data set contained eight normal urinary bladder samples and 76 urinary bladder tumor samples; when performed on GPL5186, this data set contained seven normal urinary bladder samples and 91 urinary bladder tumor samples; when performed on GPL4723, this data set contained 103 urinary bladder tumor samples. So GSE19915 performed on GPL3883 and GPL5186 was used to screen DEGs. Data set GSE13507 performed on Illumina human-6 v2.0 expression beadchip included 10 normal bladder mucosae and 165 primary BC samples. GSE31684, performed on GPL570, included 93 BC samples.

2.2 | Data preprocessing

Figure S1a showed the research process of this study. Four hundred and eight bladder urothelial carcinoma samples from TCGA-BLCA were included for WGCNA analysis. We calculated the variances of probes across all samples and top 8,000 probes with highest variances were selected for the WGCNA analysis.

2.3 | Screening of DEGs

By means of GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/; Davis & Meltzer, 2007), we selected DEGs between BC and normal bladder tissues. We considered genes as DEGs when they met the following standards: Adjust $p < 0.05$, and $|\log_2 FC| \geq 1.5$. The common DEGs in GSE19915 performed on GPL3883 and GSE19915 performed on GPL5186 were screened for subsequent analysis.

2.4 | Construction of a coexpression network

At the beginning, expression profile of the top 8,000 genes was checked if they were suitable for the WGCNA by using gsg (goodSamplesGenes) method. We further tested the expression profile using sample network methods to distinguish outlying samples. Samples with Z.Ku < −2.5 were regarded as outliers. All the outliers were removed from the data. After that, a coexpression network was constructed by using R package "WGCNA." Based on the scale-free topology criterion, β (soft threshold power β) was selected (Zhang & Horvath, 2005). Considering with the function of topological overlap matrix (TOM; Li & Horvath, 2009), we transformed adjacency into TOM. Moreover, three different branch cutting methods (manual [interactive] branch cutting approach, automatic single block analysis, and two-block analysis) were conducted for classifying genes into gene modules. A relatively large minimum module size of minClusterSize = 30, and a medium sensitivity (deepSplit = 2) was chosen to branch splitting. The dissimilarity of module eigengenes (MEs) was calculated to set a cut line for merging some modules. Additionally, a TOM plot of all genes and a classical multidimensional scaling (MDS) plot were plotted.

2.5 | Identification of clinical key modules

In the present study, we used two methods to identify key module associated with the trait (histologic grade) we were interested in. On the one hand, gene significance (GS) was calculated to quantify the relationship between genes and trait. On the other hand, module membership (MM) was also defined to describe the correlation between ME and gene expression profile. Moreover, module significance (MS, the average GS of all the genes in a module) was defined. Eventually, the module, which was most positively correlated with histologic grade, was regarded as the key module.

2.6 | Construction of protein–protein interaction (PPI) networks

By means of the Search Tool for the Retrieval of Interacting Genes (STRING; Szklarczyk et al., 2015), we constructed the PPI networks of common DEGs and genes in hub module. Parameters setting: network scoring: degree cutoff = 2; cluster finding: node score cutoff = 0.2, k-core = 2, and maximum depth = 100. In this study, we calculated the degree of genes by network analyzer (a tool in Cytoscape software (https://cytoscape.org/). Genes with degree greater than or equal to 10 were considered to be hub genes in the PPI network.

2.7 | Identification of hub genes

In the present study, a key module was chosen. Hub genes in coexpression network were identified under the threshold of |MM| > 0.50 and |GS| > 0.20. The common hub genes in coexpression network, PPI network of key module, and PPI network of DEGs were picked out for follow-up analysis. Then we used five different methods to identify hub genes among these genes using two GEO data sets (GSE13507 and GSE31684). Survival analysis was performed by R package "survival" (Therneau, 2015) using GSE13507, and we split 165 BCs into two groups based on genes expression (high group, $n = 82$; low group, $n = 83$). This package also generated Kaplan–Meier survival curve. The one-way analysis of variance (ANOVA) test and the Spearman correlation analysis were performed using GSE13507 and GSE31684. Both of the two analyses were performed using SPSS (IBM, Armonk, NY, Version 21.0). Meanwhile, by means of R package "plotROC" (Sachs, 2017), receiver operating characteristic curve (ROC) analysis was performed. In GSE13507, we calculated the area under curve (AUC) to distinguish BC samples from normal tissues. In GSE31684, we used AUC to differentiate BC of high grade from BC of low grade. After that, we compared the genes expression levels between BCs and normal bladder tissues using GSE13507 and TCGA-BLCA data. The boxplots were drawn using R package "ggstatsplot" (Patil, & Powell, 2018) and gene expression profiling interactive analysis (GEPIA; Tang et al., 2017). Genes satisfied the conditions ($p < 0.05$ in all analyses and AUC ≥ 0.80) were considered to be hub genes in the study. An upset plot was also performed using R package "UpSetR" (Conway, Lex, & Gehlenborg, 2017) to overlap genes in these five analyses. Moreover, the Pearson correlation between hub genes and marker of proliferation Ki-67 (MKi67) were performed based on the TCGA-BLCA data. Venn diagram was performed by online tool Venn-diagrams (http://bioinformatics.psb.ugent.be/beg/tools/venn-diagrams).

2.8 | Validation of hub genes

T stage (Ta, T1, T2, T3, and T4) boxplots and tumor grade (low and high) boxplots were performed using "ggstatsplot." In addition, we validated the messenger RNA (mRNA)-level and translational-level expressions of the hub genes based on the Oncomine (http://www.oncomine.org/; Rhodes et al., 2004) and The Human Protein Atlas database (https://www.proteinatlas.org/; Uhlén et al., 2015).

2.9 | Genetical alteration of hub genes

Visualization and analysis of cancer genomic data sets can be realized by using CBio Cancer Genomics Portal (http://www.cbioportal.org/; Cerami et al., 2012; Gao et al., 2013). In the present study, CBioPortal was used to explore the genetic alterations of the hub genes and the relationships between genes and drugs.

2.10 | Investigation of the associations between the clinical features of patients with BC and the hub gene expression levels

Based on GSE13507, we evaluated the median of hub genes expression levels. After that, 165 BCs from GSE13507 were divided into two groups. The associations between the clinicopathological features of BC patients and the hub gene expression levels were analyzed by the χ^2 test through SPSS (IBM, Version 21.0).

2.11 | Gene ontology (GO) and kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis

We performed GO (Ashburner et al., 2000) and KEGG pathway enrichment analysis (Kanehisa & Goto, 2000) for DEGs and genes in key module by using R package "clusterProfiler" (Yu, Wang, Han, & He, 2012). In this study, we only showed the results of biological process (BP) and KEGG. Gene sets at $p < 0.05$ were considered to be significantly enriched.

2.12 | Gene set enrichment analysis (GSEA) and guilt of association of hub genes

With the same method we mentioned before, the 165 BC samples from GSE13507 were classified into two groups. GSEA (Subramanian et al., 2005) was conducted between the two groups. Signaling pathways reached the standards (nominal $p < 0.05$; |ES| > 0.6; gene size ≥ 100; FDR < 25%) were considered significant in the present study. Also, we performed batch Spearman correlation analysis of hub genes using the TCGA-BLCA data. Correlation coefficient absolute values were calculated, and we selected the top 500 genes of each hub gene using R packages "dplyr" and "tidyr." Moreover, functional enrichment analysis was performed by using "clusterProfiler." According to the results, we predicted the lurking functions of hub genes. We called this method "guilt of association." Then, we compared the results between GSEA and guilt of association.

3 | RESULTS

3.1 | Screening of DEGs

By using GEO2R, 1,322 DEGs (269 upregulated and 1,053 downregulated) were identified in GSE19915 performed on GPL3883 (Figure 1a), and 592 DEGs (127 upregulated and 465 downregulated) were identified in GSE19915 performed on GPL5186 (Figure 1b). The common DEGs, including 258 genes (33 upregulated and 225 downregulated), were used to construct a PPI network (Figure 1c,d). The real hub genes were also labeled on the volcano plots.

3.2 | Construction of a coexpression network and identification of key modules

Two different methods were used to identify outlier samples. After that, a total of 28 samples were removed from the subsequent analysis (Figure S1b,c). A total of 380 TCGA-BLCA samples and their clinical information were included for the WGCNA analysis. In the present study, $\beta = 5$ (scale-free $R^2 = 0.84$) was set as the soft thresholding for further adjacency calculation (Figure S2). After classifying genes into gene modules and merging modules (pairwise correlation of modules > 0.75), 33 modules were generated in total (Figure S3a). Genes in gray module were removed from the further processing. Furthermore, the ME of the yellow module ($p = 6 \times 10^{-8}$; $R^2 = 0.27$) revealed a high correlation with histologic grade (low and high) compared with other modules (Figure 2a). The MS of yellow module also made sense (Figure S3b). Yellow module, the most positively associated module with histologic grade of BC, was regarded as the key module. The relationship between MM and GS in yellow module is shown in Figure 2b. The network heatmap was shown in Figure 2c. A classical MDS plot was created, and each dot (gene) is colored by the module assignment (Figure 2d).

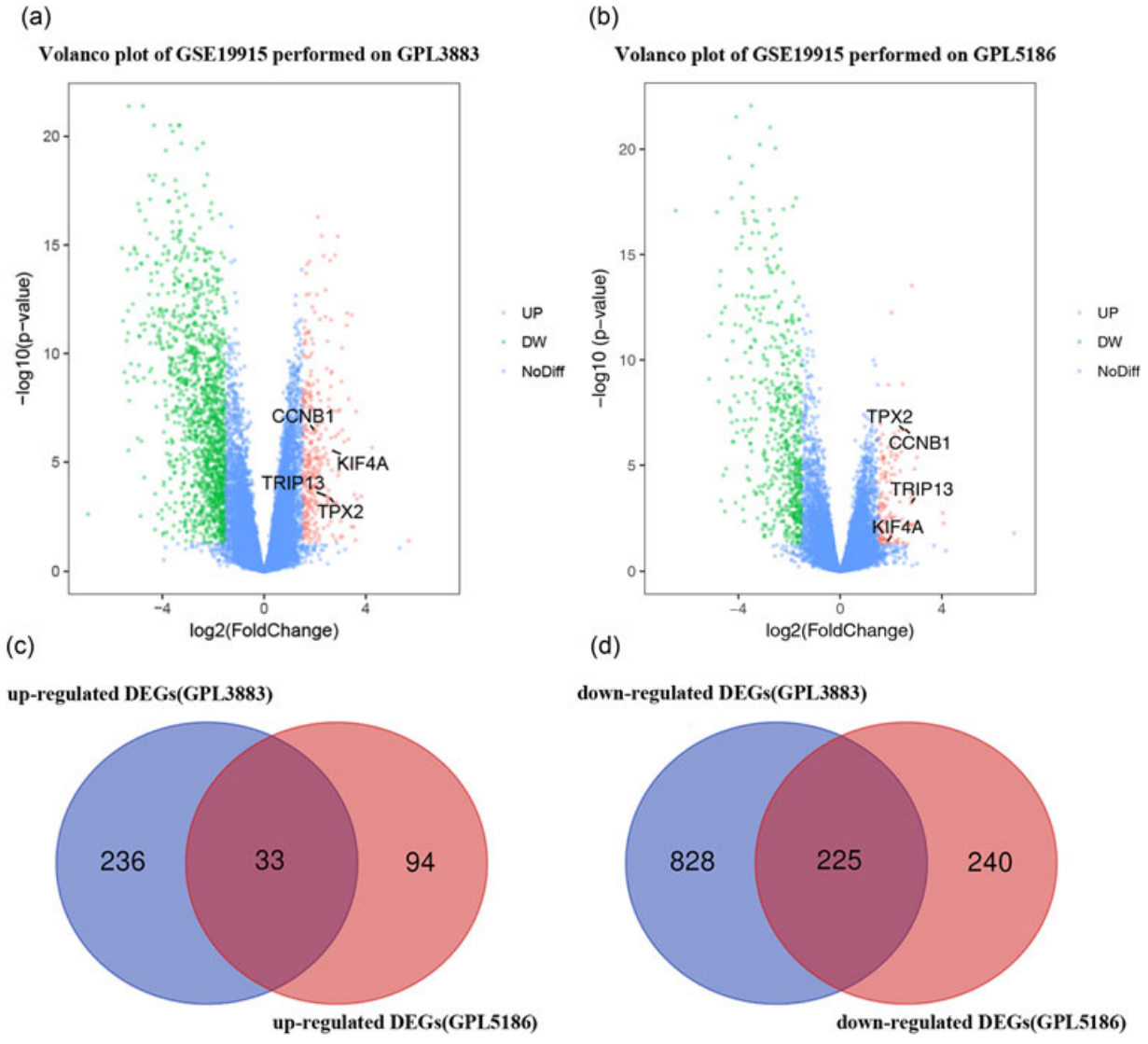

FIGURE 1 Identification of common DEGs. (a) Volcano plot visualizing DEGs in GSE19915 performed on GPL3883. (b) Volcano plot visualizing DEGs in GSE19915 performed on GPL5186. (c) Identification of common upregulated DEGs. (d) Identification of common upregulated DEGs. DEG: differentially expressed gene [Color figure can be viewed at wileyonlinelibrary.com]

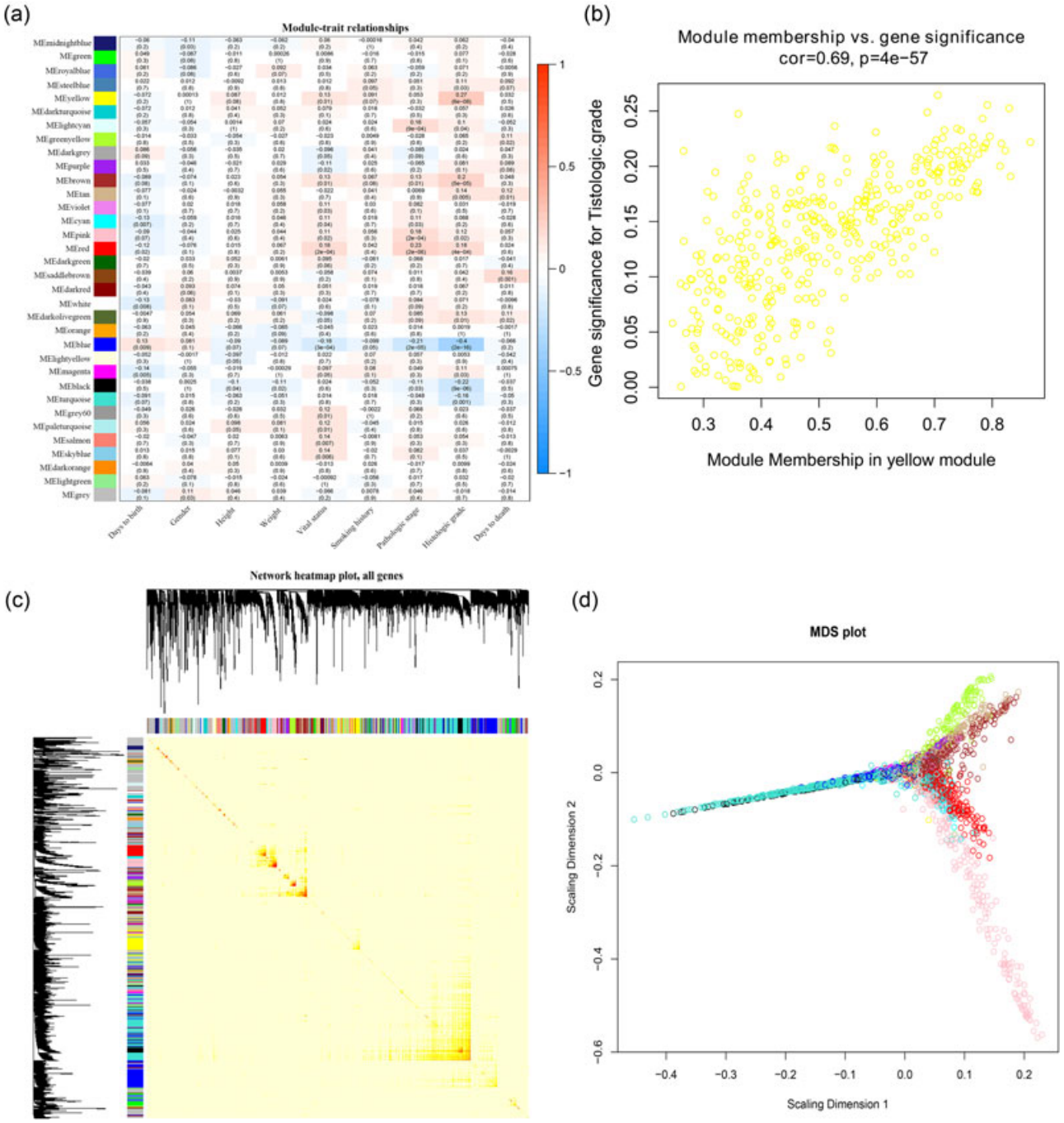

FIGURE 2 Identification of modules associated with the clinical traits of BC and construction of a classical MDS plot. (a) Heatmap of the correlation between MEs and different clinical information of BC (days to birth, gender, height, weight, vital status, smoking history, pathologic stage, histologic grade, and days to death). (b) Scatter plot of module eigengenes in the yellow module. (c) Interaction relationship analysis of coexpression genes. Different colors of horizontal axis and vertical axis represent different modules. The brightness of yellow in the middle represents the degree of connectivity of different modules. There was no significant difference in interactions among different modules, indicating a high-scale independence degree among these modules. (d) Classical MDS plot whose input is the TOM dissimilarity. Each dot (gene) is colored by the module assignment. BC: bladder cancer; MDS: multidimensional scaling; ME: module eigengene; TOM: topological overlap matrix [Color figure can be viewed at wileyonlinelibrary.com]

3.3 | Identification of hub genes

First, we constructed a PPI network of the common 258 genes. Fifty-eight genes that reached the cutoff criterion (degree ≥ 10) were regarded as hub genes in DEGs (Figure 3a). Also, we constructed a PPI network of genes in the yellow module, with the same cutoff criterion, and 279 genes were identified as hub genes in the yellow module PPI network (Figure 3b). The whole PPI networks of genes in

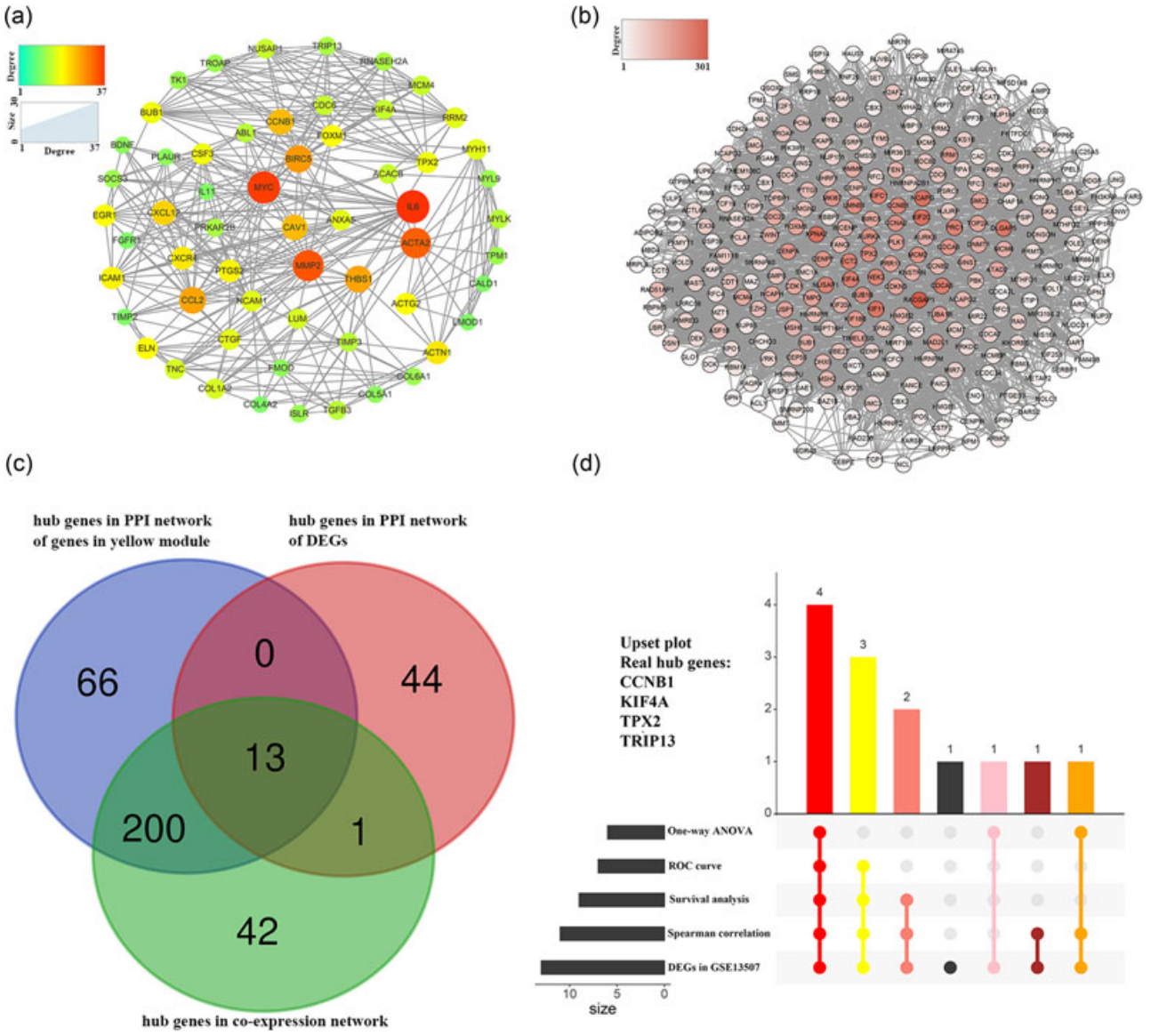

FIGURE 3 Construction of PPI networks, Venn plot of candidate hub genes, and Upset plot of hub genes. (a) PPI network of hub genes in DEGs. (b) PPI network of hub genes in the yellow module. (c) Common hub genes in the coexpression network and PPI network. (d) Common genes with significant *p* value in survival analysis, ROC curve analysis, one-way ANOVA, the Spearman correlation analysis, and DEG analysis (an Upset plot). ANOVA: analysis of variance; DEG: differentially expressed gene; PPI: protein–protein interaction; ROC: receiver operating characteristic curve [Color figure can be viewed at wileyonlinelibrary.com]

yellow module and DEGs were shown in Figures S4 and S5. Considering with module connectivity, 256 hub genes in coexpression network were identified. Based on the three methods, a total of 13 genes were common in them, which were screened out preliminarily for further validation (Figure 3c). According to the result of the one-way ANOVA (Table S1), seven genes were picked out in both data sets. The results of the Spearman correlation analysis suggested that 10 genes showed significant *p* values in both GSE31684 and GSE13507 (Table S2). According to the results of survival analysis, nine genes made sense in both overall survival and cancer-specific survival (Table S3). By using GSE31684 and GSE13507, ROC curve was plotted (Table S4). Seven genes which reached the standard of AUC ≥ 0.80 were finally screened out. Moreover, the gene expressions of BC and normal bladder tissues were compared based on GSE13507 and GEPIA, and 13 of which were differentially expressed (Figure S8a,b only showed the results of real hub genes). Four genes were eventually identified (CCNB1, KIF4A, TPX2, and TRIP13; Figure 3d) because they showed significant *p* values in all the analyses. Thus, we considered them as hub genes associated with progression and prognosis of BC in our study. Figures 4–5 showed the results of survival analysis. Recently, MKi67 was regarded as a biomarker for the growth of multiple tumors (Felix et al., 2015). Thus, we also performed Pearson's correlation between MKi67 and hub genes (Figure S6).

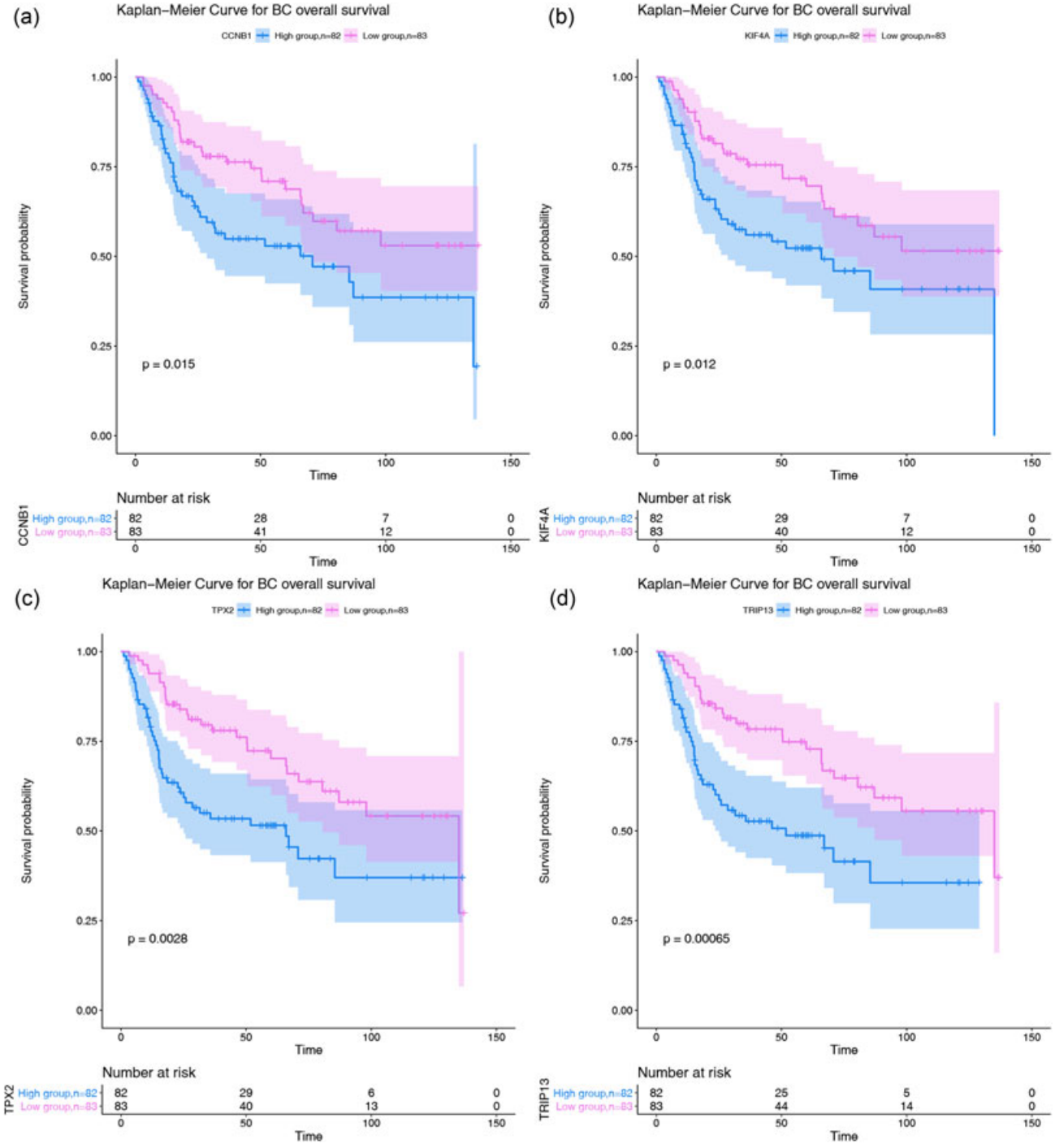

FIGURE 4 Survival analysis of the association between the expression levels of hub genes and overall survival time in BC (based on GSE13507). (a) CCNB1, (b) KIF4A, (c) TPX2, and (d) TRIP13. BC: bladder cancer [Color figure can be viewed at wileyonlinelibrary.com]

3.4 | Validation and genetical alteration of hub genes

Based on GSE13507 and GSE31684, the grade and stage boxplots of hub genes were shown in Figure S7. In addition, mRNA expression levels were all significantly higher in BC tissues than those in normal bladder tissues (Figure S8c), which was suggested by Oncomine database. Figure S9 showed the translational-level expression of hub genes. These results made the hub genes we screened out reliable. As for genetical alteration, four hub genes altered in 159 (39%) of 412 patients (Figure 6b). As shown in Figure 6a, TRIP13 altered most (23%) and the main type was mRNA upregulation. A network containing 58 genes (three real hub genes and 55 most variant genes) is shown in Figure 6c. TP53 was significantly vital in the network. As for the relationship between anticancer drugs and hub genes, we found CCNB1 was the target of cancer drugs. But there was no drug

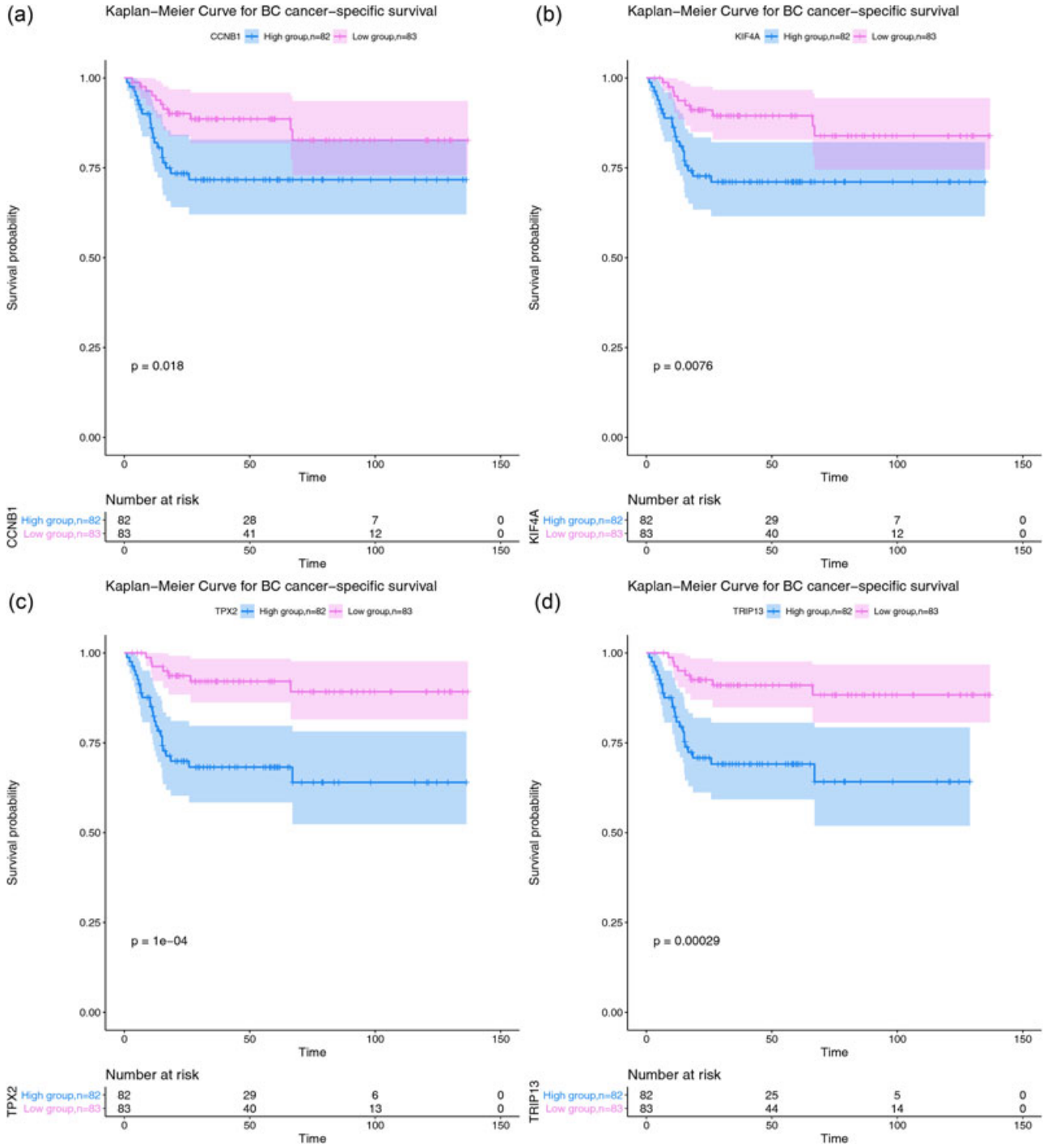

FIGURE 5 Survival analysis of the association between the expression levels of hub genes and cancer-specific survival time in BC (based on GSE13507). (a) CCNB1, (b) KIF4A, (c) TPX2, and (d) TRIP13. BC: bladder cancer [Color figure can be viewed at wileyonlinelibrary.com]

targeting to the rest three hub genes, which might be novel therapeutic targets for patients with BC.

3.5 | Associations between hub gene expression levels and clinical features

Based on GSE13507, we collected the clinical information of 165 patients. We brought seven clinical factors (age, gender, progression, grade, T stage, N stage, and M stage) into this step. The results suggested that CCNB1 expression level in BC was statistically related to age, progression, grade, T stage, and N stage (Table 1); KIF4A expression level in BC was statistically related to age, gender, progression, grade, T stage, and N stage (Table 1); TPX2 expression level in BC was statistically related to age, gender, progression, grade, T stage, and N stage (Table 1); TRIP13 expression level in BC was statistically related to age, progression, grade, T stage, and

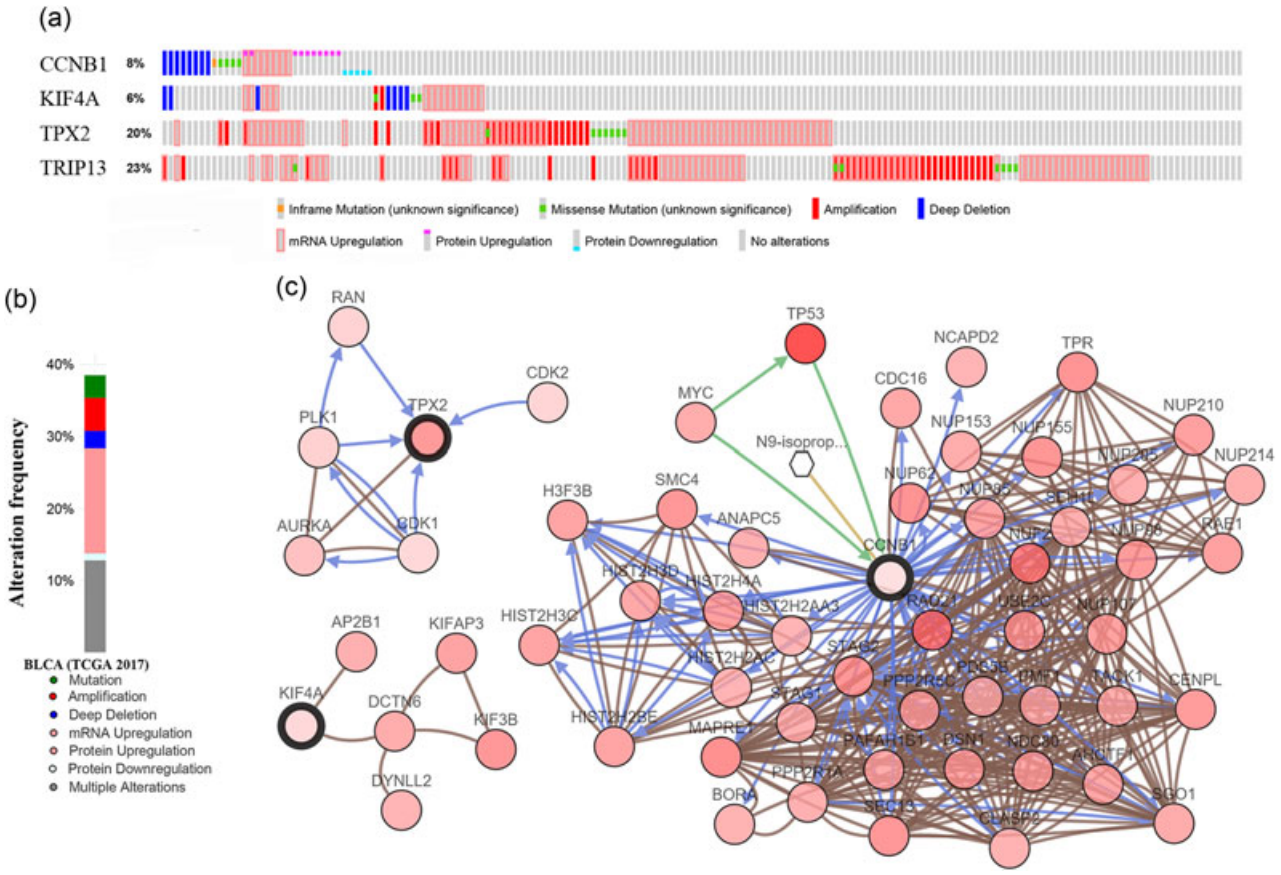

FIGURE 6 Genetic alterations associated with hub genes in TCGA-BLCA. (a) A visual summary across on a query of four hub genes showing genetic alteration of hub genes in TCGA-BLCA patients. (b) The total alteration frequency of four hub genes in TCGA-BLCA is illustrated. (c) The network contains 58 nodes, including our four query genes and the 55 most frequently altered neighbor genes (only three out of four were correlated with the 55 genes). Relationship between hub genes and tumor drugs is also illustrated. mRNA: messenger RNA [Color figure can be viewed at wileyonlinelibrary.com]

N stage (Table 1). All *p* values were available in Table 1. All in all, high expression levels of the four genes were relevant to the tumor progression of BC probably.

3.6 | Functional and pathway enrichment analysis

GO analysis results showed that DEGs were enriched in various BPs, the top 10 terms were extracellular matrix organization, extracellular structure organization, muscle system process, muscle contraction, cell-substrate adhesion, regulation of muscle system process, regulation of peptidase activity, regulation of cell-substrate adhesion, regulation of actin filament-based process, and regulation of muscle contraction (Figure 7a). And genes in the yellow module were enriched in chromosome segregation, sister chromatid segregation, nuclear chromosome segregation, mitotic nuclear division, mitotic sister chromatid segregation, nuclear division, organelle fission, DNA replication, sister chromatid cohesion, and DNA conformation change (Figure 7c).

As for the pathways, the results of KEGG enrichment analysis showed that DEGs were enriched in PI3K-Akt signaling pathway, focal adhesion, vascular smooth muscle contraction, cGMP-PKG signaling pathway, proteoglycans in cancer, AGE-RAGE signaling pathway in diabetic complications, human cytomegalovirus infection, ECM–receptor interaction, malaria, and amebiasis (Figure 7b). As for genes in the yellow module, they enriched in cell cycle, mismatch repair, RNA transport, DNA replication, spliceosome, oocyte meiosis, nucleotide excision repair, base excision repair, progesterone-mediated oocyte maturation, and one carbon pool by folate (Figure 7d).

3.7 | GSEA and guilt of association

With the cutoff criteria we set before, a total of eight CCNB1-related signaling pathways were enriched, and the top three pathways were "spermatogenesis," "mTORC1 signaling," and "G2M checkpoint" (Figure 8a). Totally nine KIF4A-related signaling pathways were enriched, and the top three pathways were "spermatogenesis," "unfolded protein response," and "G2M checkpoint" (Figure 8). Totally nine TPX2-related signaling pathways were enriched, and the top three pathways were "spermatogenesis," "G2M checkpoint," and "unfolded protein response" (Figure 8c). Totally nine TRIP13-related signaling pathways were enriched, and

TABLE 1 Associations between real hub gene expression and clinicopathological factors of patients with BC (based on GSE13507)

	CCNB1 expression				KIF4A expression				TPX2 expression				TRIP13 expression			
Characteristics	Low (%); *n* = 83	High (%); *n* = 82	χ^2	*p* Value	Low (%); *n* = 83	High (%); *n* = 82	χ^2	*p* Value	Low (%); *n* = 83	High (%); *n* = 82	χ^2	*p* Value	Low (%); *n* = 83	High (%); *n* = 82	χ^2	*p* Value
Age																
≤65	45 (54.2)	29 (35.4)	5.926	0.015	45 (54.2)	29 (35.4)	5.926	0.015	48 (57.8)	26 (31.7)	11.381	0.001	45 (54.2)	29 (35.4)	5.926	0.015
>65	38 (45.8)	53 (64.6)			38 (45.8)	53 (64.6)			35 (42.2)	56 (68.3)			38 (45.8)	53 (64.6)		
Sex																
Male	70 (84.3)	65 (79.3)	0.712	0.399	73 (88.0)	62 (77.5)	4.224	0.04	73 (88.0)	62 (77.5)	4.224	0.04	72 (86.7)	63 (76.8)	2.727	0.099
Female	13 (15.7)	17 (20.7)			10 (12.0)	20 (22.5)			10 (12.0)	20 (22.5)			11 (13.3)	19 (23.2)		
Progression																
Yes	10 (12.0)	21 (25.6)	4.972	0.026	8 (9.6)	23 (28.0)	9.163	0.002	8 (9.6)	23 (28.0)	9.163	0.002	8 (9.6)	23 (28.0)	9.163	0.002
No	73 (88.0)	61 (74.4)			75 (90.4)	59 (72.0)			75 (90.4)	59 (72.0)			75 (90.4)	59 (72.0)		
Grade																
High	15 (18.1)	45 (54.9)	24.147	<0.001	13 (15.7)	47 (57.3)	30.928	<0.001	11 (13.3)	49 (59.8)	38.548	<0.001	10 (12.0)	50 (61.0)	42.672	<0.001
Low	68 (81.9)	37 (45.1)			70 (84.3)	35 (42.7)			72 (86.7)	33 (40.2)			73 (88.0)	32 (39.0)		
T stage																
Ta–T1	62 (74.7)	42 (51.2)	9.758	0.002	61 (73.5)	43 (52.4)	8.341	0.004	63 (75.9)	41 (50.0)	11.878	0.001	65 (78.3)	39 (47.6)	16.740	<0.001
T2–T4	21 (25.3)	40 (48.8)			22 (26.5)	39 (47.6)			20 (24.1)	41 (50.0)			18 (21.7)	43 (52.4)		
N stage																
N0	78 (94.0)	71 (86.6)	3.596	0.058	79 (95.2)	70 (85.4)	5.944	0.015	80 (96.4)	69 (84.1)	6.189	0.013	79 (95.2)	70 (85.4)	5.944	0.015
N1–N3	4 (6.0)	11 (13.4)			3 (3.6)	12 (14.6)			3 (3.6)	12 (15.9)			3 (3.6)	12 (14.6)		
M stage																
M0	79 (95.2)	79 (96.3)	0.137	0.711	80 (96.4)	78 (95.1)	<0.001	0.987	80 (96.4)	78 (95.1)	<0.001	0.987	80 (96.4)	78 (95.1)	<0.001	0.987
M1	4 (4.8)	3 (3.7)			3 (3.6)	4 (4.9)			3 (3.6)	4 (4.9)			3 (3.6)	4 (4.9)		

Note. BC: bladder cancer.

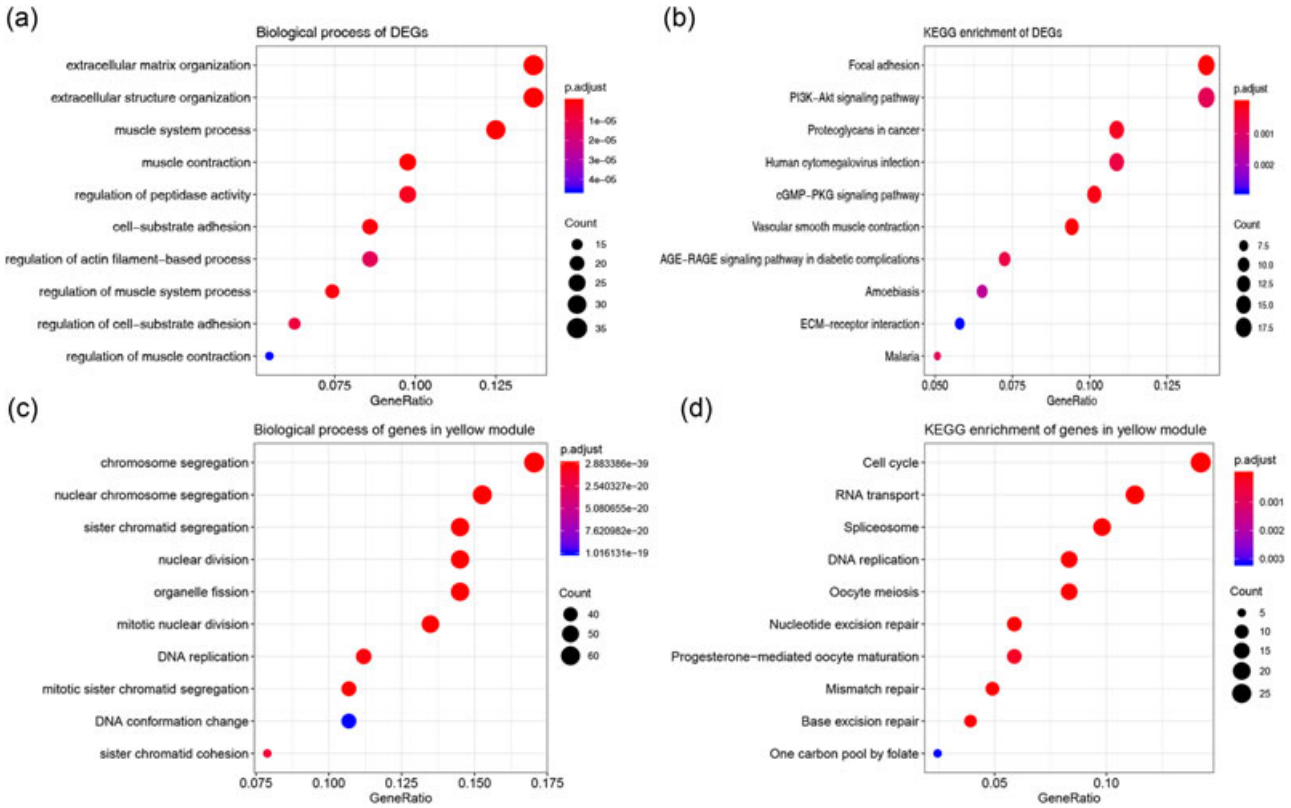

FIGURE 7 Bioinformatics analysis of DEGs and genes in yellow module. (a) GO analysis of DEGs. (b) KEGG pathway enrichment of DEGs. (c) GO analysis of genes in yellow module. (d) KEGG pathway enrichment of genes in yellow module. DEG: differentially expressed gene; ECM: extracellular matrix; GO: gene ontology; KEGG: kyoto encyclopedia of genes and genomes [Color figure can be viewed at wileyonlinelibrary.com]

the top three pathways were "spermatogenesis," "unfolded protein response," and "G2M checkpoint" (Figure 8d). Moreover, "spermatogenesis" and "G2M checkpoint" were the common signaling pathways in all of real hub gene-related pathways. According to the results of guilt of association, we found that all real hub genes played important roles in many events during cell mitosis, such as chromosome segregation, sister chromatid segregation, nuclear chromosome segregation, mitotic nuclear division, nuclear division, and organelle fission (Figure S10).

4 | DISCUSSION

BC, the commonest urinary malignancy, can occur at any age. BC is particularly liable to relapse after combined therapy (Rye, Nustad, & Stigbrand, 2003). What's worse, BC carries a very poor prognosis when developing into infiltrating BC (Rye et al., 2003). Although radical cystoprostatectomy is the most efficient therapy for patients with infiltrating BC nowadays, it will cause many bad consequences and weaken the quality of life of its victims (Stein et al., 2001). Histologic grade of malignant tumors reflects the internal characteristics of tumors, which is of great value for evaluation of the degree of tumor differentiation and prediction of prognosis. Given that most BCs are diagnosed with advanced stages, the prognosis of patients with BC remains extremely poor. Therefore, we expect to find some sensitive and novel biomarkers that can predict tumor grade and prognosis of BC patients. And we also determine to screen out a few promising targets of new cancer drugs.

In the present study, we screened out 258 DEGs by using GSE19915. Moreover, we constructed a PPI network of DEGs to find out hub genes in DEGs. By means of WGCNA analysis, we constructed a coexpression network by using TCGA-BLCA data. We brought nine clinical features (days to birth, gender, height, weight, vital status, smoking history, pathologic stage, histologic grade, and days to death) into this study. We also attempted to find a key module which was most relevant to the histologic grade of BC. Thirty-three modules were identified. Among them, the yellow module was most positively correlated with the histologic grade. Four hundered and seventy-two genes made up this module, 256 of which were regarded as hub genes in the coexpression network. Relying on the STRING, 279 genes with degree ≥ 10 were considered as hub genes in PPI network of yellow module. With the same cutoff criterion, 58 genes were screened out in PPI network of DEGs. Preliminarily, 13 genes which were common in these networks were identified. To find out hub genes among the 13 genes, five different analyses were performed. Four hub genes (CCNB1, KIF4A, TPX2, and TRIP13) related to tumor grade and poor prognosis of BC were finally identified. These hub genes were also highly expressed in tumors comparing with normal tissues using GSE13507 and TCGA-BLCA data.

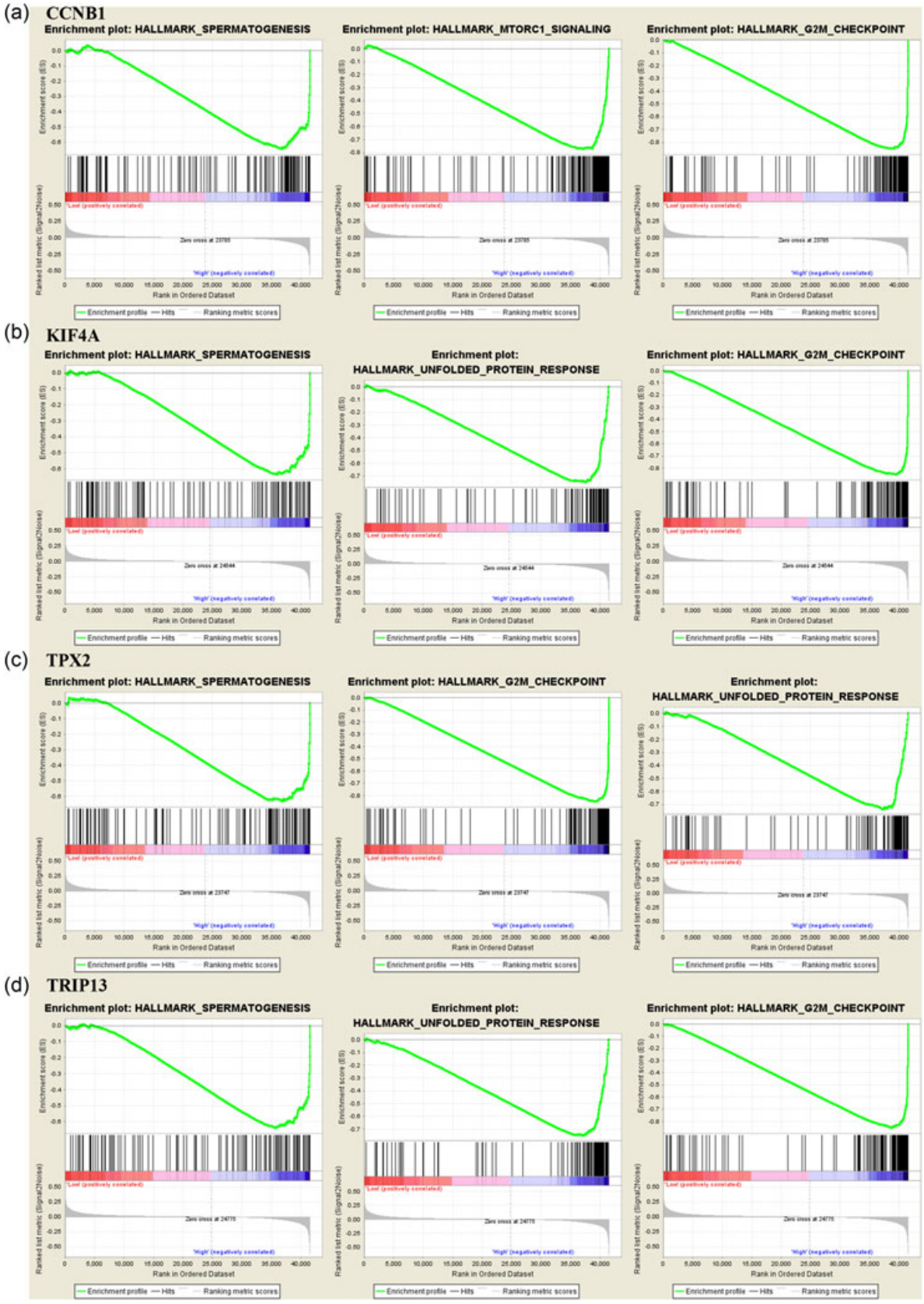

FIGURE 8 Gene set enrichment analysis. The top three functional genes sets enriched in bladder cancer with CCNB1 (a), KIF4A (b), TPX2 (c), and TRIP13 (d) [Color figure can be viewed at wileyonlinelibrary.com]

The validations of hub genes in mRNA level and translational level were carried out using two databases. After that, we used CBioPortal to explore the relationship between the hub genes and drugs aiming at finding new targets for anticancer drugs. We found that CCNB1 was already the target for anticancer drugs, which meant the remaining genes (KIF4A, TPX2, and TRIP13) probably became potent drug targets. To improve the reliability of results, based on GSE13507, we brought seven factors into the analysis of clinical characteristics among BC patients. The associations between the clinicopathological features of BC patients and the hub genes expression levels were analyzed by using the same data set GSE13507. The results suggested that expression levels of all the four hub genes in BC tissues were associated with age, progression, grade, T stage, and N stage. Although we did not perform this analysis in TCGA-BLCA data (the data used for WGCNA analysis), this might be a statistical limitation of this part of the study. It might be better for us to do this part in TCGA-BLCA data to keep the consistency of patient cohort. But make our results more universal and representative, we chose to do this part by using GSE13507. The gene expression data and the clinical information were all got from GSE13507. GSE13507 was also a representational data set with many BC samples, which made these results reliable. When talking about T stage, recent research showed that BC stage was an independent prognostic factor of muscle-invasive BC (MIBC) patients. MIBC (T stage ≥ T2) is related to poor clinical prognosis. This coincides with outcomes in this study.

According to the GO BP analysis, DEGs in GSE19915 were majorly participated in extracellular matrix organization, extracellular structure organization, muscle system process, and so on; meanwhile, genes in yellow module were enriched in chromosome segregation, sister chromatid segregation, nuclear chromosome segregation, and so on. As for KEGG analysis, DEGs in GSE19915 were enriched in focal adhesion, vascular smooth muscle contraction, cGMP-PKG signaling pathway, and so on; meanwhile, genes in yellow module were obviously relevant to cell cycle, DNA replication, mismatch repair, and so on. To further understand the signaling pathways regulated by real hub genes, we performed GSEA analysis by the aid of GSE13507. All the hub genes were relevant to spermatogenesis and G2M checkpoint. According to the results of guilt of association, we firmly believed that hub genes played important roles in many events during cell mitosis, such as chromosome segregation, sister chromatid segregation, nuclear chromosome segregation, mitotic nuclear division, nuclear division, and organelle fission.

For a deeper and better understanding of the four hub genes, a literature review was carried out. G2/mitogen-specific cyclin B1, encoded by CCNB1 in human, was a regulatory protein involved in mitosis (Sartor, Ehlert, Grzeschik, Mu Ller, & Adolph, 1992). CCNB1 could form a complex called cyclin B1/cdc2 (maturation-promoting factor [MPF]) with p34 (cdc2; Milatovich & Francke, 1992). This complex was involved in early mitosis, such as chromosome condensation, nuclear membrane decomposition, and spindle pole assembly. Activation of cdc2 caused cells to enter the M phase from G2 phase (Kimura, Hirano, Kobayashi, & Hirano, 1998). With the disintegration of MPF complex, the expression of cyclin B1 decreased rapidly. Cdc2 activity was inhibited, and cells further entered the anaphase of mitosis. Kinesin family member 4, a kind of newly discovered kinesins, included KIF4A, KIF4B, KIF4C, and KIF4D (Ha et al., 2000). Kumiko et al. (2012) showed that KIF4A was a multifunctional dynein molecule, which could interact with other dynein molecules. This dynein molecule participated in chromosome segregation, chromosome condensation (Mazumdar, Sundareshan, & Misteli, 2004), spindle formation (Wandke et al., 2012), and DNA damage response (Wu et al., 2008) during cell mitosis. TPX2, a microtubule-associated protein, played a crucial role in cell mitosis, especially in spindle formation (Neumayer, Belzil, Gruss, & Nguyen, 2014). Microtubule was the main component of spindle (Wei, Zhang, Wynn, & Seemann, 2015). TPX2 played an important role in spindle assembly and maintaining the integrity of spindle (Cocchiola et al., 2014; Gruss et al., 2002). As for TRIP13, it has already been identified as oncogene (Wang et al., 2014). Overexpression of TRIP13 led to many human cancers, such as head and neck squamous cell carcinoma (Banerjee et al., 2016) and prostate cancer (Larkin et al., 2012). During mitosis, TRIP13 could affect the function of spindle assembly checkpoint (Musacchio & Salmon, 2007; Wang et al., 2014). Combining this with our study, we believed that all the hub genes played crucial roles in many events during cell mitosis, especially in chromosome segregation and spindle formation. Thus, we forecasted the four hub genes played important roles in the cause and development of BC by affecting mitosis boldly.

Some limitations of our study also should be discussed. First, the correlations between the four hub genes (CCNB1, KIF4A, TPX2, and TRIP13) and clinical features were performed on GSE13507 instead of TCGA-BLCA data (the data set used for coexpression network). This might destroy the consistency of patient cohort and we will validate the correlations in our further research. Second, although we designed this bioinformatic study well and used strict thresholds for each database mining and subsequent analysis, the major drawback in this study was the lack of in vivo and in vitro validation. Therefore, we will further use a variety of cell lines, tissues, and animal experiments to carry out systematic verification after we have made prospective design and preparation (especially when tissue testing requires informed consent of patients and medical ethics approval).

5 | CONCLUSIONS

To sum up, the present study used various bioinformatics analysis tools to identify four novel hub genes, which may serve key roles in the tumorigenesis, progression and prognosis of human BC. Meanwhile, we predicted the potential function of the grade- and prognosis-related hub genes, which participated in cell mitosis. The four hub genes might be novel biomarkers of BC and three of them might be novel potential drug targets. However, the lack of in vivo and in vitro experiments is a limitation of the present study, further

molecular biological experiments are required to confirm the present findings, and confirm the role and function of these hub genes in BC.

ACKNOWLEDGMENTS

We would like to acknowledge the TCGA and GEO database developed by the National Institutes of Health (NIH). This study was supported by the National Natural Science Foundation of China (81802541) and the Zhongnan Hospital of Wuhan University Science, Technology and Innovation Seed Fund (znpy2017050) and the 351 Talent Project of Wuhan University (Luojia Young Scholars: SL).

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

X. L. and S. L. conceived and designed the study. X. Y., Z. G., and X. L. performed the analysis procedures. X. Y., Z. G., Y. F., and Y. Z. analyzed the results. X. L., S. L., and T. L. contributed analysis tools. X. Y. and Z. G. contributed to the writing of the manuscript. All authors reviewed the manuscript.

ORCID

Xin Yan http://orcid.org/0000-0001-9253-3290
Sheng Li http://orcid.org/0000-0003-4070-7345

REFERENCES

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. The gene ontology consortium. *Nature Genetics*, *25*(1), 25–29.

Banerjee, R., Russo, N., Liu, M., Basrur, V., Bellile, E., Palanisamy, N., … D'Silva, N. J. (2016). TRIP13 promotes error-prone nonhomologous end joining and induces chemoresistance in head and neck cancer. *Nature Communications*, *5*, 4527.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., … Schultz, N. (2012). The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*, *2*(5), 401–404.

Clarke, C., Madden, S. F., Doolan, P., Aherne, S. T., Joyce, H., O'Driscoll, L., … Clynes, M. (2013). Correlating transcriptional networks to breast cancer survival: A large-scale coexpression analysis. *Carcinogenesis*, *34*(10), 2300–2308.

Cocchiola, R., Grillo, C., Altieri, F., Chichiarelli, S., Turano, C., & Eufemi, M. (2014). Upregulation of TPX2 by STAT3: Identification of a novel STAT3 binding site. *PLOS One*, *9*(11), e113096.

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics*, *33*(18), 2938–2940.

Davis, S., & Meltzer, P. S. (2007). GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*, *23*(14), 1846–1847.

Ebrahimi, H., Amini, E., Pishgar, F., Moghaddam, S. S., Nabavizadeh, B., Rostamabadi, Y., & Daneshmand, S. (2019). Global, regional, and national burden of bladder cancer, 1990–2016: Results from the Global Burden of Disease Study 2016. *Journal of Urology*. [Epub ahead of print]. https://doi.org/10.1097/JU.0000000000000025

Emerson, R. E., & Cheng, L. (2005). Immunohistochemical markers in the evaluation of tumors of the urinary bladder: A review. *Analytical and Quantitative Cytology and Histology*, *27*(6), 301–316.

Felix, B., Jens, W., Wolfgang, S., Matthias, K., Vanessa, W., Fulvia, D., … Marcus, Q. (2015). Major prognostic role of Ki67 in localized adrenocortical carcinoma after complete resection. *Journal of Clinical Endocrinology and Metabolism*, *100*(3), 841–849.

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., … Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling*, *6*(269), l1.

Gruss, O. J., Wittmann, M., Yokoyama, H., Pepperkok, R., Kufer, T., Silljé, H., … Vernos, I. (2002). Chromosome-induced microtubule assembly mediated by TPX2 is required for spindle formation in HeLa cells. *Nature Cell Biology*, *4*(11), 871–879.

Ha, M. J., Yoon, J., Moon, E., Lee, Y. M., Kim, H. J., & Kim, W. (2000). Assignment of the kinesin family member 4 genes (KIF4A and KIF4B) to human chromosome bands Xq13.1 and 5q33.1 by in situ hybridization. *Cytogenetics and Cell Genetics*, *88*(1-2), 41–42.

Hussein, A. A., Dibaj, S., Hinata, N., Field, E., O'Leary, K., Kuvshinoff, B., … Guru, K. A. (2016). Development and validation of a quality assurance score for robot-assisted radical cystectomy: A 10-year analysis. *Urology*, *97*, 124–129.

Ivliev, A. E., Pa, T. H., & Sergeeva, M. G. (2010). Coexpression network analysis identifies transcriptional modules related to proastrocytic differentiation and sprouty signaling in glioma. *Cancer Research*, *70*(24), 10060–10070.

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopaedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30.

Kim, W. J., Kim, E. J., Kim, S. K., Kim, Y. J., Ha, Y. S., Jeong, P., … Bae, S. C. (2010). Predictive value of progression-related gene classifier in primary non-muscle invasive bladder cancer. *Molecular Cancer*, *9*, 3.

Kimura, K., Hirano, M., Kobayashi, R., & Hirano, T. (1998). Phosphorylation and activation of 13S condensin by Cdc2 in vitro. *Science*, *282*(5388), 487–490.

Kumiko, S., Itaru, S., Paola, V., Hiromi, O., Giulia, V., Kelly, D. A., … Hudson, D. F. (2012). Mitotic chromosomes are compacted laterally by KIF4 and condensin and axially by topoisomerase IIα. *Journal of Cell Biology*, *199*(5), 755–770.

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 559.

Larkin, S. E. T., Holmes, S., Cree, I. A., Walker, T., Basketter, V., Bickers, B., … Aukim-Hastie, C. (2012). Identification of markers of prostate cancer progression using candidate gene expression. *British Journal of Cancer*, *106*(1), 157–165.

Lee, J. S., Leem, S. H., Lee, S. Y., Kim, S. C., Park, E. S., Kim, S. B., … Chu, I. S. (2010). Expression signature of E2F1 and its associated genes predict superficial to invasive progression of bladder tumors. *Journal of Clinical Oncology*, *28*(16), 2660–2667.

Li, A., & Horvath, S. (2009). Network module detection: Affinity search technique with the multi-node topological overlap measure. *BMC Research Notes*, *2*(1), 142.

Lindgren, D., Frigyesi, A., Gudjonsson, S., Sjödahl, G., Hallden, C., Chebil, G., … Höglund, M. (2010). Combined gene expression and genomic profiling define two intrinsic molecular subtypes of urothelial carcinoma and gene signatures for molecular grading and outcome. *Cancer Research*, *70*(9), 3463–3472.

Mazumdar, M., Sundareshan, S., & Misteli, T. (2004). Human chromokinesin KIF4A functions in chromosome condensation and segregation. *Journal of Cell Biology*, *166*(5), 613–620.

Milatovich, A., & Francke, U. (1992). Human cyclin B1 gene (CCNB1) assigned to chromosome 5 (q13-qter). *Somatic Cell and Molecular Genetics*, *18*(3), 303–307.

Musacchio, A., & Salmon, E. D. (2007). The spindle-assembly checkpoint in space and time. *Nature Reviews Molecular Cell Biology*, *8*(5), 379–393.

Neumayer, G., Belzil, C., Gruss, O. J., & Nguyen, M. D. (2014). TPX2: Of spindle assembly, DNA damage response, and cancer. *Cellular and Molecular Life Sciences*, *71*(16), 3027–3047.

Patil I, & Powell C (2018). ggstatsplot: 'ggplot2' based plots with statistical details. https://CRAN.R-project.org/package=ggstatsplot

Rhodes, D. R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., … Chinnaiyan, A. M. (2004). ONCOMINE: A cancer microarray database and integrated data-mining platform. *Neoplasia*, *6*(1), 1–6.

Riester, M., Taylor, J. M., Feifer, A., Koppie, T., Rosenberg, J. E., Downey, R. J., & Michor, F. (2012). Combination of a novel gene expression signature with a clinical nomogram improves the prediction of survival in high-risk bladder cancer. *Clinical Cancer Research*, *18*(5), 1323–1333.

Riester, M., Werner, L., Bellmunt, J., Selvarajah, S., Guancial, E. A., Weir, B. A., … Rosenberg, J. E. (2014). Integrative analysis of 1q23.3 copy-number gain in metastatic urothelial carcinoma. *Clinical Cancer Research*, *20*(7), 1873–1883.

Rye, P. D., Nustad, K., & Stigbrand, T. (2003). Tumor marker workshops. *Tumor Biology*, *24*(4), 165–171.

2017>Sachs, M. C. (2017). plotROC: A tool for plotting ROC curves. *Journal of Statistical Software*,, *79*(2), 1–19.

Sartor, H., Ehlert, F., Grzeschik, K. H., Mu Ller, R., & Adolph, S. (1992). Assignment of two human cell cycle genes, CDC25C and CCNB1, to 5q31 and 5q12, respectively. *Genomics*, *13*(3), 911–912.

Shadpour, P., Emami, M., & Haghdani, S. (2016). A Comparison of the progression and recurrence risk index in non-muscle-invasive bladder tumors detected by narrow-band imaging versus white light cystoscopy, based on the EORTC scoring system. *Nephro-Urology Monthly*, *8*(1), e33240.

Siegel, R. L., Miller, K. D., & Jemal, A. (2018). Cancer statistics, 2018. *CA*, *68*(1), 7–30.

Stein, J. P., Lieskovsky, G., Cote, R., Groshen, S., Feng, A. C., Boyd, S., … Skinner, D. G. (2001). Radical cystectomy in the treatment of invasive bladder cancer: Long-term results in 1,054 patients. *Journal of Clinical Oncology*, *19*(3), 666–675.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., … Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huertacepas, J., … von Mering, C. (2015). STRING v10: Protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, *43*(Database issue), D447–D452.

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., & Zhang, Z. (2017). GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Research*, *45*(W1), W98–W102.

Therneau, T. M. (2015). Survival: Survival analysis. *Technometrics*, *46*(1), 111–112.

Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., … Ponten, F. (2015). Tissue-based map of the human proteome. *Science*, *347*(6220), 1260419.

Wandke, C., Barisic, M., Sigl, R., Rauch, V., Wolf, F., Amaro, A. C., … Geley, S. (2012). Human chromokinesins promote chromosome congression and spindle microtubule dynamics during mitosis. *Journal of Cell Biology*, *198*(5), 847–863.

Wang, K., Sturtgillespie, B., Hittle, J. C., Macdonald, D., Chan, G. K., Yen, T. J., & Liu, S. T. (2014). Thyroid hormone receptor interacting protein 13 (TRIP13) AAA-ATPase is a novel mitotic checkpoint-silencing protein. *Journal of Biological Chemistry*, *289*(34), 23928–23937.

Wei, J. H., Zhang, Z. C., Wynn, R. M., & Seemann, J. (2015). GM130 regulates golgi-derived spindle assembly by activating TPX2 and capturing microtubules. *Cell*, *162*(2), 287–299.

Wu, G., Zhou, L., Khidr, L., Guo, X. E., Kim, W., Lee, Y. M., … Chen, P. L. (2008). A novel role of the chromokinesin Kif4A in DNA damage response. *Cell Cycle*, *7*(13), 2013–2020.

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. *Omics*, *16*(5), 284–287.

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*(1), e17.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Yan X, Guo Z-X, Liu X-P, et al. Four novel biomarkers for bladder cancer identified by weighted gene coexpression network analysis. *J Cell Physiol*. 2019;1–15. https://doi.org/10.1002/jcp.28546