Clinical Trial/Experimental Study

Medicine®

OPEN

Identification of grade-related genes and construction of a robust genomic-clinicopathologic nomogram for predicting recurrence of bladder cancer

Xiqi Peng, MM[a,b], Jingyao Wang, MM[a], Dongna Li, MM[b], Xuan Chen, MM[a,b], Kaihao Liu, MM[a,c], Chunduo Zhang, MM[a], Yongqing Lai, PhD[a,*]

Abstract

Background: Bladder cancer (BC) is a common tumor in the urinary system with a high recurrence rate. The individualized treatment and follow-up after surgery is the key to a successful outcome. Currently, the surveillance strategies are mainly depending on tumor stage and grade. Previous evidence has proved that tumor grade was a significant and independent risk factor of BC recurrence. Exploring the grade-related genes may provide us a new approach to predict prognosis and guide the post-operative treatment in BC patients.

Methods: In this study, the weighted gene co-expression network analysis was applied to identify the hub gene module correlated with BC grade using GSE71576. After constructing a protein–protein interaction (PPI) network with the hub genes inside the hub gene module, we identified some potential core genes. TCGA and another independent dataset were used for further validation.

Results: The results revealed that the expression of AURKA, CCNA2, CCNB1, KIF11, TTK, BUB1B, BUB1, and CDK1 were significantly higher in high-grade BC, showing a strong ability to distinguish BC grade. The expression levels of the 8 genes in normal, paracancerous, tumorous, and recurrent bladder tissues were progressively increased. By conducting survival analysis, we proved their prognostic value in predicting the recurrence of BC. Eventually, we constructed a prognostic nomogram by combining the 8-core-gene panel with clinicopathologic features, which had shown great performance in predicting the recurrence of BC.

Conclusion: We identified 8 core genes that revealed a significant correlation with the tumor grade as well as the recurrence of BC. Finally, we proved the value of a novel prognostic nomogram for predicting the relapse-free survival of BC patients after surgery, which could guide their treatment and follow-up.

Abbreviations: BC = bladder cancer, BPs = biological processes, CCs = cellular components, DAVID = Database for Annotation, Visualization and Integrate Discovery, DCA = decision curve analysis, GEO = Gene Expression Omnibus, GO = Gene Ontology, GS = gene significance, GSEA = Gene set enrichment analysis, KEGG = Kyoto Encyclopedia of Genes and Genomes, KM = Kaplan–Meier, MEs = Module eigengenes, MFs = molecular functions, MS = Module significance, PPI = protein-protein interaction, RFS = relapse-free survival, ROC = receiver operating characteristic, STRING = Search Tool for the Retrieval of Interacting Genes, TCGA = The Cancer Genome Atlas, TOM = topological overlap matrix, TPM = transcripts per million reads, WGCNA = Weighted gene co-expression network analysis.

Keywords: bladder cancer, recurrence, weighted gene co-expression network analysis, prognostic nomogram

Editor: Jia Jing.

XP and JW contributed equally to this work.

This study was supported by Basic Research Project of Peking University Shenzhen Hospital (JCYJ2017001, JCYJ2017004, JCYJ2017005, JCYJ2017006, JCYJ2017007, JCYJ2017012), Clinical Research Project of Peking University Shenzhen Hospital (LCYJ2017001), Science and Technology Development Fund Project of Shenzhen (no. JCYJ20180507183102747) and Clinical Research Project of Shenzhen Health Commission (no. SZLY2018023).

The authors declare no conflict of interest.

Supplemental Digital Content is available for this article.

The datasets generated during and/or analyzed during the current study are publicly available.

[a] Guangdong and Shenzhen Key Laboratory of Male Reproductive Medicine and Genetics, Peking University Shenzhen Hospital, Shenzhen, [b] Shantou University Medical College, Shantou, Guangdong, [c] Anhui Medical University, Hefei, Anhui, China.

[*] Correspondence: Yongqing Lai, Guangdong and Shenzhen Key Laboratory of Male Reproductive Medicine and Genetics, Peking University Shenzhen Hospital, 1120 Lianhua Road, Shenzhen, Guangdong 518036, China (e-mail: yqlord@163.com).

How to cite this article: Peng X, Wang J, Li D, Chen X, Liu K, Zhang C, Lai Y. Identification of grade-related genes and construction of a robust genomic-clinicopathologic nomogram for predicting recurrence of bladder cancer. Medicine 2020;99:47(e23179).

Received: 2 June 2020 / Received in final form: 24 September 2020 / Accepted: 6 October 2020

http://dx.doi.org/10.1097/MD.0000000000023179

1. Introduction

Bladder cancer (BC) is one of the most common malignant tumors in the urinary system. In 2018, about 81,190 new cases and 17,240 deaths of BC were expected in the United States.[1] Although the 5-year survival rate of early-stage BC patients reaches 95.7%,[2] about 70% of them will suffer from recurrence, and approximately 10% to 20% will undergo tumor progression,[3] even followed by metastasis. Once tumor progress, the prognosis of patients will be poor, and limited effective treatment will be available. The individualized treatment and follow-up after surgery is the key to a successful outcome in BC patients. Previous studies have proved that tumor grade was a significant and independent risk factor of BC recurrence.[4,5] Exploring the grade-related genes may provide us a new approach to the prediction of prognosis and the post-operative treatment in BC patients.

Microarray and high-throughput sequencing have served as powerful weapons on the research of molecular mechanisms and therapeutic targets. Many studies have applied these gene expression profiles to identify genes related to malignant tumors.[6,7] Since the initiation, progression, and recurrence of BC are regulated by multiple genes, identifying crucial genes based on gene expression analysis, which only focuses on an individual gene, is deficient. Weighted gene co-expression network analysis (WGCNA), a systematic biology algorithm, have been served as a tool to describe correlation among genes, determine modules with highly correlated gene expression in microarray samples, and relate modules to particular clinical phenotype.[8,9] The most central and connected genes inside the chosen module will be regarded as core genes, which usually function crucially. Based on this method, we attempt to create a co-expression network of relationships between genes and identify the genes undertaking critical roles in BC. We further validated the core genes we selected using independent data from TCGA and GEO database and explored their potential as prognostic biomarkers in clinical use.

2. Material and methods

2.1. Data collection

Gene expression profile and corresponding clinical data of GSE71576[10] were acquired from Gene Expression Omnibus (GEO) database, which was based on Affymetrix Human Gene 1.0 ST Array, containing 44 primary bladder cancer tissues. GSE71576 was applied to create co-expression networks and exploring core genes in the present study. As for the external validation of the core genes, the RNA-sequencing data of 412 BC patients were downloaded from The Cancer Genome Atlas (TCGA) database (https://www.cancergenome.nih.gov) was obtained. And we download another GEO datasets GSE13507[11] based on Illumina human-6 v2.0 expression beadchip, which included 165 primary bladder cancer tissues, 58 normal looking bladder mucosae surrounding cancer, 23 recurrent bladder tumor tissues and 10 normal bladder mucosae. All the data mentioned above are open access and thus ethical approval is not necessary. The information of all samples used in this study was uploaded in the supplementary files, including samples from GSE71576, GSE13507, and TCGA datasets (see Table, http://links.lww.com/MD/F185, Supplemental Content, which includes all sample information used in this study).

2.2. Co-expression network construction

Normalized data from dataset GSE71576 were acquired from the GEO database. As those genes with little variation mostly represent noise, we selected the top 50% most variant genes by median absolute deviation for subsequent analysis. Using the "WGCNA" package in R,[12] we created a scale-free co-expression network of the filtered genes. We used the adjacency function in the WGCNA package to check if outlier samples exist for ensuring the reliability of the network. Then, goodSamplesGenes function in the WGCNA package was used to eliminate samples and genes containing too many missing values. The soft threshold power β was determined according to the scale-free topology criterion from Zhang and Horvath.[9] Pearsons correlations between each gene pair were calculated to determine the concordance of gene expression to create a matrix of adjacencies, which was later transformed into a topological overlap matrix (TOM).[13] The hierarchical clustering function was applied to gather genes with similar expression profiles into modules with a minimum gene size of 30 and a medium sensitivity of 2.

2.3. Identification of clinical significant modules

Two methods were utilized to recognize modules related to clinical traits of BC. Firstly, the log10 transformation of the *P* value ($GS = \lg p$) in the linear regression between clinical traits and gene expression was defined as gene significance (GS). The average GS of all genes in one module was defined as module significance (MS). The module with the highest MS value was usually regarded as the one most correlated with clinical traits. Module eigengenes (MEs) were defined as the predominant component in the principal component analysis of each gene module, and the expression of MEs was considered as a representative of all genes in a given module. The correlation between MEs and clinical traits was calculated to identify the clinically significant module. The most correlative module was selected for subsequent analysis.

2.4. Identification of candidate grade-related hub genes

Genes with high clinical trait relationship and module connectivity were considered to be candidate grade-related genes, selected by the absolute value of the Pearsons correlation ($|\text{cor.GeneModuleMenbership}| > 0.8$ and $|\text{cor.GeneTraitSignificance}| > 0.2$). Later, the candidate genes were uploaded to Database for Annotation, Visualization and Integrate Discovery (DAVID) (http://david.abcc.ncifcrf.gov/) for functional annotation, including gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. We chose the top 10 terms with the lowest *P* value ($P < .05$ as the cut-off criteria) to be listed. Subsequently, all these candidate genes were uploaded to the Search Tool for the Retrieval of Interacting Genes (STRING) database[14] to build a protein–protein interaction (PPI) network. After calculating the connectivity degree of all genes in the PPI network by cytoHubba in Cytoscape software (version 3.7.1), we selected 11 genes that were ranked top10 (connectivity degree >170) as grade-related core genes for further analysis.

2.5. Grade-related core gene validation and clinical value evaluation

To validate the candidate grade-related hub genes, the RNA sequencing data and corresponding clinical information of BC

were downloaded from The Cancer Genome Atlas Project database (TCGA, https://cancergenome.nih.gov/). The gene expression level was measured as transcripts per million reads (TPM). We compared the gene expression of subgroups classified by our aimed clinical trait. Another dataset GSE13507 from the GEO database was applied to confirm our results by expression difference analysis. Those significant genes were further confirmed by Kaplan–Meier (KM) analysis using TCGA data and receiver operating characteristic curve (ROC) using GSE13507 data. Additionally, we performed the validation of immunochemistry staining on Human Protein Atlas (http://www.proteinatlas.org).

2.6. Construction of nomogram

The prognostic value of each core grade-related gene was evaluated using Cox univariate regression analysis in the TCGA cohort, and the statistically significant genes were combined into a gene panel. We randomly divided TCGA data into 2 equal groups, naming as training group, and validation group. Based on the 8-core-gene panel, a risk score for each patient in the training group was calculated using their coefficients of Cox regression analysis and separated them into high- and low-risk groups with the median as the cut-off. Kaplan–Meier curve and Cox regression analysis were conducted to confirm the prognostic value of the risk score. A nomogram comprising of the risk score and clinicopathologic features was constructed. We evaluated the model using Calibration plots, decision curve analysis (DCA), time-dependent receiver operating characteristic (ROC) curve, and Harrells concordance index (c-index). All these processes were performed on R software (version 3.5.3).

2.7. Gene set enrichment analysis (GSEA)

To investigate the potential mechanisms of the core genes, GSEA[15] was conducted based on the level of risk scores to detect whether a series of prior defined biological processes were enriched in the gene rank derived from DEGs between the high and low-risk groups. We selected the collection of annotated gene sets of c2.cp.kegg.v6.0.symbols.gmt in Molecular Signatures Database (MSigDB, http://software.broadinstitute.org/gsea/msigdb/index.jsp) as the reference gene sets. The cut-off criteria were Nominal $P < .05$, $|ES| > 0.6$ and FDR < 25%.

2.8. Statistical analysis

The differences between the 2 groups were estimated by the unpaired Student *t* test in normally distributed variables and the Mann–Whitney *U* test in non-normally distributed variables. The comparisons of 2 more groups were analyzed via Kruskal-Wallis and one-way ANOVA test as non-parametric and parametric methods, respectively. The correlation was measured using Spearman correlation analysis. Survival rates were computed using the Kaplan–Meier method, and the log-rank test was used to evaluate the differences between the survival curves. Uni- and multivariate analyses were performed utilizing Cox proportional hazard models with the stepwise method "LR forward". The construction and validation of the nomogram were conducted according to lasonos guide.[16] All statistical analyses were processed using the R software (version 3.5.3) and SPSS software (version 22.0). *P* values were two-tailed. $P < .05$ was set as statistical significance.

3. Results

3.1. Data preprocessing and construction of co-expression network

Normalized data of GSE71576 was downloaded from GEO database, and probes inside were mapping to a total of 18,275 genes. The top 50% most variant genes (9137 genes) measuring by MAD values were used for the subsequent analysis. Two samples were excluded for lacking clinical information. After outlier detection, no outlier sample was found (Fig. 1A). To ensure the scale-free of the network, we selected the power of $\beta = 9$ as the soft-thresholding parameter (see Fig. S1, http://links.lww.com/MD/F182, Supplemental Content, which illustrates the determination of soft-thresholding power in the co-expression network.). After classifying genes into different modules by hierarchical clustering, 9 modules named with unique colors were identified (Fig. 1B). Two approaches were conducted to determine the most correlative module with clinical traits (stage and grade). Among all the modules, the ME of the blue module ($P = 2e\text{-}15$; $R^2 = 0.89$) revealed a strong correlation with tumor grade (Fig. 1C). Additionally, the MS of the blue module was the highest among all modules (Fig. 1D). The relationship between MM and GS of the blue module is shown in Fig. 1E. Therefore, the blue module, which included 1351 genes, was considered as the hub module that was significantly correlated with tumor grade. Two hundred thirty nine genes in the blue module were filtered as hub genes by the cut-off criteria ($|MM| > 0.8$ and $|GS| > 0.2$) and put into the following analysis.

3.2. Functional enrichment analysis of hub genes

The 239 filtered hub genes in the blue module were uploaded to DAVID (http://david.abcc.ncifcrf.gov/). The results of the GO analysis revealed that these genes were enriched in various biological processes (BPs), cellular components (CCs), and molecular functions (MFs) (Fig. 2A-C). The top 10 GO terms of BP included cell division, mitotic nuclear division, DNA replication, sister chromatid cohesion, DNA repair, G1/S transition of mitotic cell cycle, chromosome segregation, DNA replication initiation, microtubule–based movement and mitotic sister chromatid segregation (Fig. 2A). The results of CCs revealed that these genes were mainly enriched in the nucleus and nucleoplasm (Fig. 2B). The top 10 GO terms of MF included protein binding, ATP binding, DNA binding, microtubule binding, chromatin binding, microtubule motor activity, ATPase activity, single–stranded DNA binding, single–stranded DNA–dependent ATPase activity, and ATP–dependent microtubule motor activity, plus–end–directed (Fig. 2C). As for KEGG analysis, they were mainly enriched in cell cycle, DNA replication, oocyte meiosis, Fanconi anemia pathway, progesterone–mediated oocyte maturation, systemic lupus erythematosus, pyrimidine metabolism, homologous recombination, p53 signaling pathway, and mismatch repair (Fig. 2D).

3.3. Identification of grade-related core genes

The 239 filtered genes in the blue module were uploaded to the STRING database 3[14] to build a protein–protein interaction (PPI) network. After calculating the connectivity degree of all genes in the PPI network, we selected 11 genes that were ranked top10 (connectivity degree >170) as grade-related core genes, including AURKA, CCNA2, CCNB1, KIF11, TTK, BUB1B,

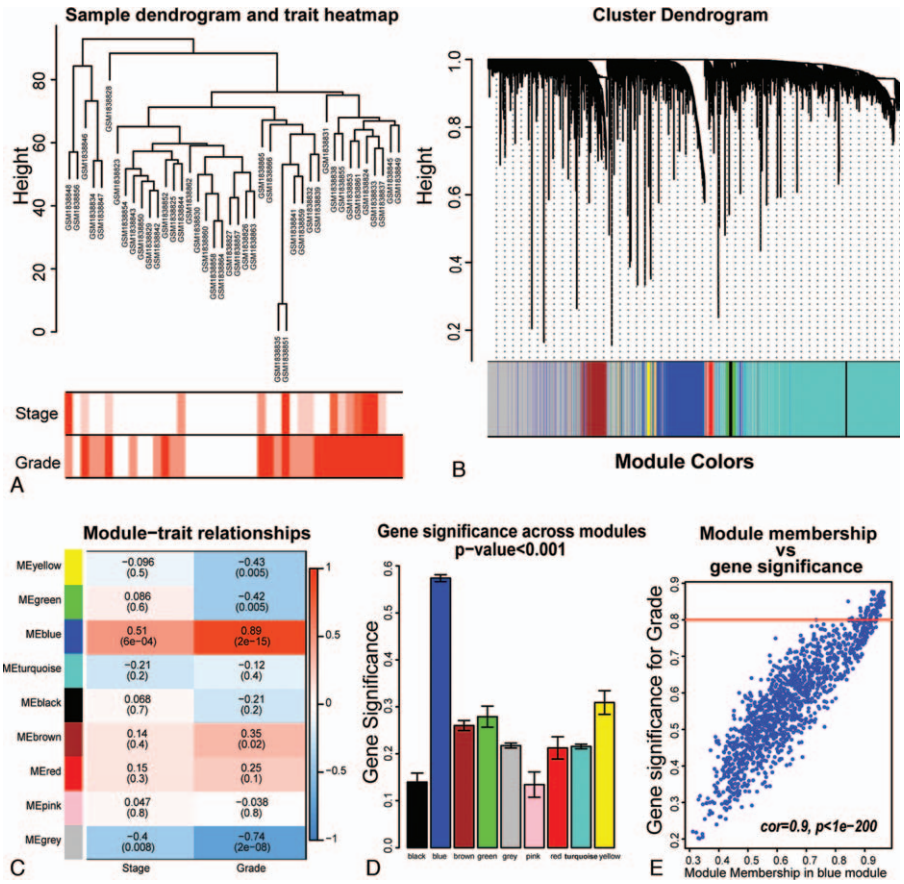

Figure 1. Weighted gene co-expression network analysis. (A) Clustering dendrogram of 42 BC samples and clinical trait heatmap. (B) Dendrogram of the top 50% most variant genes clustered based on a dissimilarity measure (1-TOM). (C) Heatmap of the correlation values between MEs and different clinical traits of BC (tumor stage and grade). Red for positive correlation and Blue for negative correlation with *P* values printed below the correlations. (D) Distribution of average gene significance in the modules associated with tumor grades of BC. (E) Scatter plot of GS for tumor grade versus kME for blue module. BC = bladder cancer, MEs = module eigengenes, GS = Gene Significance, kME = module membership.

BUB1, CDK1, AURKB, CCNB2, and CDC20. The PPI network of grade-related genes was shown in Figure 2E. The core genes were surrounded by others in the middle of the network.

3.4. Grade-related core gene validation in TCGA

To validate the 11 grade-related genes we found, we download the RNA sequencing data and corresponding clinical information of 412 BC tissue from the TCGA database. We compared the expression level of these genes between high-grade and low-grade BC patients. The results showed that the expression levels of all these genes were significantly higher in the high-grade group than the low-grade (Fig. 3).

3.5. Grade-related core gene clinical value assessment

We conducted survival analysis using KM analysis and Cox univariate regression analysis of these genes using TCGA data. In univariate regression analysis, AURKA, CCNA2, CCNB1, KIF11, BUB1B, BUB1, TTK, and CDK1 were significantly associated with relapse-free survival (RFS) of BC patients. The *P* value of univariate regression analysis of them was 0.00635, 0.00756, 0.0255, 0.00752, 0.00301, 0.0364, 0.00679, and 0.0174, respectively. The KM curves revealed that CCNA2, CCNB1, KIF11, BUB1B, TTK, and AURKA were significantly negatively associated with the RFS of BC patients according to quartile cut-off (Fig. 4). Although AURKA and BUB1 were not significantly associated with RFS in KM analysis, we kept them in our further analysis on account of the results of univariate regression analysis.

3.6. Grade-related core gene multi-analysis with independent GEO dataset

To find more evidence for supporting our results, we obtained data of another dataset GSE13507 from GEO database, which included 165 primary bladder cancer samples, 58 normal looking bladder mucosae surrounding cancer, 23 recurrent non-muscle invasive tumor tissues and 10 normal bladder mucosae. Similar results in tumor grade were shown in this dataset (Fig. 5). Interestingly, comparing the expression level of these genes in normal, paracancerous, tumorous and recurrent groups, markedly progressive increase was present (Fig. 6). Furthermore, we constructed ROC curves of the 8 grade-related core genes to explore their ability of differentiating high grade and low grade patients. The results indicated that AURKA, CCNA2, CCNB1, KIF11, TTK, BUB1B, BUB1 and CDK1 all exhibit excellent

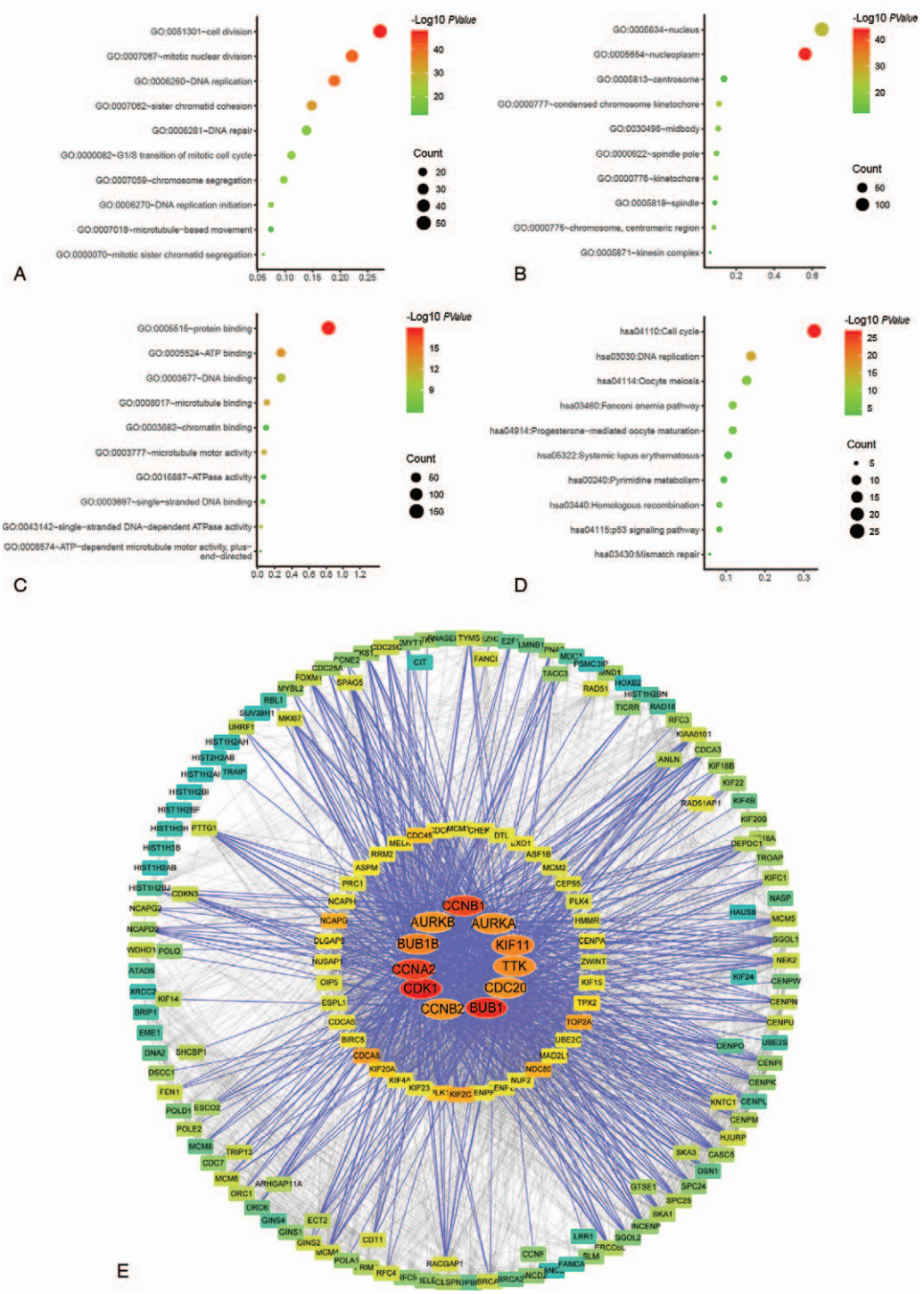

Figure 2. Functional Enrichment Analysis and PPI network of the 239 hub genes from blue module. (A) Biological process analysis. (B) Cellular component analysis. (C) Molecular function analysis. (D) KEGG pathway analysis. (E) PPI networks of 239 hub genes in blue module. PPI, protein–protein interaction.

discriminatory ability (Fig. 7). The area under curve (AUC) of them were 0.735 (CI 0.654-0.816, Sensitivity = 73.3%, Specificity = 72.4%), 0.757 (CI 0.681-0.833, Sensitivity = 66.7%, Specificity = 78.1%), 0.722 (CI 0.639-805, Sensitivity = 80.0%, Specificity = 61.0%), 0.737 (CI 0.658-0.817, Sensitivity = 80.0%, Specificity = 65.7%), 0.77 (CI 0.693-0.848, Sensitivity = 65.0%, Specificity = 81.9%), 0.767 (CI 0.687-0.846, Sensitivity = 75.0%, Specificity = 80.0%), 0.763 (CI 0.685-0.840, Sensitivity = 75.0%, Specificity = 72.4%), and 0.796 (CI 0.720-0.871, Sensitivity = 85.0%, Specificity = 71.4%). The cut-off points were chosen at the point with the largest Youden index (sensitivity + specificity-1).

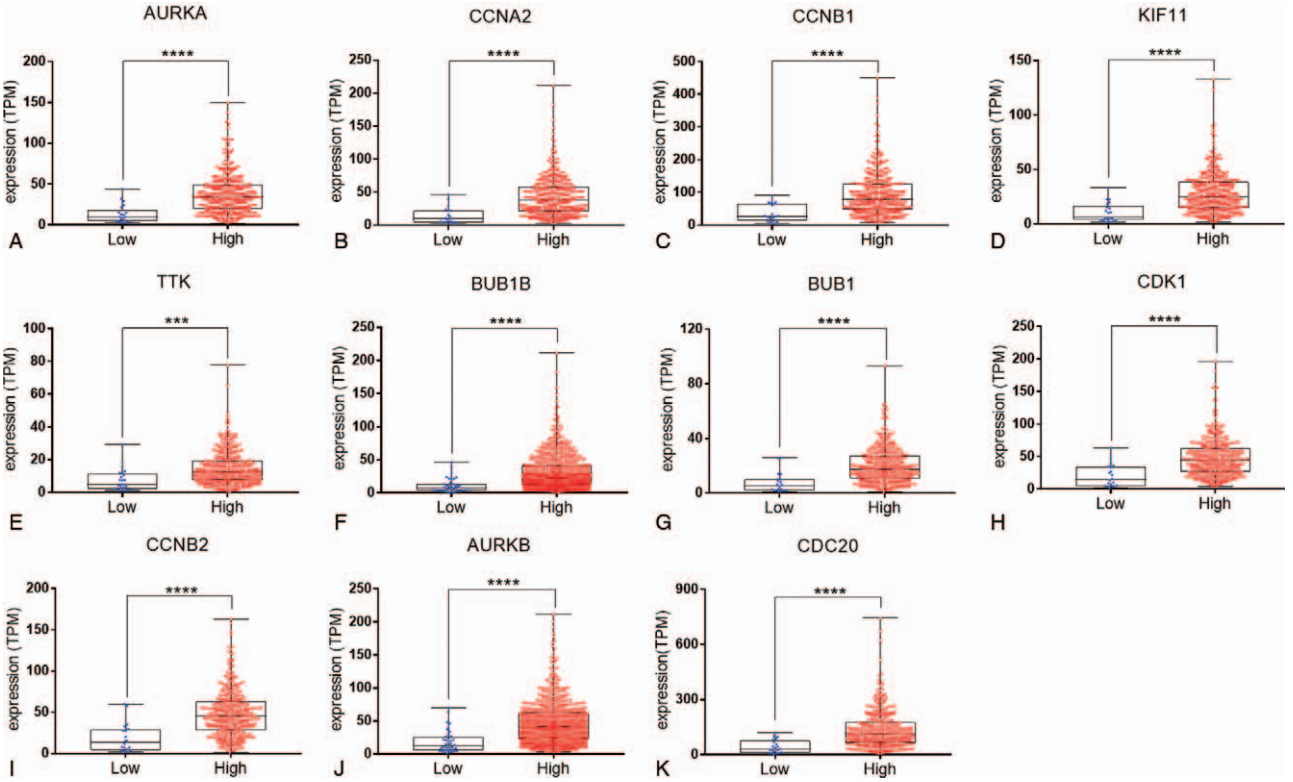

Figure 3. Expression of grade-related core genes between high- and low-grade BC based on TCGA data. (A) AURKA (B) CCNA2 (C) CCNB1 (D) KIF11 (E) TTK (F) BUB1B (G) BUB1 (H) CDK1. BC = bladder cancer, TCGA = the Cancer Genome Atlas Project database.

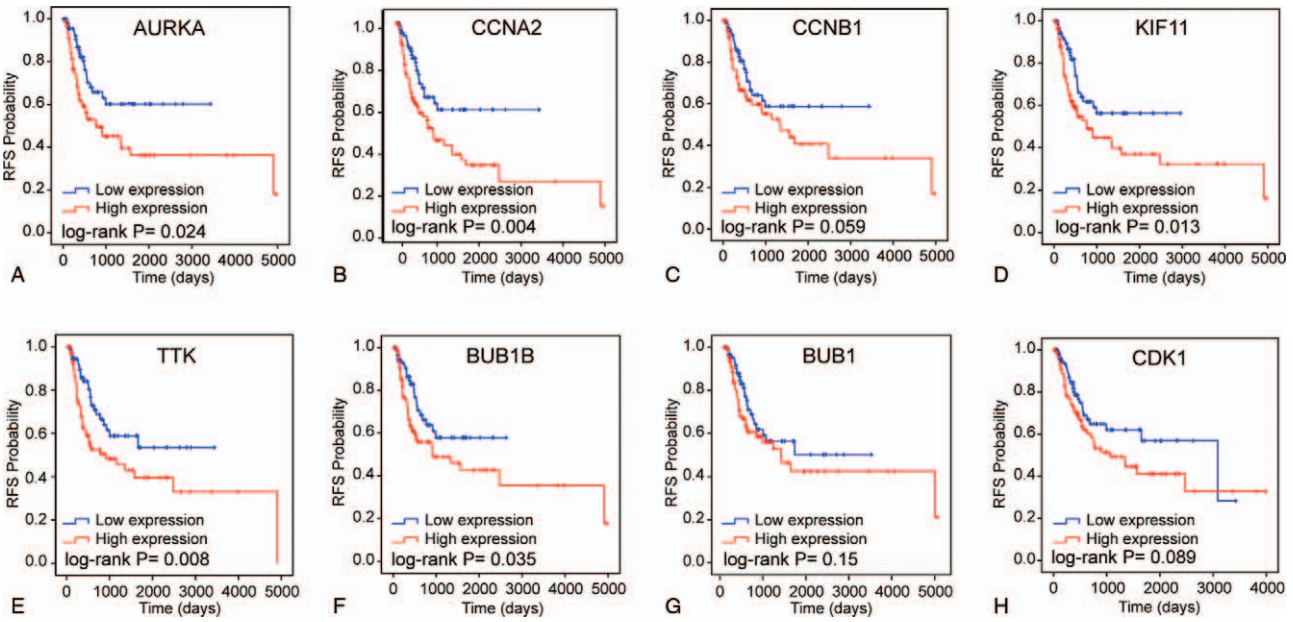

Figure 4. Relapse-free survival (RFS) of grade-related core genes in BC patients based on TCGA data. Patients were classified into high expression and low expression group with median cut-off. (A) AURKA (B) CCNA2 (C) CCNB1 (D) KIF11 (E) TTK (F) BUB1B (G) BUB1 (H) CDK1. BC = bladder cancer, TCGA = the Cancer Genome Atlas Project database.

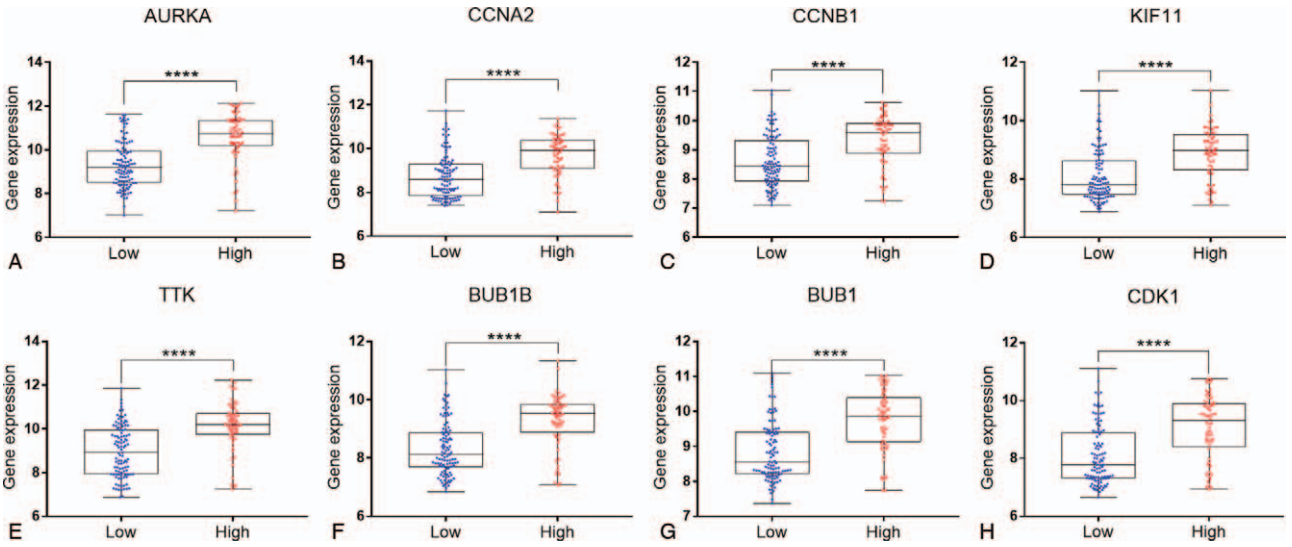

Figure 5. Expression of grade-related core genes between high and low grade BC based on GEO datasets GSE13507. (A) AURKA (B) CCNA2 (C) CCNB1 (D) KIF11 (E) TTK (F) BUB1B (G) BUB1 (H) CDK1. BC = bladder cancer; GEO = Gene Expression Omnibus (GEO) database.

3.7. Immunohistochemistry validation of the grade-related core genes

The protein levels of AURKA, CCNA2, CCNB1, KIF11, TTK, and CDK1 were markedly higher in tumor tissues than normal tissues according to the results of the Human Protein Atlas database, except that no data of BUB1 and BUB1B was found (see Fig. S2, http://links.lww.com/MD/F183, Supplemental Content, which demonstrates the immunohistochemistry of the grade-related core genes from the Human Protein Atlas database).

3.8. Nomogram construction and assessment

After calculating the risk score of the 8-core-gene panel for each BC patients in the training group from TCGA, we divided them into high- and low-risk groups. The result of the Kaplan-Meier curve showed that patients in the high-risk group had a conspicuously higher probability of recurrence than the low-risk group (Fig. 8A). Using Cox univariate and multivariate analysis, we found that risk score ($P < .001$) and tumor stage ($P < .001$) were probably independent predictive factors for BC patients while gender ($P = .42$) and age ($P = .511$) were highly irrelevant (Table 1). Considering that smoking ($P = .285$) and tumor grade ($P = .134$) were of some value (Table 1), we combined them with risk score and tumor stage to construct a prognostic nomogram (Fig. 8B). The calibration plots of the training group and validation group both performed well when compared with the ideal model (Fig. 8C-D). The nomogram also exhibited a better predictive accuracy than only using the clinicopathologic features (stage, grade, smoking), in the results of time-dependent ROC (Fig. 8E-F), decision curve (Fig. 8G-L) and C-index (Table 2).

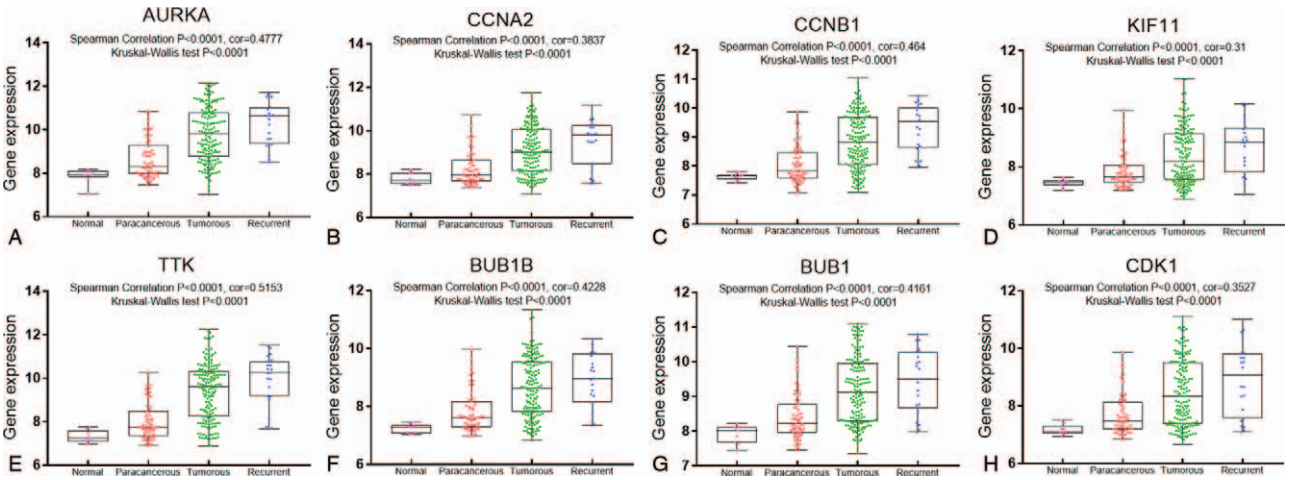

Figure 6. Expression of grade-related genes in different kind of tissues including normal bladder, paracancerous tissue, tumorous tissue and recurrent tumor tissue. (A) AURKA (B) CCNA2 (C) CCNB1 (D) KIF11 (E) TTK (F) BUB1B (G) BUB1 (H) CDK1.

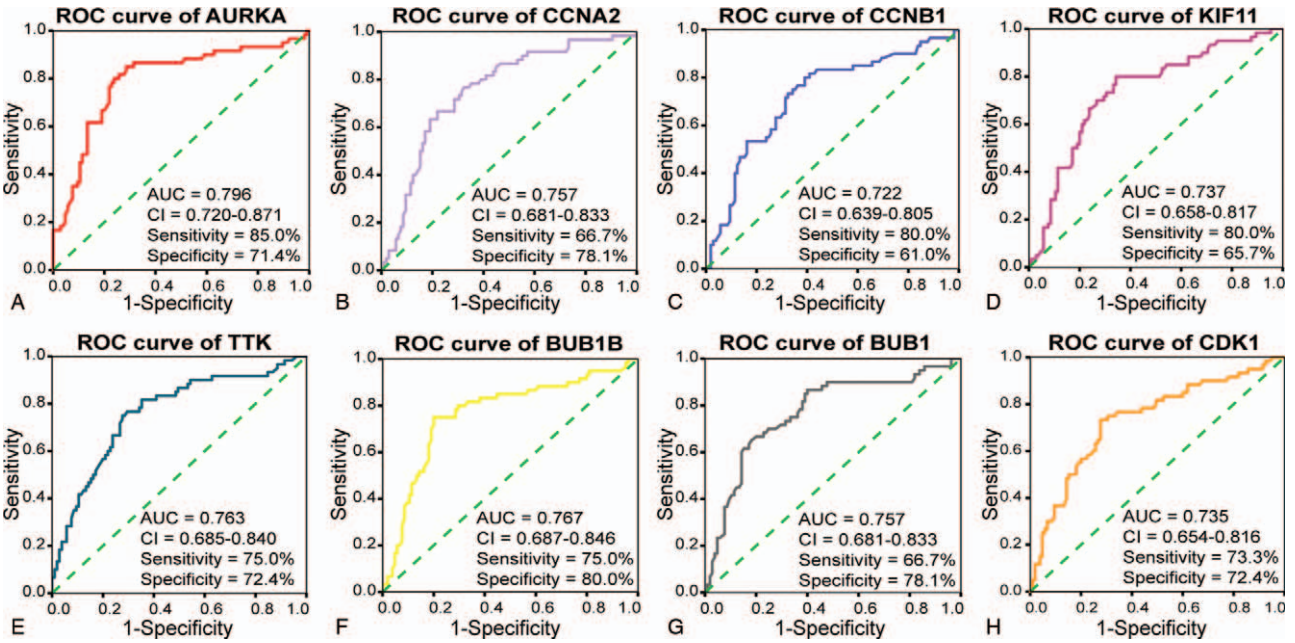

Figure 7. Receiver operating characteristic curve (ROC) of grade-related core genes for differentiating tumor grade of BC. (A) AURKA (B) CCNA2 (C) CCNB1 (D) KIF11 (E) TTK (F) BUB1B (G) BUB1 (H) CDK1. BC, bladder cancer.

3.9. Gene set enrichment analysis (GSEA)

To explore the potential mechanisms that the 8-core-gene panel was relevant within BC, GSEA was performed to search the enriched KEGG pathways based on the level of the risk score in the whole TGCA cohort with the cut-off of the median. The results revealed that genes highly expressed in the high-risk group were significantly enriched in multiple pathways. The most significantly correlative pathways were basal transcriptional factors, cell cycle, DNA replication, base excision repair, and nucleotide excision repair (Fig. 9).

4. Discussion

Due to the high recurrence rate of both non-muscle-invasive and muscle-invasive BC and the short time to progression and death in patients with metastasis, surveillance is the key to a better outcome. The management and follow-up of BC are both complicated and high-cost, which is causing a dilemma in clinical work. Currently, the surveillance strategies are mainly determined by tumor grade and stage, which is not dissatisfactory enough.[17] Actually, unlike many other tumors, the survival rate of BC has not increased over the past 3 decades.[18] In recent years, some researchers have developed models for predicting BC recurrence and progression. Among them, EORTC risk table, a predictive tool based on clinical and pathological factors, is one of the best tools by far.[19] However, the predictive accuracy of EORTC risk table is still not good enough to meet the current demand in clinical work.[20]

In this study, we performed integrated bioinformatics analyses based on microarray and high-throughput sequencing data, and finally identified 8 core genes that were closely related to tumor grade and recurrence of BC. The nomogram comprising of the 8-core-gene panel and clinicopathologic features exhibit a great performance in the prediction of RFS in BC patients, revealing its potential to guide our clinical decision. The 8-core-gene panel as a supplement of EORTC risk table will be our important research direction in future.

To gain a deeper understanding of the function of the 8 grade-related core genes in BC, GSEA analysis was performed. According to the results of GSEA, BC samples with higher expression of the core genes were mostly enriched in genes sets that were linked with pathways of basal transcriptional factors, cell cycle, DNA replication, base excision repair, and nucleotide excision repair, which were the key pathways for tumorigenesis and development. For a deeper insight into the 8 core genes, we carried out a literature review of them.

AURKA encodes a cell cycle-regulated kinase that associates with the centrosome and the spindle microtubules during mitosis and plays a vital role in various mitotic events. The alteration of this protein may result in aberrant mitotic spindles, leading to tumorigenesis. Previous studies on TTK had verified its role in several cancers, such as hepatocellular carcinoma,[21] pancreatic cancer,[22] breast cancer,[23] and leukemia.[24] As for bladder cancer, several studies had demonstrated the roles of AURKA in tumor cell function and its correlation with prognosis, which were powerful support of our study results.[25–27]

CCNA2 gene encodes cyclin A2 (CCNA2), which regulates cell cycle progression by interacting with CDK kinase and may participate in tumorigenesis. CCNA has been proved to be highly expressed in multiple kinds of tumors such as liver cancer,[28] cervical cancer,[29] and breast cancer.[30] By targeting cyclin A2, miR-27b can suppress the proliferation of leukemia.[31] Similarly, miR-449a and miR-424 can suppress osteosarcoma by targeting cyclin A2.[32] As for bladder cancer, Li J demonstrated that CCNA2 played a regulatory role in modulating CDK6 and MET-mediated cell-cycle pathway and EMT progression.[33]

G2/Mitotic-Specific Cyclin-B1, a regulatory protein encoded by CCNB1 in humans and involved in mitosis, is essential for the control of G2/M transition phase in cell cycle. By targeting the

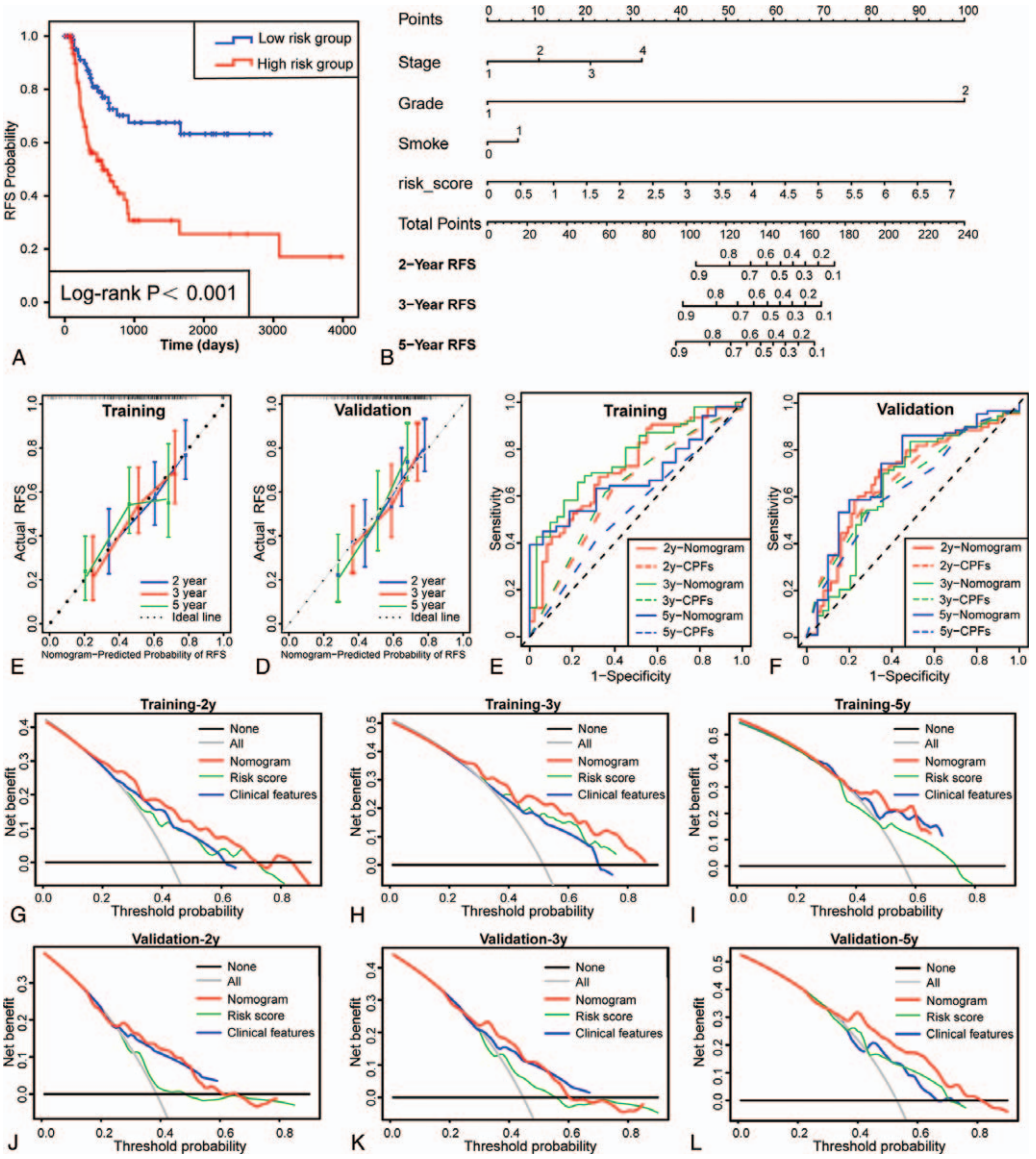

Figure 8. Construction and assessment of Nomogram. (A) RFS of the risk score based on the 8 core-gene panel with median cut-off. (B) Nomogram combining the 8-core-gene panel with CPFs. (C-D) Calibration plots of the training group and validation group in observed time of 2 year, 3 year and 5 year. (E-F) Time-dependent ROC of the training group and validation group for the predictive ability of recurrence using the nomogram or CPFs. (G-I) DCA of the training group for 2-year, 3-year and 5-year risk. (J-L) DCA of the validation group for 2-year, 3-year and 5-year risk. ROC = receiver operating characteristic curve, DCA = decision curve analysis.

Table 1

The association between the 8-core-gene panel and clinicopathologic features with RFS of BC patients.

	Univariate analysis		Multivariate analysis	
Variable	**HR (95% CI)**	***P* value**	**HR (95% CI)**	***P* value**
Gender (male vs female)	1.164 (0.805–1.681)	.42		
Age	1.005 (0.989–1.022)	.511		
Smoking (yes vs no)	1.202 (0.858–1.682)	.285	1.230 (0.878–1.723)	.229
Tumor grade (high vs low)	2.917 (0.72–11.82)	.134	1.304 (0.3122–5.444)	.716
Tumor stage	1.704 (1.374–2.113)	**<.001**	1.671 (1.343–2.080)	**<.001**
Gene panel risk score	2.735 (1.703–4.165)	**<.001**	2.537 (1.587–4.056)	**<.001**

BC = bladder cancer, HR = hazard ratio, 95% CI = 95% confidence interval, RFS = Relapse-free survival.

CCNB1-related pathway, Ngan AWL found that the proliferation of cervical cancer cells could be regulated, revealing a potential therapeutic target.[34] In addition, the study of Yan X had demonstrated that CCNB1 was a key gene and associated with the prognosis of BC.[35]

Table 2

Harrell concordance indexes of the 8-core-gene panel, clinicopathologic features*, and nomogram in different cohorts.

Cohort	8-core-gene panel	Clinicopathologic features	Nomogram
Training	0.651	0.661	**0.708**
Validation	0.610	0.603	**0.659**
Entire	0.59	0.635	**0.667**

* including tumor stage, tumor grade, and smoking.

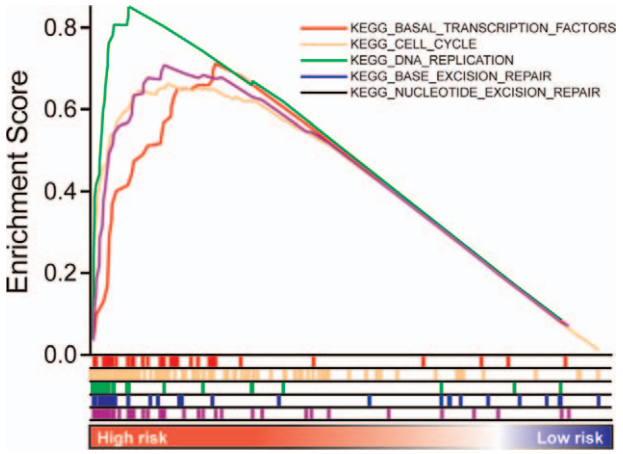

Figure 9. Gene set enrichment analysis based on the level of risk score calculating by the 8-core-gene panel with median cut-off.

Kinesin family member 11 (KIF11) takes essential roles in the formation of bipolar spindle and maintenance in the early prometaphase during mitosis.[36,37] KIF11 had been proved to be correlated with various malignancies, including prostate cancer,[38] breast cancer,[39] malignant pleural mesothelioma,[40] pancreatic cancer,[41] laryngeal squamous cell carcinoma,[42] and gastric cancer .[43] Currently, there is no clear evidence in the relationship between KIF11 and BC.

TTK encodes a protein that acts a basic role in alignment at the centromere during mitosis and is necessary for centrosome duplication. It plays critical roles in aneuploidy and genomic integrity in various kinds of tumors. In breast cancer, TTK had already been proved to be available therapeutic target in several studies.[44,45] The association of TTK with BC are remained to be investigated.

BUB1 Mitotic Checkpoint Serine/Threonine Kinase B (BUB1B), a pivotal component of the spindle assembly checkpoint requiring for accurate segregation of chromosomes, is functional in the mitotic checkpoint and the establishment of proper microtubule–kinetochore attachments.[46] Aberrant expression or mutations of BUB1B is one of the important causes of aneuploidy. In fact, the role of BUB1B in different cancers is still controversial. Previous studies have proved the high expression of BUB1B was related to the progression and recurrence of several tumors, including gastric cancer,[47] hepatocellular carcinoma,[48] prostate cancer,[49] esophageal squamous cancer[50] and breast cancer,[51] while some tumors were proved to be related with low expression of BUB1B.[52,53] So far, little definite evidence is known about the correlation between BUB1B and BC.

BUB1 gene encodes a serine/threonine-protein kinase that plays a vital role in mitosis. Mutations in this gene are relevant to aneuploidy and several forms of cancer.[54,55] In breast cancer, it has been reported that BUB1 was associated with cancer progression[56] and cancer stem cell potential.[57] The correlation between BUB1 and bladder cancer is still unclear yet.

The protein encoded by CDK1 is essential for G1/S and G2/M phase transitions of the eukaryotic cell cycle. The formation of complex CDK1/cyclin B1 is essential for the G2/M phase transition of cell cycle. One of the latest research reported that upregulation of CDK1 was essential for nasal natural killer/T-cell lymphoma progression and targeting CDK1 might benefit the treatment of NNKTL.[58] Similar results were reported in cervical cancer by Li.[59] In bladder cancer, it is demonstrated that Glaucocalyxin A regulated the expression of CDK1 to induces G2/M cell cycle arrest and apoptosis.[60]

Taken all current results together, we believed that AURKA, CCNA2, CCNB1, KIF11, TTK, BUB1B, BUB1, and CDK1 were correlated with tumor grade of BC and were probably important participants in the initiation and recurrence of BC, by participating in pathways of basal transcriptional factors, cell cycle, DNA replication, base excision repair and nucleotide excision repair. Meanwhile, they have the potential to be practicable biomarkers for identifying BC patients at high risk, guiding the post-operative treatment and follow-up. Actually, the critical role of AURKA, CCNA2, CCNB1, and CDK1 in BC had already been reported before. The present study not only providing new evidence for them, but also identify 4 more new genes with great significance.

However, the limitations of the present study should be acknowledged. Firstly, our results were based on microarray and RNA-sequencing data, lacking in support of in vitro and in vivo experiments. Moreover, although we had combined several datasets to make our results more convincing, a larger sample size of external validation is required, which is the main direction of our future research.

5. Conclusions

In conclusion, the present study had conducted integrated bioinformatics analyses, eventually identifying 8 grade-related core genes that probably act crucially in the recurrence of BC. Furthermore, we proved their potential to be available prognostic biomarkers by constructing a nomogram for predicting recurrence of BC patients. Hopefully, further research could push these genes into clinical use.

Acknowledgments

We gratefully acknowledge Gene Expression Omnibus (GEO) database, The Cancer Genome Atlas (TCGA) database and Human Protein Atlas database, which provided public available data of bladder cancer.

Author contributions

Conceptualization: Xiqi Peng, Jingyao Wang, Yongqing Lai.
Data curation: Dongna Li, Xuan Chen, Kaihao Liu, Chunduo Zhang.
Formal analysis: Xiqi Peng.
Funding acquisition: Yongqing Lai.
Methodology: Xiqi Peng, Jingyao Wang, Xuan Chen.
Project administration: Dongna Li, Kaihao Liu, Chunduo Zhang, Yongqing Lai.
Resources: Xiqi Peng, Dongna Li, Kaihao Liu.
Software: Xiqi Peng, Xuan Chen.
Supervision: Jingyao Wang, Dongna Li, Yongqing Lai.
Validation: Jingyao Wang.
Writing – original draft: Xiqi Peng.
Writing – review & editing: Jingyao Wang.

References

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018;68:7–30.
[2] Antoni S, Ferlay J, Soerjomataram I, et al. Bladder cancer incidence and mortality: a global overview and recent trends. Eur Urol 2017;71:96–108.

[3] Kaufman DS, Shipley WU, Feldman AS. Bladder cancer. Lancet 2009;374:239–49.
[4] Burger M, van der Aa MN, van Oers JM, et al. Prediction of progression of non-muscle-invasive bladder cancer by WHO 1973 and 2004 grading and by FGFR3 mutation status: a prospective study. Eur Urol 2008;54:835–43.
[5] Chen Z, Ding W, Xu K, et al. The 1973 WHO Classification is more suitable than the 2004 WHO Classification for predicting prognosis in non-muscle-invasive bladder cancer. PLoS One 2012;7:e47199.
[6] Liu J, Li S, Liang J, et al. ITLNI identified by comprehensive bioinformatic analysis as a hub candidate biological target in human epithelial ovarian cancer. Cancer Manag Res 2019;11:2379–92.
[7] Zhu Q, Yang H, Cheng P, et al. Bioinformatic analysis of the prognostic value of the lncRNAs encoding snoRNAs in hepatocellular carcinoma. Biofactors 2019;45:244–52.
[8] Yip AM, Horvath S. Gene network interconnectedness and the generalized topological overlap measure. BMC Bioinformatics 2007;8:22.
[9] Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 2005;4: Article17.
[10] Pineda S, Real FX, Kogevinas M, et al. Integration analysis of three omics data using penalized regression methods: an application to bladder cancer. PLoS Genet 2015;11:e1005689.
[11] Lee JS, Leem SH, Lee SY, et al. Expression signature of E2F1 and its associated genes predict superficial to invasive progression of bladder tumors. J Clin Oncol 2010;28:2660–7.
[12] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559.
[13] Li A, Horvath S. Network module detection: affinity search technique with the multi-node topological overlap measure. BMC Res Notes 2009;2:142.
[14] Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019;47(D1):D607–13.
[15] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.
[16] Iasonos A, Schrag D, Raj GV, et al. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol 2008;26:1364–70.
[17] Kamat AM, Hahn NM, Efstathiou JA, et al. Bladder Cancer. Lancet 2016;388:2796–810.
[18] Berdik C. Unlocking bladder cancer. Nature 2017;551:S34–5.
[19] Sylvester RJ, van der Meijden AP, Oosterlinck W, et al. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. European Urol 2006;49: 466-465; discussion 475-467.
[20] Chen MK, Chen ZJ, Xiao KH, et al. Predictive value of cadherin-11 for subsequent recurrence and progression in non-muscle invasive bladder cancer. Jpn J Clin Oncol 2020;50:456–64.
[21] Liang XD, Dai YC, Li ZY, et al. Expression and function analysis of mitotic checkpoint genes identifies TTK as a potential therapeutic target for human hepatocellular carcinoma. PLoS One 2014;9:e97739.
[22] Kaistha BP, Honstein T, Muller V, et al. Key role of dual specificity kinase TTK in proliferation and survival of pancreatic cancer cells. Br J Cancer 2014;111:1780–7.
[23] King JL, Zhang B, Li Y, et al. TTK promotes mesenchymal signaling via multiple mechanisms in triple negative breast cancer. Oncogenesis 2018;7:69.
[24] Goldenson B, Crispino JD. The aurora kinases in cell cycle and leukemia. Oncogene 2015;34:537–45.
[25] Guo M, Lu S, Huang H, et al. Increased AURKA promotes cell proliferation and predicts poor prognosis in bladder cancer. BMC Syst Biol 2018;12(Suppl 7):118.
[26] Mobley A, Zhang S, Bondaruk J, et al. Aurora kinase A is a biomarker for bladder cancer detection and contributes to its aggressive behavior. Sci Rep 2017;7:40714.
[27] Zhou N, Singh K, Mir MC, et al. The investigational Aurora kinase A inhibitor MLN8237 induces defects in cell viability and cell-cycle progression in malignant bladder cancer cells in vitro and in vivo. Clin Cancer Res 2013;19:1717–28.
[28] Wen DY, Lin P, Pang YY, et al. Expression of the Long Intergenic Non-Protein Coding RNA 665 (LINC00665) gene and the cell cycle in hepatocellular carcinoma using the cancer genome Atlas, the gene expression omnibus, and quantitative real-time polymerase chain reaction. Med Sci Monit 2018;24:2786–808.
[29] Wu Y, Li H, Wang H, et al. MSK2 promotes proliferation and tumor formation in squamous cervical cancer via PAX8/RB-E2F1/cyclin A2 axis. J Cell Biochem 2019;Feb 12. doi: 10.1002/jcb.28421. Epub ahead of print.
[30] Gao T, Han Y, Yu L, et al. CCNA2 is a prognostic biomarker for ER+ breast cancer and tamoxifen resistance. PLoS One 2014;9:e91771.
[31] Wang B, Li D, Kovalchuk A, et al. Ionizing radiation-inducible miR-27b suppresses leukemia proliferation via targeting cyclin A2. Int J Radiat Oncol Biol Phys 2014;90:53–62.
[32] Shekhar R, Priyanka P, Kumar P, et al. The microRNAs miR-449a and miR-424 suppress osteosarcoma by targeting cyclin A2 expression. J Biol Chem 2019;294:4381–400.
[33] Li J, Ying Y, Xie H, et al. Dual regulatory role of CCNA2 in modulating CDK6 and MET-mediated cell-cycle pathway and EMT progression is blocked by miR-381-3p in bladder cancer. FASEB J 2019;33:1374–88.
[34] Ngan AWL, Grace Tsui M, So DHF, et al. Novel nuclear partnering role of EPS8 With FOXM1 in regulating cell proliferation. Front Oncol 2019;9:154.
[35] Yan X, Guo ZX, Liu XP, et al. Four novel biomarkers for bladder cancer identified by weighted gene coexpression network analysis. J Cell Physiol 2019;234:19073–87.
[36] Chauviere M, Kress C, Kress M. Disruption of the mitotic kinesin Eg5 gene (Knsl1) results in early embryonic lethality. Biochem Biophys Res Commun 2008;372:513–9.
[37] Fernandez JP, Aguero TH, Vega Lopez GA, et al. Developmental expression and role of Kinesin Eg5 during Xenopus laevis embryogenesis. Dev Dyn 2014;243:527–40.
[38] Piao XM, Byun YJ, Jeong P, et al. Kinesin family member 11 mRNA expression predicts prostate cancer aggressiveness. Clin Genitourin Cancer 2017;15:450–4.
[39] Jin Q, Huang F, Wang X, et al. High Eg5 expression predicts poor prognosis in breast cancer. Oncotarget 2017;8:62208–16.
[40] Kato T, Lee D, Wu L, et al. Kinesin family members KIF11 and KIF23 as potential therapeutic targets in malignant pleural mesothelioma. Int J Oncol 2016;49:448–56.
[41] Sun XD, Shi XJ, Sun XO, et al. Dimethylenastron suppresses human pancreatic cancer cell migration and invasion in vitro via allosteric inhibition of mitotic kinesin Eg5. Acta Pharmacol Sin 2011; 32:1543–8.
[42] Lu M, Zhu H, Wang X, et al. The prognostic role of Eg5 expression in laryngeal squamous cell carcinoma. Pathology 2016;48:214–8.
[43] Imai T, Oue N, Nishioka M, et al. Overexpression of KIF11 in gastric cancer with intestinal mucin phenotype. Pathobiology 2017;84:16–24.
[44] Maire V, Baldeyron C, Richardson M, et al. TTK/hMPS1 is an attractive therapeutic target for triple-negative breast cancer. PLoS One 2013;8: e63712.
[45] Zhu D, Xu S, Deyanat-Yazdi G, et al. Synthetic lethal strategy identifies a potent and selective TTK and CLK1/2 inhibitor for treatment of triple-negative breast cancer with a compromised G1-S checkpoint. Mol Cancer Ther 2018;17:1727–38.
[46] Baker DJ, Dawlaty MM, Wijshake T, et al. Increased expression of BubR1 protects against aneuploidy and cancer and extends healthy lifespan. Nat Cell Biol 2013;15:96–102.
[47] Ando K, Kakeji Y, Kitao H, et al. High expression of BUBR1 is one of the factors for inducing DNA aneuploidy and progression in gastric cancer. Cancer Sci 2010;101:639–45.
[48] Liu AW, Cai J, Zhao XL, et al. The clinicopathological significance of BUBR1 overexpression in hepatocellular carcinoma. J Clin Pathol 2009;62:1003–8.
[49] Fu X, Chen G, Cai ZD, et al. Overexpression of BUB1B contributes to progression of prostate cancer and predicts poor outcome in patients with prostate cancer. Onco Targets Ther 2016;9:2211–20.
[50] Tanaka K, Mohri Y, Ohi M, et al. Mitotic checkpoint genes, hsMAD2 and BubR1, in oesophageal squamous cancer cells and their association with 5-fluorouracil and cisplatin-based radiochemotherapy. Clin Oncol (R Coll Radiol) 2008;20:639–46.
[51] Yuan B, Xu Y, Woo JH, et al. Increased expression of mitotic checkpoint genes in breast cancer cells with chromosomal instability. Clin Cancer Res 2006;12:405–10.
[52] Park HY, Jeon YK, Shin HJ, et al. Differential promoter methylation may be a key molecular mechanism in regulating BubR1 expression in cancer cells. Exp Mol Med 2007;39:195–204.

[53] Shichiri M, Yoshinaga K, Hisatomi H, et al. Genetic and epigenetic inactivation of mitotic checkpoint genes hBUB1 and hBUBR1 and their relationship to survival. Cancer Res 2002;62:13–7.
[54] Pinto M, Vieira J, Ribeiro FR, et al. Overexpression of the mitotic checkpoint genes BUB1 and BUBR1 is associated with genomic complexity in clear cell kidney carcinomas. Cell Oncol 2008;30:389–95.
[55] Zhao Q, Bian AP, Zhang Y, et al. Expression of budding uninhibited by benzimidazoles-1 and mitotic arrest deficient-2 in endometrial carcinoma and its significance. Eur J Gynaecol Oncol 2014;35:44–7.
[56] Wang Z, Katsaros D, Shen Y, et al. Biological and clinical significance of MAD2L1 and BUB1, genes frequently appearing in expression signatures for breast cancer prognosis. PLoS One 2015;10:e0136246.
[57] Han JY, Han YK, Park GY, et al. Bub1 is required for maintaining cancer stem cells in breast cancer cell lines. Sci Rep 2015;5: 15993.
[58] Nagato T, Ueda S, Takahara M, et al. Cyclin-dependent kinase 1 and survivin as potential therapeutic targets against nasal natural killer/T-cell lymphoma. Lab Invest 2019;99:612–24.
[59] Li H, Jia Y, Cheng J, et al. LncRNA NCK1-AS1 promotes proliferation and induces cell cycle progression by crosstalk NCK1-AS1/miR-6857/CDK1 pathway. Cell Death Dis 2018;9:198.
[60] Lin W, Xie J, Xu N, et al. Glaucocalyxin A induces G2/M cell cycle arrest and apoptosis through the PI3K/Akt pathway in human bladder cancer cells. Int J Biol Sci 2018;14:418–26.