Received: 1 March 2019 | Revised: 23 July 2019 | Accepted: 29 August 2019

DOI: 10.1002/mgg3.982

ORIGINAL ARTICLE

Molecular Genetics & Genomic Medicine Open Access WILEY

Weighted gene coexpression network analysis identifies a new biomarker of CENPF for prediction disease prognosis and progression in nonmuscle invasive bladder cancer

Jiawei Shi[1] | **Pu Zhang**[1] | **Lilong Liu**[1] | **Xiaobo Min**[2] | **Yajun Xiao**[1]

[1]Department of Urology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]Department of Hepatology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Correspondence
Yajun Xiao, Department of Urology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China.
Email: 15607123366@163.com

Abstract

Background: The dreadful prognosis of nonmuscle invasive bladder cancer mainly results from the delay in recognition of individuals with a high risk of progression. Thus, the emphasis of this work lies in developing valuable biomarkers that is conducive to accurately predicting the progression of NMIBC.

Methods: Microarray data from GSE32894 including 209 NMIBC samples were performed by weighted gene coexpression network analysis (WGCNA), which could find modules of highly correlated genes and relate modules to external sample traits. Besides, we constructed a protein–protein interaction to facilitate screening the hub gene. At last, we used RNA-seq and microarray data and clinical information from ArrayExpress (E-MTAB-4321) and GSE13507 to select and validate the candidate gene.

Results: In current paper, blue module of 13 gene coexpression clusters we identified was selected as the key modules. Seven genes namely: *CDCA8, CENPF, MCM6, MELK, PRC1, STIL,* and *TPX2* have been identified as candidate genes. Notably, among them, only elevated CENPF in NIMBC tissue was closely associated with low progression-free survival (PFS) and overall survival (OS) rate in three datasets and had a large area under receiver operating characteristic (ROC) curve. Finally, *CENPF* was identified as an effective biomarker in NMIBC.

Conclusion: Therefore, our findings submit a new progressive and prognostic molecular marker and therapeutic target for NMIBC. Moreover, these genes that deserve to be further researched may improve the comprehension about the occurrence and development of superficial bladder cancer.

KEYWORDS

biomarkers, *CENPF*, nonmuscle invasive bladder cancer (NMIBC), progression, weighted gene coexpression network analysis (WGCNA)

Jiawei Shi and Pu Zhang contributed equally to this work.

1 | INTRODUCTION

Bladder cancer is the ninth most common cancer in the world, with the 13th most common cause of cancer-associated mortality (Ferlay et al., 2015). Depending on the guidelines of European Association of Urology (EAU) and American Urological Association (AUA), bladder cancer (BCa) can be divided into muscle invasive bladder cancer (MIBC) and nonmuscle invasive bladder cancer (NMIBC)—which are comprised of Ta (noninvasive papillary carcinoma), Tis (carcinoma in situ: "flat tumor"), and T1 (tumor invades subepithelial connective tissue) and represented the majority of primary BCa with approximately 85% (Babjuk et al., 2017; Chang et al., 2016). The prevailing management strategy for NMIBC is complete transurethral resection of the bladder tumor (TURBT) and adjuvant intravesical treatment (Babjuk et al., 2017; Chang et al., 2016). Despite this, it is reported that the progression rate of NIMBC can reach 10%–30%, especially those patients with T1G3 (Cambier et al., 2016). Van den Bosch and Alfred pointed out that two-thirds of the secondary MIBC from progression died to BCa within 48 months even after radical treatment, and cancer-specific survival is significantly dreadful (van den Bosch & Alfred, 2011). Fortunately, European Organization for Research and Treatment of Cancer (EORTC) risk tables were recommended for predicting the progression of NMIBC after TURBT (Babjuk et al., 2017; Chang et al., 2016). However, there are certain deficiencies in these risk tables. Numerous researches in recent years have established that remarkable the molecular heterogeneity has been shown in bladder cancer (Cancer Genome Atlas Research Network, 2014; Choi et al., 2014; Sanchez-Carbayo, Socci, Lozano, Saint, & Cordon-Cardo, 2006; Sjodahl et al., 2012; van Kessel et al., 2018). NMIBC is a heterogeneous set of tumors with thoroughly different oncologic outcomes. Furthermore, variability in staging and grading assessment is a recognized problem (Bol et al., 2003). Unfortunately, EORTC risk tables do not take these factors into account.

Due to rapid technological advancements in the next-generation sequencing, whole-genome and RNA sequencing have been used to research pathological mechanisms and related biomarkers of BCa (Cancer Genome Atlas Research Network, 2014; Choi et al., 2014; Sanchez-Carbayo et al., 2006; Sjodahl et al., 2012). According to the latest research, certain molecular markers can improve the accuracy of predictive progression of EAU risk stratification (van Kessel et al., 2018). Even though molecular profiles of NMIBC subtypes and considerable diagnostic indicators have been investigated as biomarkers of the risk of disease progression (Choi et al., 2014; Ding et al., 2014; Kim et al., 2010; Sanchez-Carbayo et al., 2006; Sjodahl et al., 2012) and play a role in clinical management, none are able to accurately predict the behavior of NMIBC. Hence, identifying molecular biomarkers or therapeutic targets to address those problems is imminent. Unlike most other studies that focus on the screening of differentially expressed genes, this paper explores the affiliation between gene sets and clinical features by WGCNA—a systems biology method and enables the description of the correlation patterns between microarray samples and gene expression profiles (Langfelder & Horvath, 2008). Here, by this method, we find some candidate biomarkers associated with NMIBC progression and prognosis or therapeutic targets.

2 | MATERIALS AND METHODS

2.1 | Ethical compliance and study design

We retrospectively analyzed the gene expression profiles of four public BCa cohorts, therefore ethical approval is not required. This study design was performed in a flow diagram (Figure 1).

2.2 | Data collection

For consistency in the microarray platform, we searched in GEO database (http://www.ncbi.nlm.nih.gov/geo/) for the BCa-related gene expression datasets measured by the Illumina HumanHT-12 V3.0 expression beadchip (Illumina, Inc.) and obtained two independent datasets. For one, non-normalized microarray gene expressing profiles and corresponding clinical data were downloaded from GSE32894 ($n = 213$) (Sjodahl et al., 2012) and used to screen probesets, construct of coexpression networks and identify hub genes in the present study. For another, the processed expression matrix of GSE48075 ($n = 54$) (Choi et al., 2014) was obtained for module preservation analysis. Meanwhile, to verify the analysis results, expression data and related clinicopathologic information were acquired from GSE13507 ($n = 103$) (Kim et al., 2010) and ArrayExpress (https://www.ebi.ac.uk/arrayexpress/) ($n = 460$, E-MTAB-4321) (Hedegaard et al., 2016) which was performed on Illumina human-6 v2.0 expression beadchip (48K) and paired-end RNA-Seq (101 + 7 + 101 bp) analysis on an Illumina HiSeq 2000 (Illumina, Inc.), respectively. In addition, the ONCOMINE database (www.oncomine.org) (Rhodes et al., 2004; Sanchez-Carbayo et al., 2006) were used for further analysis to validating the result of mRNA levels compared with normal tissue. All NMIBC (Ta/Tis/T1, 2009, 7th edition) patients were saved.

2.3 | Data preprocessing and genes screening

Non-normalized data sets from GSE32894 and ArrayExpress (E-MTAB-4321) were performed log2 transformation and standardized by quantile normalization used

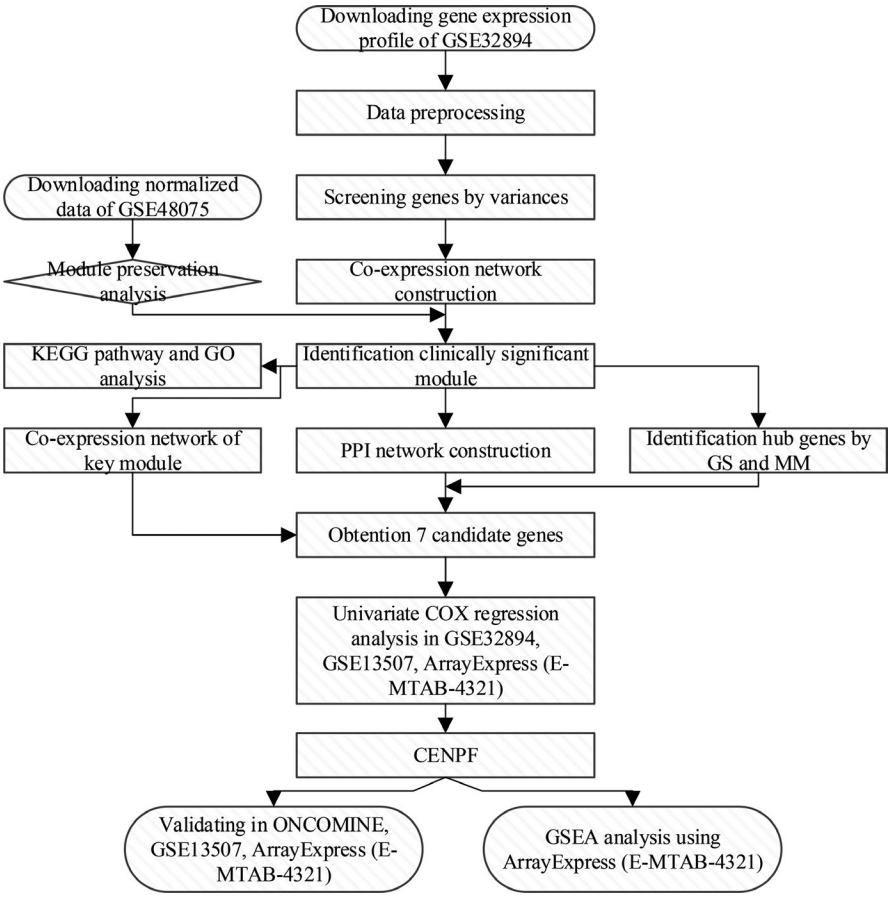

FIGURE 1 Flowchart of data collection, preparation, processing, analysis, and validation in this report

"preprocessCore" package (Bolstad, 2018) in R. Then, the reference genome (GRCh38.86) was used to convert Gene ID in the later dataset. Processed gene expression profiles from other datasets were utilized directly in the further analysis. Furthermore, according to the variance of the probe sets in all samples, the genes relevant to the probes of ranking in the top 20% were chosen for subsequent analysis. All the above operations are conducted on the programming software R and could be found in Appendix Code S1.

2.4 | Weighted coexpression network construction

Following the instructions below, expression matrix of 9,761 probes was used to build coexpression network by the "WGCNA" packages in R (Langfelder & Horvath, 2008). Firstly, a gene coexpression resemblance measure (absolute value of the Pearson correlation) was utilized to estimate for all pairwise gene–gene relationship. Next, a weighted adjacency matrix was constructed using a "soft" power adjacency function $a_{ij} = |cor\ (x_i, x_j)|^{\beta}$. a_{ij} indicates the weighted Pearson's correlation coefficient that measures the connection strength between gene i and gene j. The soft threshold power β, is the lowest integer where the resulting gene coexpression networks satisfy approximate scale-free topology (Zhang & Horvath, 2005). The adjacency matrix was converted to topological overlap matrix (TOM) subsequently, which could evaluate the direct correlation of gene pairs and the degree of agreement in association with other genes in the data set. (Yip & Horvath, 2007). After that, average linkage hierarchical clustering was conducted in accordance with the TOM-based dissimilarity measure. An appropriate minimum gene module size for the gene dendrogram was set to classify similar genes into one module (Ravasz, Somera, Mongru, Oltvai, & Barabasi, 2002). Moreover, module eigengene (ME) that can be regarded as a representative of the gene expression profiles of a module, is defined as the first principal component of an interesting module (Langfelder & Horvath, 2008). Similar modules would be merged by the calculation of ME. More detailed WGCNA method processing can be found in these references (Langfelder & Horvath, 2008; Zhang & Horvath, 2005) and Appendix Code S1.

2.5 | Module preservation analyses

To evaluate whether a module is robust and reproducible across other datasets, a permutation test was carried out by applying the modulePreservation function from the WGCNA packages. According to WGCNA authors (Langfelder & Horvath, 2008; Langfelder, Luo, Oldham, & Horvath, 2011), the module with a Z score below 10 indicates a low

preservation, and is not recommended to use it for the following analysis.

2.6 | Identification of clinically significant modules

As cited previously, ME could summarize the expression patterns of all genes into a single characteristic expression profile within a given module. The researchers calculated the correlation between each ME and clinical traits including age, gender, tumor grade, tumor stage and progression as one of the factors to identify the key module. Under description of the author of the WGCNA package, the equation $GS_i = |cor(x_i, T)|$ was used to quantify the gene i of gene significance (GS) (Langfelder & Horvath, 2008). We refer to x_i as gene i expression profile and T as one of sample trait. Afterwards, module significance (MS), as another factor identified in key modules, is defined as the average absolute GS measure for all genes in a specific module. In general, clinically significant modules were modules with higher correlation between the ME and external traits and bigger MS.

2.7 | Functional annotation and hub genes and candidate genes identification

To further reveal the mechanism underlying of the connection between genes in interesting modules and clinical features, "clusterProfiler" package (version "3.10.1") (Yu, Wang, Han, & He, 2012) in R was implemented to perform GO function annotation and KEGG pathway enrichment analysis on these genes.

It is a robust approach that integration of protein–protein interactions and coexpression networks for gene selection may provide greater insight in disease-related biological process (Dutta, Saha, & Gulati, 2019; Mahapatra, Mandal, & Swarnkar, 2018). Generally, the hub gene is a highly connected central node that is the core of the network architecture and has been proven have major biological functions. For each gene, module membership (MM) was determined by correlating a given gene expression profile with the ME of a specific module. Therefore, the key modules' subnetwork with weighted edges above 0.15 was extracted from the weighted coexpression network to discern hub node which have the intramodular connectivity (also interpreted as a measure of module membership) above 15. Additionally, we constructed the protein–protein interaction (PPI) network of the key module genes based on the Search Tool for the Retrieval of Interacting Genes (STRING) database (http://www.string-db.org/). The genes interaction with a combined score of ≥0.7 was analyzed and visualized by the Cytoscape software (https://cytoscape.org/). Genes with a connectivity degree above 15 were treated as hub genes in PPI network. Usually, Hub genes in the designated module highly correlate with a particular trait tends to imply the genes' high GS and MM. Thus, in the PPI network and the weighted coexpression network, obtained hub genes with $|GS| > 0.2$ and $|MM| > 0.8$ were determined to the candidate genes for further analysis.

2.8 | Candidate genes validation

Patients and clinical covariates for the NMIBC are summarized in Appendix Table S1–S3. In present study, the progression-free survival (PFS) was defined as the time from the date of resection to the date of progression, death or last follow-up. OS was measured from the date of resection to the date of death or last follow-up. To figure out the association of candidate genes expression with PFS time, these genes were performed with univariate and multivariate Cox regression analysis in the GSE32894 Cohort (Sjodahl et al., 2012), ArrayExpress (E-MTAB-4321) Cohort (Hedegaard et al., 2016) and the GSE13507 Cohort (Kim et al., 2010). The analysis was done by packages "survival" (Goel, Khanna, & Kishore, 2010) in R. Among these candidate genes, the mRNA with p-value <0.05 in three data sets' univariate Cox regression analysis was chosen as key genes to subsequent analysis. Kaplan–Meier analysis and the receiver operating characteristics (ROC) analysis were performed by "survminer" package in R-studio and SPSS (SPSS version 22.0, Inc.), respectively. ROC plot was visualized by R package "ggplot2" and all the tests above were done two-sidedly. Notably, the area under the ROC curve (AUC) for each ROC curve represents the ability of the model to predict endpoint events. What is more, a higher AUC means better model performance (with AUC = 0.5 demonstrating random performance). Afterwards, in SPSS, we calculated the Spearman correlation between *CENPF* and clinicopathological features, *MKI67* (OMIM *176741; as a control object), and *FOXM1* (OMIM *602341), respectively, to further verify their association with these features. Only values of $p < .05$ are considered statistically significant, using a two-sided t test. The package "ggcorrplot" in R was used to visualize. Then, the ONCOMINE database (www.oncomine.org) was used to validate the differential expression of the genes. We also investigate the key gene relative expression among subgroup of clinical traits in "ggpbur" package. The expression characteristics were compared between groups using a one-way analysis of variance (ANOVA). All R packages version information was displayed in Appendix Figure S1.

2.9 | Gene set enrichment analysis (GSEA)

In RNA-seq expression profiles of ArrayExpression cohort, the top 30% and the last 30% of the samples, given the problem of noise, were classified into high- and low expression groups according to CENPF relative expression. Afterward, we executed the groups differential expression analysis with

"limma" package in R and construct preranked whole gene list by log2 fold change (Mootha et al., 2003). Next, GSEA (Subramanian et al., 2005) was performed to determine Kyoto encyclopedia of genes and genomes (KEGG) pathways of dysregulated CENPF. In addition, preranked gene list was mapped into MSigDB C6 Oncogenic Signatures gene sets. "clusterProfiler" (Yu et al., 2012) and "ggplot2" package were utilized to analyze and visualize the above. The parameters were adapted to default and the results of p adjust $< .05$ and minimal gene set above 30 considered significantly enriched. The p-value was adjusted using the Benjamini–Hochberg Procedure.

3 | RESULTS

3.1 | Genes screening, weighted coexpression network construction, and module preservation analysis

WGCNA were performed in a total of 9,761 probes (Appendix Table S6) selected by their variance. GSM814108, GSM814148, GSM814236, and GSM814239 were excluded for incomplete clinicopathologic information. GSM814116 and GSM814242 were outlier samples which were distant from other samples in the clustering via the average linkage method. Those samples were removed when constructing a scale-free coexpression network to guarantee the accuracy of the analysis results. A total of 207 samples with related clinical data were incorporated into the weighted coexpression analysis (Appendix Figure S2). The genes with similar expression patterns were placed together with some modules via the average linkage hierarchical clustering. In our research, considering the fitness of the scaleless topology model and the mean connectivity of the network, the power of $\beta = 4$ (scale free $R^2 = 0.91$) was selected as the soft-thresholding (Figure 2a–c). The other parameters as needed were set with minimum module size 30, the module detection sensitivity deepSplit 2, and cut height for merging of modules 0.25 which means that the modules whose eigengenes are correlated above 0.75 will be merged. In the end we got 13 modules (Figure 2d). Furthermore, the result of module preservation analysis on an independent datasets GSE48075 indicate that blue and greenyellow modules preservation statistics Zsummary is far exceeding 10 (Figure 3a). Consequently, there is no evidence showing that the detected modules are not stable in this dataset.

3.2 | Identification of key modules

According to the two approaches described in the method section, the blue and greenyellow module should be identified as clinically significant modules. As it is shown in Figure 3b, tumor grade has the highest correlation coefficient with the ME of blue module which showed the highest MS (Figure 3c), and was also crucially associated with tumor T stage as well as progression. In addition, we also observed that the ME of greenyellow modules demonstrated the secondly high MS (Figure 3c) and the robust correlations with multiple features, such as histological grade, tumor stage, and progression (Figure 3b). Remarkably, although the greenyellow module has a high MS value and a significant correlation between clinical features and ME, their weight values in the coexpression network are too low (weight values of all edge below 0.15). This means that the correlation between gene pairs in the network is not robust, thus, we abandoned the greenyellow module considering the identification of hub genes and only blue module was selected as key ones.

3.3 | Functional enrichment analysis and hub genes and candidate genes identification

Initially, 1934 probe genes in the blue module were mapped into the GO terminology and the KEGG pathway to bring their implied functions and mechanisms to the surface. Genes of blue module were mainly enriched in mitotic processes (Figure 4a), and mapped into the pathways of "Cell cycle," "DNA replication," "Cellular senescence," "p53 signaling pathway," which has been proved to be related to the onset of cancer (Figure 4b). The detailed analysis results are presented in Appendix Table S7 and S8. Later, the blue modules genes were upload to the STRING database (http://www.string-db.org/) to build a PPI network, which could calculate and visualize the intramodular connectivity. In the blue module, centromere protein *F* (*CENPF*, OMIM *600236), cell division cycle associated 8 (*CDCA8*, OMIM *609917), minichromosome maintenance complex component 6 (*MCM6*, OMIM *601806), protein regulator of cytokinesis 1 (*PRC1*, OMIM *603483), the parental embryonic leucine zipper kinase (*MELK*, OMIM *607025), centriolar assembly protein (*STIL*, OMIM *181590), and targeting protein for xenopus kinesin-like protein 2 (*TPX2*, OMIM *605917) were taken as candidate genes after filtering by three formerly described methods (Figure 4c). The detailed information of these genes was summarized in Table 1. However, no genes met the criteria of weighted edges above 0.15 in greenyellow modules' coexpression network (Figure 4d). Lastly, we showed PPI subnetwork composed of first neighbors of seven candidate genes and the correlation between MM and GS of each gene in blue module (Appendix Figure S3 and S4).

3.4 | Candidate genes validation

The validation of *CENPF* was performed on three independent BCa cohorts. At first, considering the accuracy of the analysis, six cases with missing follow-up information were

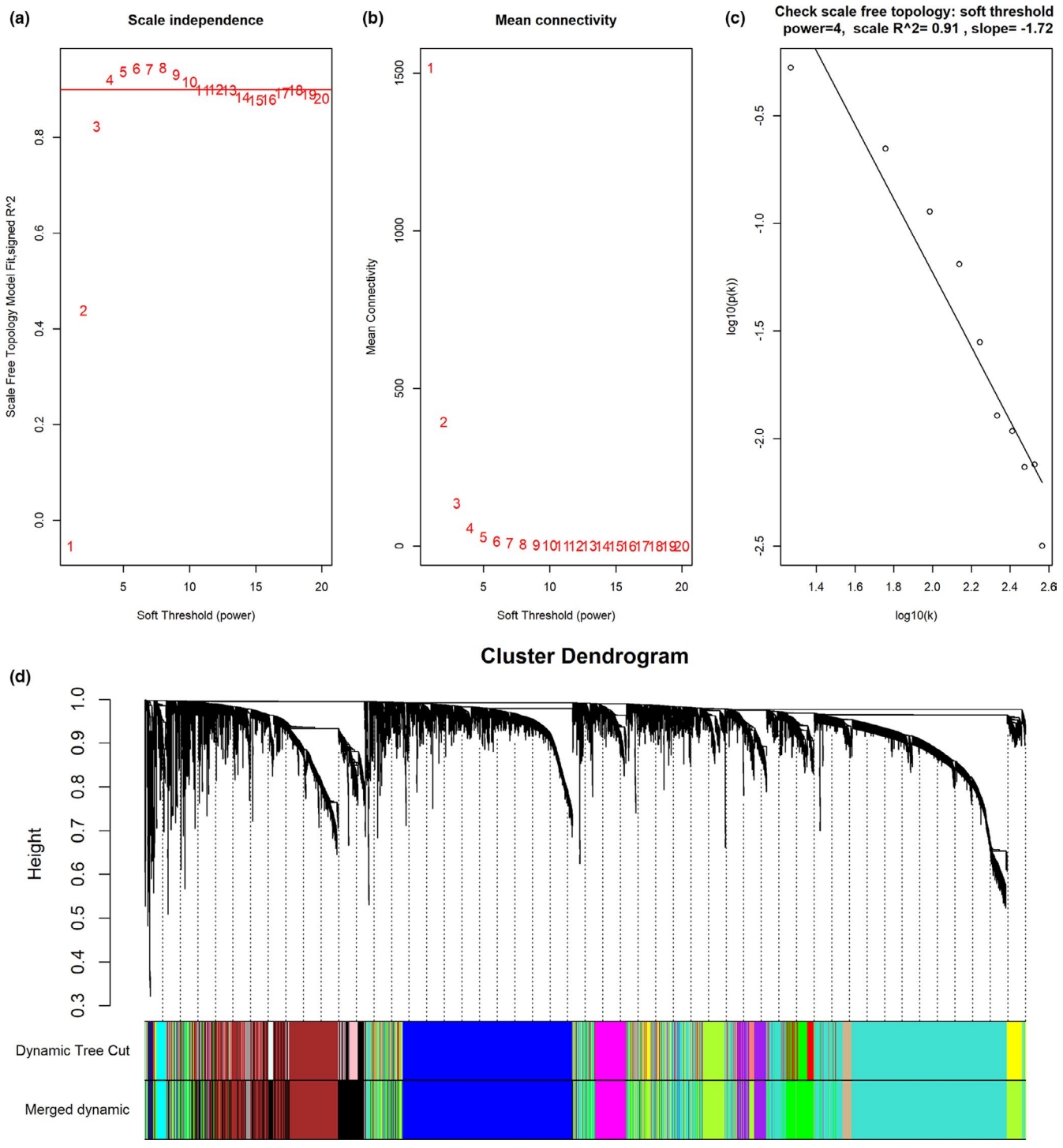

FIGURE 2 (a, b) Determination of soft-thresholding power in WGCNA. According to analysis of the scale-free fit index and the mean connectivity for various soft-thresholding powers (β), β is finally determined to be 4. (c) Checking the scale-free topology when β = 4. (d) Hierarchical cluster analysis dendrogram used to detect coexpression clusters along with corresponding color assignments. In total, 12 modules ranging from 51 to 3,007 probes in size were identified. The 12 probes that were not coexpressed in the data set were assigned to the gray group

removed from the ArrayExpress cohort (E-MTAB-4321) (Hedegaard et al., 2016). The rest of 454 cases were reserved for following analysis. These seven genes in the univariate Cox regression analysis of three data sets, only *CENPF* has been proven to be the most significant factors affecting PFS (Table 2). Therefore, *CENPF* was chosen as a unique candidate for subsequent analysis. On the base of spearman correlation analysis (two-sided test), it is evident that the results obtained here are in pleasurable agreement with the found by the WGCNA analysis, which is a reliable connection

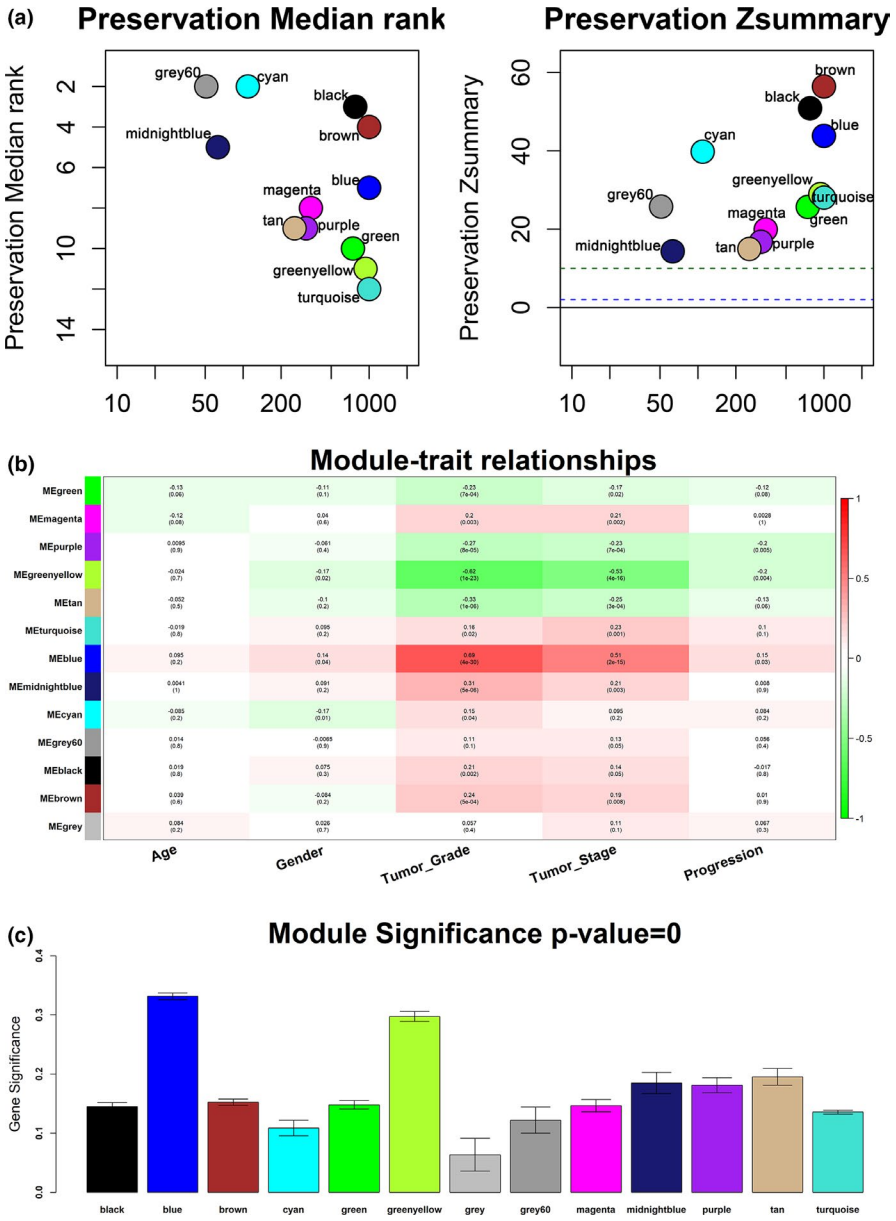

FIGURE 3 (a) The medianRank and Zsummary statistics of the module preservation of the genes modules. The left panel shows the medianRank of the modules, which close to zero indicates a high degree of module preservation. On the right panel, the blue module is above the dashed green lines (Z = 10). (b) Heatmap of the Pearson correlation coefficient between MEs and clinicopathological variables. Each cell contains the corresponding correlation and *p*-value. (c) Bar plots of eigengenes' module significance associated with histological grade of NMIBC across all modules

between *CENPF* and tumor grade, stage and progression (Figure 5a). Soon, we tried to elucidate the expression status of *CENPF* in NMIBC utilizing the ONCOMINE database. In this work, *CENPF* in superficial bladder cancer was significantly higher at mRNA levels compared to normal tissues (Figure 5b-c). As presented in Appendix Figure S5, *CENPF* is not only highly expressed in bladder cancer but also in many other tumors. Data set filters were set as *p*-value <0.001 (independent sample *t* test), and fold changes ≥2, and other parameters are default values. On the other hand, in ArrayExpress cohort, we observed that *CENPF* expression was significantly increased, compared with NMIBC with low grade (PUNLMP and low-grade papillary urothelial carcinoma), low stage (Ta), no progression, and no carcinoma in situ (Figure 6a–d). In the GEO13507 and GEO32894 database (Kim et al., 2010; Sjodahl et al., 2012), the *CENPF* expression of high histological grade or tumor stage BCa are more than that of low grade or tumor stage, which is parallel to the previous experimental data (Figure 6e–h).

Subsequently, a two-sided log-rank test for Kaplan–Meier curve was used to assess the significance of the difference in overall and progression-free survival profiles between different *CENPF* expression levels (the median expression as the cutoff value). Compared with downregulated *CENPF* expression, upregulated expression of *CENPF* is accompanied by poorer progression-free survival (PFS) among these cohorts (*p*-value range 0–.038; Figure 7a–c) and poorer overall survival (OS) in GSE13507 cohort ($p < .038$; Figure 7d). What is more, on the basis of data from the ArrayExpress cohort (E-MTAB-4321), classifier cut by *CENPF* median expression also enables us to distinguish the progression-free survival time between different subgroups of patients sorted

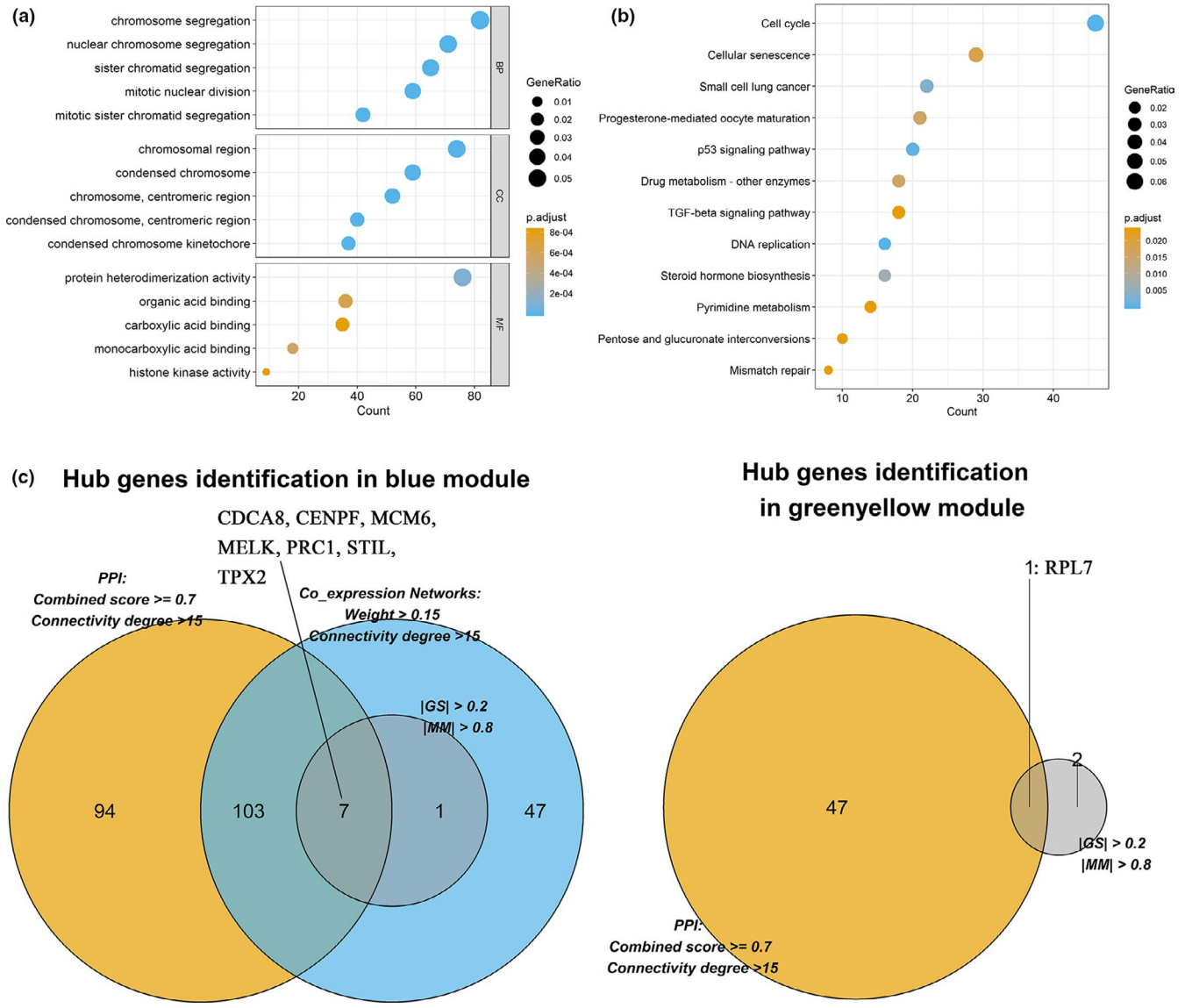

FIGURE 4 (a)The bubble chart of the gene ontology enrichment analysis based on the genes of blue module, showing only the representative results. (b) A bubble chart showing the results of the KEGG pathway enrichment analysis. The size of the bubbles shows the count of the enriched genes, while the color indicates the enrichment significance (*p*-value). In blue (c) and greenyellow (d) modules, number of hub genes identified with PPI network (orange circle), coexpression network (sky blue circle) as well as gene significance and module membership (grey circle); and the overlapping part of the three indicates the number of candidate genes

TABLE 1 The detailed information of candidate genes

Gene symbol	Probe ID	GS T-stage (*p*-value)	GS grade (*p*-value)	GS progression (*p*-value)	MM (*p*-value)	PPI network[a]	Coexpression network[b]
CDCA8	ILMN_1709294	0.40 (2.20E-09)	0.54 (8.80E-17)	0.23 (8.71E-04)	0.81 (6.66E-50)	90	146
CENPF	ILMN_1664516	0.45 (1.42E-11)	0.56 (7.46E-19)	0.20 (3.26E-03)	0.80 (7.43E-48)	85	134
MCM6	ILMN_1798654	0.39 (4.53E-09)	0.56 (7.74E-19)	0.21 (2.49E-03)	0.85 (7.89E-60)	52	121
MELK	ILMN_2212909	0.44 (2.24E-11)	0.57 (2.87E-19)	0.22 (1.17E-03)	0.87 (1.86E-66)	85	201
PRC1	ILMN_1728934	0.40 (1.63E-09)	0.55 (2.87E-19)	0.22 (1.21E-03)	0.82 (1.31E-51)	85	193
STIL	ILMN_2413650	0.48 (4.29E-13)	0.58 (2.87E-20)	0.21 (2.73E-03)	0.89 (1.33E-70)	19	187
TPX2	ILMN_1796949	0.48 (1.73E-13)	0.61 (1.88E-22)	0.20 (3.32E-03)	0.89 (1.47E-70)	84	208

[a]Connectivity degree in protein–protein interaction network,
[b]Connectivity degree in coexpression network

TABLE 2 Univariate analysis of seven candidate genes in three cohorts

	GSE32894 Cohort			ArrayExpress Cohort (E-MTAB-4321)			GSE13507 Cohort		
Variables	**HR**	**95% CI**	***p*-Value**	**HR**	**95% CI**	***p* -Value**	**HR**	**95% CI**	***p* -Value**
CDCA8	4.25	0.91–19.74	.065	5.7	2.19–14.86	**<.001**	1.66	0.87–3.18	.125
CENPF	4.42	1.95–20.53	**.048**	10.03	3.05–32.99	**<.001**	2.46	1.24–4.86	**.01**
MCM6	2.41	0.64–9.16	.195	5.72	2.19–14.89	**<.001**	1.13	0.6–2.13	.712
MELK	4.04	0.86–18.9	.076	4.51	1.85–11	**.001**	2.15	1.1–4.22	**.025**
PRC1	2.44	0.64–9.3	.19	5.53	2.12–14.41	**<.001**	1.97	1.02–3.81	**.043**
STIL	1.47	0.42–5.09	.544	5.58	2.14–14.54	**<.001**	0.92	0.49–1.75	.805
TPX2	5.03	1.08–23.34	**.039**	7.31	2.56–20.89	**<.001**	1.55	0.81–2.96	.188

Notes: A value of $p < .05$ is bolded and indicates statistical significance.
Abbreviation: 95% CI, 95% confidence intervals; HR, hazard ratio.

through the four common molecular taxonomy including CLASS assignment, BASE47 signature, CIS signature, and 12 gene signatures (All of them $p < .001$, Appendix Figure S6). Meanwhile, the individual ability of *CENPF* and clinical covariates (age, gender, T stage, grade) to predict PFS is shown in Appendix Table S4–S5, which give hazard ratios (HRs) and 95% confidence intervals (CIs) calculated by the univariate and multivariate Cox regression analyses. Because the GSE32894 cohort uses the different WHO classification of bladder tumors (1973 version) than the other two datasets, we analyzed GSE32894 cohort alone by the approaches mentioned above, and combined the data from two other datasets which used the same 2004 World Health Organization (WHO) classification of bladder tumors. In addition, the other two data sets were combined and analyzed in view of the fewer endpoint events in the GSE13507 cohort, which enhanced the confidence of cox regression analysis. In our study, the covariates of $p < .2$ in univariate Cox regression analysis were adopted by Cox proportional-hazard models. The univariate analysis result of GEO32894 cohort indicates elevated *CENPF* expression (HR = 4.42, 95% CI = 1.95–20.53, $p = .048$) instead of high grade ($p = .972$) or T stage ($p = .182$) was predictive of unfavorable PFS. However, interestingly, high *CENPF* expression (HR = 7.53, 95% CI = 1.58–35.96, $p = .011$) and T stage (HR = 5.53, 95% CI = 1.16–26.36, $p = .032$) perfectly predicted an adverse outcome when they are evaluated by multivariate analysis (Appendix Table S4), which illustrate that *CENPF* remained an independent risk factor for PFS. In two other cohorts combined analysis, as we expected, high *CENPF* expression (HR = 8.38, 95% CI = 3.28–21.41, $p < .001$) and T stage (HR = 2.11, 95% CI = 2.11–6.74, $p < .001$) are risk factors for PFS in univariate analysis and a similar trend was seen for in multivariate analysis (Appendix Table S5).

Eventually, the AUC of *CENPF* was as high as 0.682 ($p < .05$), 0. 819 ($p < .05$), 0.622 ($p = .187$) respectively in GEO32894 cohort, ArrayExpress cohort, and GEO13507 cohort (Hedegaard et al., 2016; Kim et al., 2010; Sjodahl et al., 2012) (Figure 7e). Random error caused by fewer endpoint events may allow for a relatively low AUC in both GEO data sets. In short, *CENPF* can be used as an independent predictor of PFS and has an excellent predictive performance.

3.5 | Gene set enrichment analysis (GSEA)

To obtain the potential function of *CENPF*, GSEA was conducted to search KEGG pathways which is enriched with highly expressed samples. A total of 14 functional genomes (Appendix Table S9) meet the above criteria—most of which was strongly related with cancer. Three representative pathways including "cell cycle," "Cellular senescence," "MicroRNAs in cancer" were shown in Figure 7f. In addition, we performed GSEA to explore which pathways in oncogenic signatures gene sets was activated or repressed due to the deregulation of *CENPF*. Twenty-one and three pathways (Appendix Table S10) were activated and suppressed respectively in high *CENPF* expression group. A host of them are playing an essential role in cell cycle and the onset and development of BCa, including "E2F1_UP.V1_UP," "RB_P107_DN.V1_UP," "RB_P130_DN.V1_UP," "E2F3_UP.V1_UP," and "CYCLIN_D1_KE_.V1_DN" (Figure 7g).

4 | DISCUSSION

There has long been a major clinical challenge in the prediction in term of disease progression for patients with superficial BCa. In our study, the researchers used some public data sets to identify a pleasing biomarker, *CENPF*, to predict NMIBC progression.

The WGCNA algorithm has been proven in applications ranging from oncology (Chen et al., 2018), neurology (Voineagu et al., 2011) etc. In this report, we applied this classic and reliable algorithm to microarray gene expression data obtained from NMIBC patients, to present a global

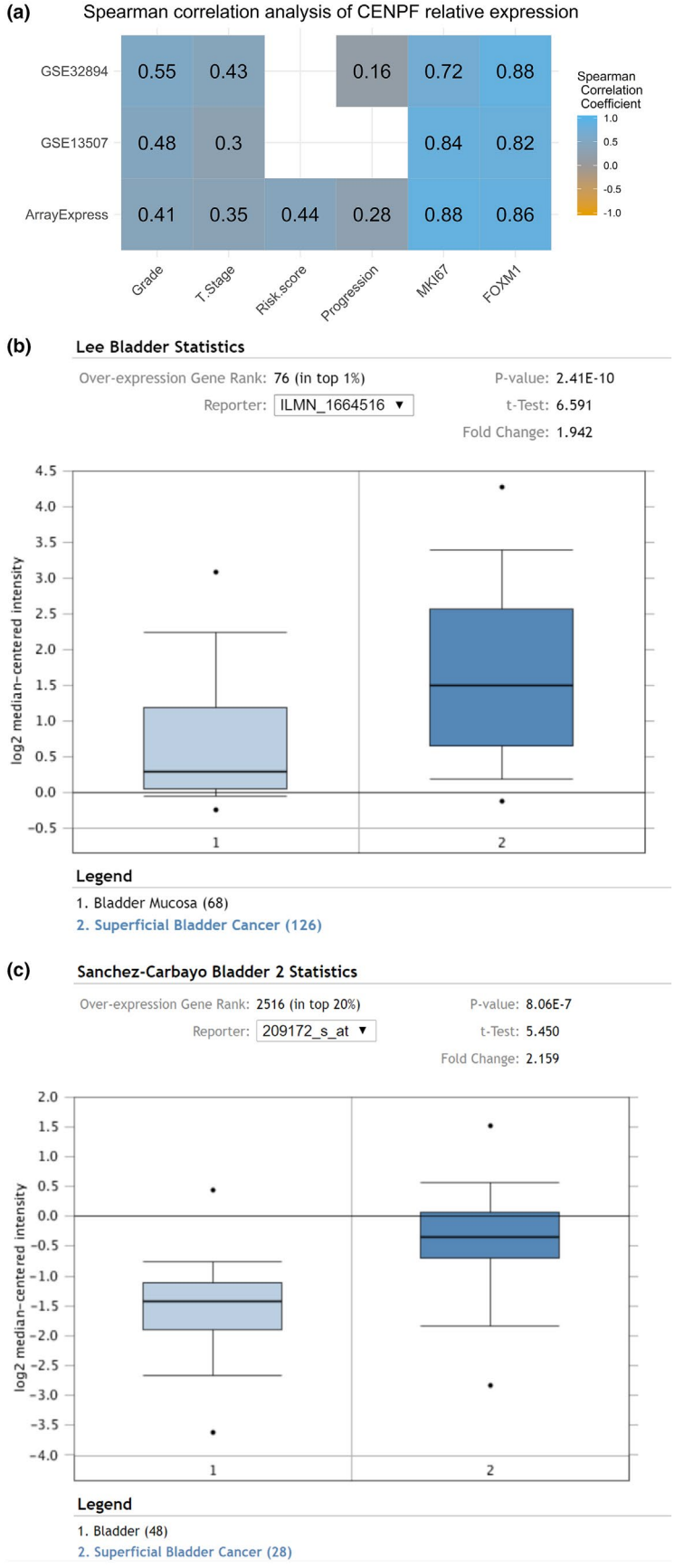

FIGURE 5 (a) The heatmap of the spearman correlation between CENPF expression and clinical features and two genes (MKI67 and FOXM1) in three data sets. Using a two-sided test, only $p < .05$ was shown. (b, c) The box plot of differential CENPF mRNA expression compared normal tissues with superficial bladder cancer in two independent datasets based on Oncomine database

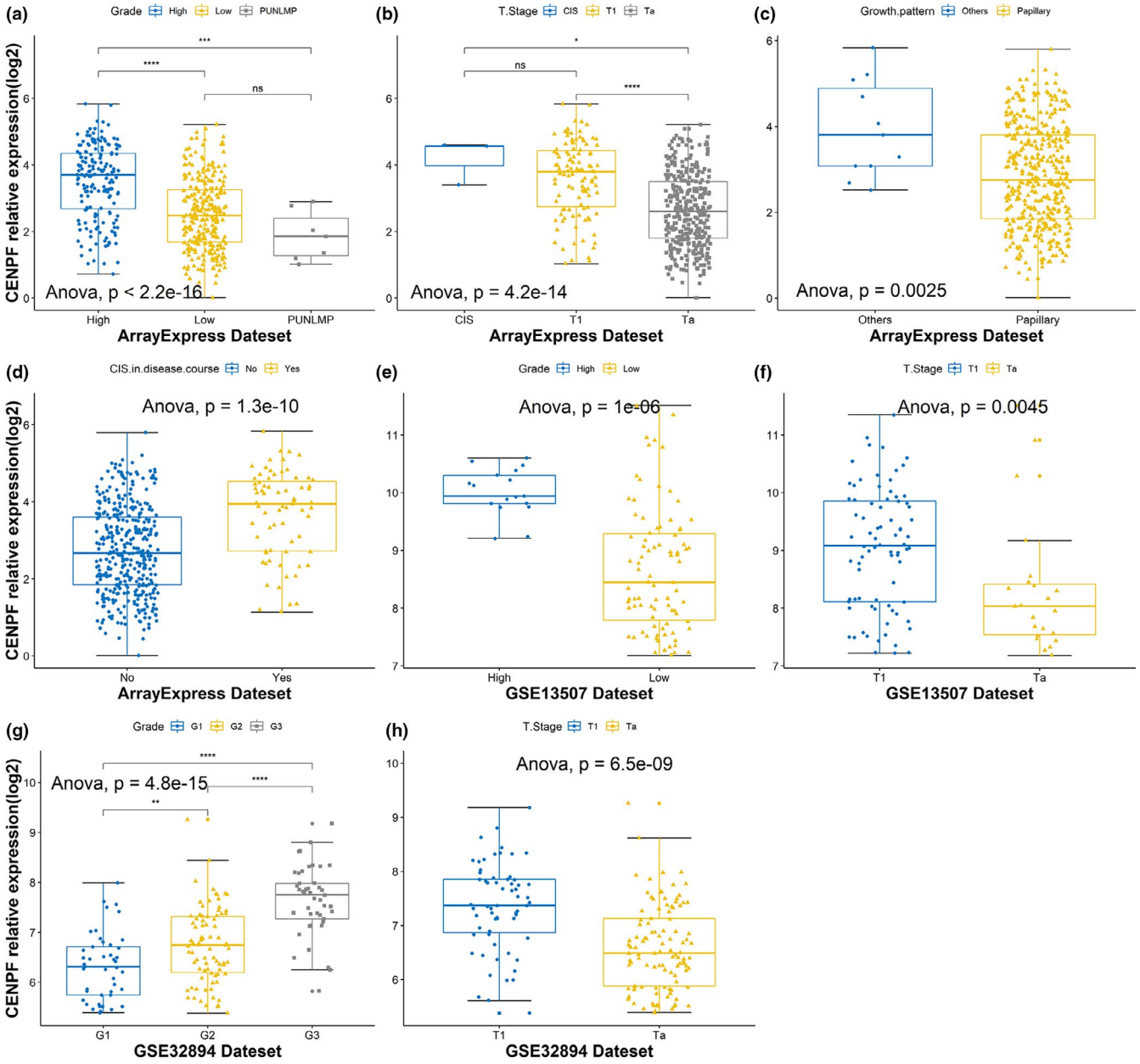

FIGURE 6 Validation of CENPF. The box plot of differential CENPF mRNA expression across distinctive subgroup of bladder cancer in three independent datasets including the ArrayExpress cohort (a–d), GEO32894 (e–f), and GEO13507 (g–h) database

interpretation of clinicopathologic characteristics and mRNA expression profiles. As recommended by the WGCNA author, we screened some probesets by variance rather than differential expression (Langfelder & Horvath, 2008). From 9,761 of the inconsistent probesets across the NMIBC sample set, we identified 13 distinct coexpression modules (ranging in size from 51 to 3,077 probesets) and only blue gene clusters as key ones. A total of seven candidate genes was selected as biomarkers which are able to significantly distinguish distinct NMIBC conditions considering the connectivity degrees of each genes in PPI network and coexpression network and their MM (the Pearson's correlation between gene and ME) and GS (the Pearson's correlation between gene and clinical traits). Then, functional annotation analysis of blue module genes indicated that they are inextricably linked to cell cycle and mitosis while the KEGG analysis results showed that they are characterized in some classical tumor pathways, including the p53 signaling pathway and TGF-beta signaling pathway, etc. Commonly, the escape from G1 checkpoint control, loss of DNA damage checkpoints (G2/M), and loss of normal DNA repair mechanisms are important ways for tumor cells to achieve immortalization. In human BCa, there are frequent events of mutations and dysregulation of cell cycle regulators (Cancer Genome Atlas Research Network, 2014; Sjodahl et al., 2012). Previous investigation also correlated cell cycle regulators (*TP53,* OMIM *191170; *RB1,*

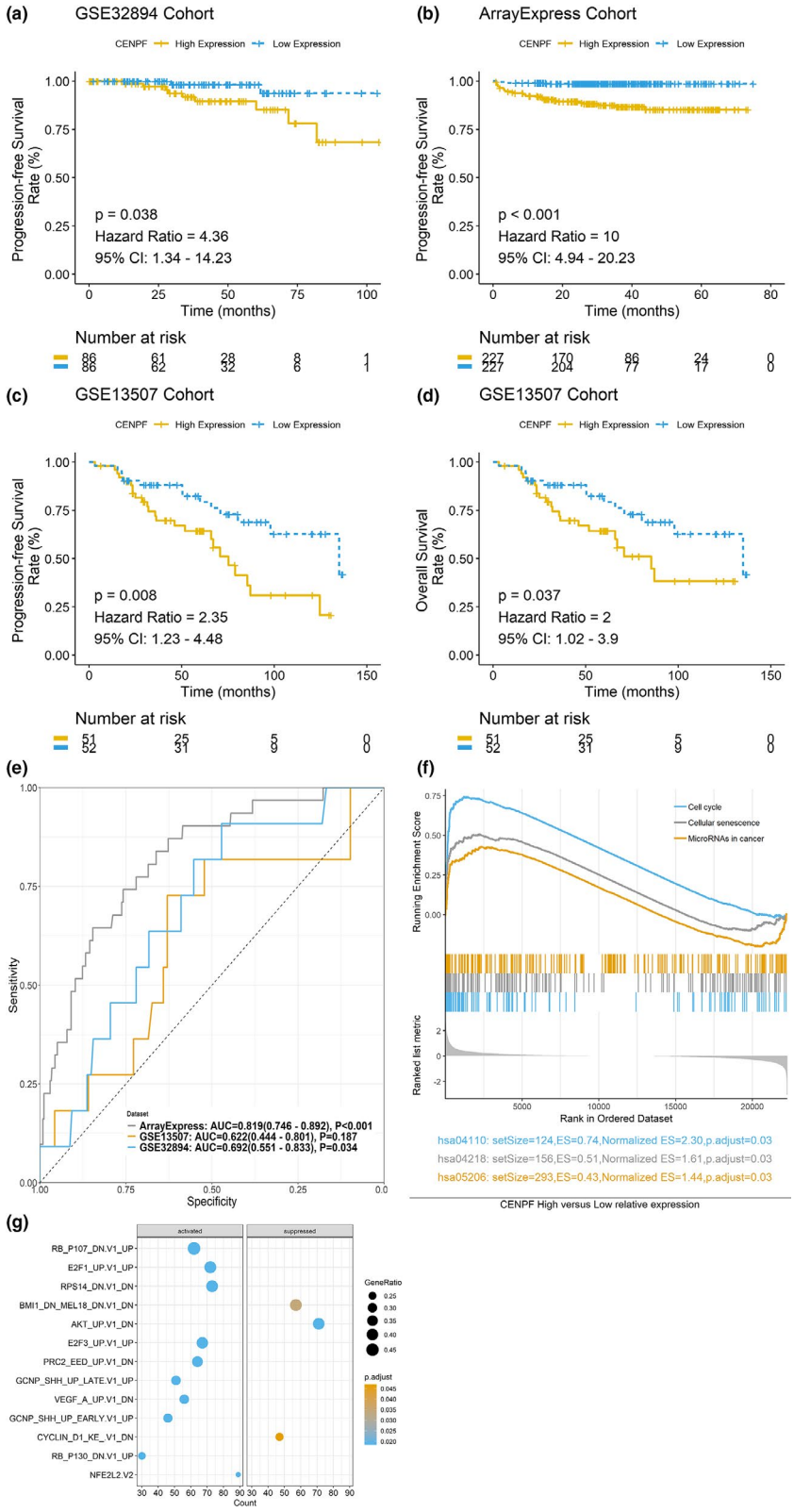

FIGURE 7 (a, b) The progression-free survival plot of GEO32894 and ArrayExpress cohort. (c–d) The progression-free survival and overall survival plot of GEO13507 cohort. (e) Receiving operating characteristic (ROC) curve of the CENPF with prediction of NMIBC progression. Summary graphs of representative gene set enrichment analysis (GSEA) plots of high- versus low CENPF expression analysis from gene sets derived from the MSigDB C2 KEGG (f) and MSigDB C6 Oncogenic Signatures (g), on basis of ArrayExpress cohort

OMIM *614041; *CCND1*, OMIM *168461; *MDM2*, OMIM *164785; *MKI67*, etc.) alterations to higher grade and stage of tumor factors which are known to associated with human urothelial carcinomas progression (Ding et al., 2014; Yurakh et al., 2006). Obviously, these studies strongly confirm our results. Overall, these blue module genes play a complex role in the development and progression of BCa undoubtedly.

As already indicated, we finally got seven candidate genes, which are *CENPF*, *CDCA8*, *MCM6*, *PRC1*, *MELK*, *STIL*, and *TPX2*. Among the seven candidate genes, only

CENPF shows a significant correlation with PFS (progression-free survival) in univariate cox regression analysis of three BCa data sets (Hedegaard et al., 2016; Kim et al., 2010; Sjodahl et al., 2012). As a fundamental member of the centromere protein family including centromere protein A (*CENPA,* OMIM *117139), *CENPB* (OMIM *117140), *CENPC* (OMIM *117141), *CENPH* (OMIM *605607) etc., *CENPF* with 400-kDa which localizes to the outer surface of the outer kinetochore plate (a part of centromere), was identified in the serum of patients with systemic diseases in 1993 (Rattner, Rao, Fritzler, Valencia, & Yen, 1993). Related studies (Lin et al., 2016; O'Brien et al., 2007) have confirmed that *CENPF* acted as a common cancer-driver gene in a variety of tumors such as breast cancer, prostate cancer, and hepatocellular carcinoma. Interestingly, consistent with Appendix Figure S5, *CENPF* expression is upregulated in these tumors. Moreover, patients with elevated *CENPF* expression often tend to have higher histological grades, tumor stage, nonpapillary tumor, and the presence of CIS during the disease course. Aside from a favorable positive correlation with these factors, it also showed a strong correlation with the expression of *FOXM1* and *MKI67* (Figure 5a). Tsai MJ, et al elucidated dysregulation of miRNAs-COUP-TFII-FOXM1-CENPF axis which plays a crucial part in drug resistance and the metastasis of prostate cancer (Lin et al., 2016). Also, *MKI-67* is an established indicator for cell proliferation of BCa (Ding et al., 2014). Usually, abnormal isolation of the chromosome will cause the generation of chromosomal diseases and even cause genomic instability, including activation of oncogenes or inactivation of tumor suppressor genes, which can lead to tumorigenesis. These may explain those results. Our results of the subsequent Kaplan–Meier analysis and univariate and multivariate analysis with clinical pathological factors are consistent with previous ones, that is, high expression of *CENPF* means poor PFS and OS. Strikingly, *CENPF* has a more excellent performance than tumor grade and T stage in term of prediction disease progression. These results highlight the usefulness of *CENPF* as biomarker of predicting for NMIBC progression and adverse outcome. This marker may be beneficial to reduce variability in the staging and grading assessment and improve the accuracy of EORTC risk tables predictions for progression. In addition, GSEA were performed in an independent RNA-seq data set and their result showed that cell cycle-related pathways are activated or inhibited when *CENPF* is aberrantly expressed. Therefore, we could be inferred that elevated *CENPF* leads to the development of bladder cancer through genes such as *TP53, RB1, CCND1,* and E2F gene family in the cell cycle signaling pathway.

Meanwhile, *CENPF* is a large multiprotein complex with a CAAX (C = cysteine, A = aliphatic, and X = any amino acid) domain, and proteins containing such domains are substrates for protein farnesylation which is a lipid posttranslational modification required for the cancer-causing activity of proteins and critical for progression but not initiation of tumorigenesis (Sebti, 2005). It seems that the modification can account for the impact of high expression *CENPF* to unfavorable PFS. Farnesyltransferase inhibitors (FTIs) were first studied for targeting the RAS oncogenes, which can also target *CENPF* resulting in its inactivation. FTIs show potent cytotoxicity as a single agent in preclinical studies and have shown clinical promise in combination with other therapeutic strategies. (Schafer-Hales et al., 2007) Although there is restricted understanding of the specific antitumor mechanisms of FTIs, it provides a new perspective for our treatment of BCa. Therefore, *CENPF* may be a significantly clinical target in BCa and *CENPF* farnesylation a useful marker of tumor response. Notably, in recent data, *CENPF* has been shown to be an independent predictor for OS survival in bladder cancer (Li et al., 2017). However, our study aims to predict disease progression in heterogeneous NMIBC patients, which provides a prospect for management and comprehensive understanding of bladder cancer.

Additionally, among other seven candidate mRNAs, some have turned out to be related to the onset and development of BCa. *CDCA8* is an indispensable regulator of mitosis and cell division. It is reported that its aberrant expression is linked to a poor prognosis for BCa (Bi et al., 2018) and lung carcinogenesis (Hayama et al., 2007). As one of the minichromosome maintenance proteins, *MCM6* are plays a significant role in the initiation of eukaryotic genome replication and associated with histological grades in various neoplastic processes (Gauchotte et al., 2012). Prior studies have evaluated *MCM6* could be the most efficient marker to detect preclinical early recurrence in meningiomas (Gauchotte et al., 2012) and hepatocellular carcinoma (Liu et al., 2018) and so on. *PRC1* is involved in multiple processes of cytokinesis, and its expression varies with the cell cycle. Nakamura, et al. study suggested that the knockdown of its expression induced a significant enrichment of multinuclear cells and succeeding cell death of BCa cells, which provides a new insight of therapeutic targets for BCa (Kanehira et al., 2007). Some papers (Chen et al., 2016; Zhang et al., 2017) have implied that *PRC1* caused hepatocellular and gastric carcinoma exerts oncogenic function via the Wnt/β-catenin signaling pathway and p53-dependent manner, respectively. *MELK* is a highly conserved serine/threonine kinase expressed in several human cancer and stem cell populations, and associated with cell cycle control, cell proliferation, apoptosis, migration, cell renewal, embryogenesis, tumorigenesis and cancer treatment resistance and relapse (Ganguly et al., 2015). There are data indicating that it plays a prominent role in breast, gastric, and lung cancer. (Gray et al., 2005) Like *CENPF, STIL* belongs to centromeric proteins and is required for

mitotic entry and cancer cell survival. It increased expression in multiple types of cancers and correlates with metastatic spread. (Erez et al., 2007) *TPX2* is a critical factor for mitosis and spindle assembly. Overexpressed in cancers, it is being established as biomarker for the diagnosis and prognosis of carcinomas. Knockdown of *TPX2* inhibits the proliferative capacity of hepatocellular carcinoma cells and pancreatic cancer cell lines, and can also induce caspase-3 mediated apoptosis in various cancer cell lines including HeLa, H1299 (lung cancer), DLD-1 (Colon cancer), and MDA-468 (breast cancer) (Neumayer, Belzil, Gruss, & Nguyen, 2014). Overall, these candidate genes are germane to the cell cycle, mitosis, and protein translation processes, and their abnormal expression contributes to the tumorigenesis and progression of certain types of cancers.

However, some limitations of the present study cannot be ignored. First, limited by the dataset we used, our analysis failed to incorporate all the known risk factors in BCa. Second, the follow-up time for the ArrayExpress cohort ranged from 0.9 to 74.9 months (the median follow-up time was 33.1 months) and GSE32894 cohort ranging from 0.2 to 104.4 months (the median follow-up time was 38.1 months), some patients in those cohort were too short to observe endpoint events of progression. Third, a total of 739 NMIBC patients in the three datasets were included in the study, yet substantial clinical cases were needed to assess the biological relevance of the findings. Fourth, some of the differences in our analysis can be explained by their different quantitative platforms (macroarray and RNA-seq). And before performing multi-platform data analysis, we converted continuous variables of expressions into ordinal categorical variables to reduce the discrepancies in multi-platform. Lastly, further research was required to illuminate the specific carcinogenic mechanisms of these genes.

In summary, we have identified 13 gene coexpression clusters from a large-scale NMIBC data set using WGCNA, which is a reliable method to screen and investigate candidate biomarkers or therapeutic targets. Then, we associated these clusters with clinical pathology variables and identify and validate candidate genes, which correlated with tumor stage, histological grade and progression of NMIBC. At last, *CENPF* was identified as a highly prognostic marker for NMIBC disease progression. It is worth mentioning that the biomarker was targeted by FTIs, further study of which may contribute to personalized therapy of NMIBC or MIBC.

ACKNOWLEDGMENTS

We thank Jimmy and the Biotrainee Forum founded by him for providing technical support. We thank the participants described in this report. We also thank the editors of the magazine and the anonymous reviewers for their help in this review.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Jiawei Shi https://orcid.org/0000-0002-1031-3276
Yajun Xiao https://orcid.org/0000-0003-0795-8273

REFERENCE

Babjuk, M., Böhle, A., Burger, M., Capoun, O., Cohen, D., Compérat, E. M., … Zigeuner, R. (2017). EAU guidelines on non-muscle-invasive urothelial carcinoma of the bladder: Update 2016. *European Urology*, *71*(3), 447–461. https://doi.org/10.1016/j.eururo.2016.05.041

Bi, Y., Chen, S., Jiang, J., Yao, J., Wang, G., Zhou, Q., & Li, S. (2018). CDCA8 expression and its clinical relevance in patients with bladder cancer. *Medicine (Baltimore)*, *97*(34), e11899. https://doi.org/10.1097/MD.0000000000011899

Bol, M. G., Baak, J. P., Buhr-wildhagen, S., Kruse, A.-J., Kjellevold, K. H., Janssen, E. A., … Øgreid, P. (2003). Reproducibility and prognostic variability of grade and lamina propria invasion in stages Ta, T1 urothelial carcinoma of the bladder. *Journal of Urology*, *169*(4), 1291–1294. https://doi.org/10.1097/01.ju.0000055471.78783.ae

Bolstad, B. (2018). preprocessCore: A collection of pre-processing functions. R package version 1.44.0, Bioconductor. https://doi.org/10.18129/B9.bioc.preprocessCore

Cambier, S., Sylvester, R. J., Collette, L., Gontero, P., Brausi, M. A., van Andel, G., … Oddens, J. (2016). EORTC nomograms and risk groups for predicting recurrence, progression, and disease-specific and overall survival in non-muscle-invasive stage Ta-T1 urothelial bladder cancer patients treated with 1–3 years of maintenance bacillus calmette-guerin. *European Urology*, *69*(1), 60–69. https://doi.org/10.1016/j.eururo.2015.06.045

Cancer Genome Atlas Research Network (2014). Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*, *507*(7492), 315–322. https://doi.org/10.1038/nature12965

Chang, S. S., Boorjian, S. A., Chou, R., Clark, P. E., Daneshmand, S., Konety, B. R., … McKiernan, J. M. (2016). Diagnosis and treatment of non-muscle invasive bladder cancer: AUA/SUO Guideline. *Journal of Urology*, *196*(4), 1021–1029. https://doi.org/10.1016/j.juro.2016.06.049

Chen, J., Rajasekaran, M., Xia, H., Zhang, X., Kong, S. N., Sekar, K., … Hui, K. M. (2016). The microtubule-associated protein PRC1 promotes early recurrence of hepatocellular carcinoma in association with the Wnt/beta-catenin signalling pathway. *Gut*, *65*(9), 1522–1534. https://doi.org/10.1136/gutjnl-2015-310625

Chen, L., Yuan, L., Qian, K., Qian, G., Zhu, Y., Wu, C.-L., … Wang, X. (2018). Identification of biomarkers associated with pathological stage and prognosis of clear cell renal cell carcinoma by co-expression network analysis. *Frontiers in Physiology*, *9*, 399. https://doi.org/10.3389/fphys.2018.00399

Choi, W., Porten, S., Kim, S., Willis, D., Plimack, E. R., Hoffman-Censits, J., … McConkey, D. J. (2014). Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell*, *25*(2), 152–165. https://doi.org/10.1016/j.ccr.2014.01.009

Ding, W., Gou, Y., Sun, C., Xia, G., Wang, H., Chen, Z., … Qiang, D. (2014). Ki-67 is an independent indicator in non-muscle invasive bladder cancer (NMIBC); combination of EORTC risk scores and Ki-67 expression could improve the risk stratification of NMIBC. *Urologic Oncology: Seminars and Original Investigations*, *32*(1), 42.e13–42.e19. https://doi.org/10.1016/j.urolonc.2013.05.004

Dutta, P., Saha, S., & Gulati, S. (2019). Graph-based hub gene selection technique using protein interaction information: application to sample classification. *IEEE Journal of Biomedical and Health Informatics*, https://doi.org/10.1109/JBHI.2019.2894374

Erez, A., Castiel, A., Trakhtenbrot, L., Perelman, M., Rosenthal, E., Goldstein, I., … Izraeli, S. (2007). The SIL gene is essential for mitotic entry and survival of cancer cells. *Cancer Research*, *67*(9), 4022–4027. https://doi.org/10.1158/0008-5472.CAN-07-0064

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., … Bray, F. (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer*, *136*(5), E359–386. https://doi.org/10.1002/ijc.29210

Ganguly, R., Mohyeldin, A., Thiel, J., Kornblum, H. I., Beullens, M., & Nakano, I. (2015). MELK-a conserved kinase: Functions, signaling, cancer, and controversy. *Clin Transl Med*, *4*, 11. https://doi.org/10.1186/s40169-014-0045-y

Gauchotte, G., Vigouroux, C., Rech, F., Battaglia-Hsu, S.-F., Soudant, M., Pinelli, C., … Bressenot, A. (2012). Expression of minichromosome maintenance MCM6 protein in meningiomas is strongly correlated with histologic grade and clinical outcome. *American Journal of Surgical Pathology*, *36*(2), 283–291. https://doi.org/10.1097/PAS.0b013e318235ee03

Goel, M. K., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. *International Journal of Ayurveda Research*, *1*(4), 274–278. https://doi.org/10.4103/0974-7788.76794

Gray, D., Jubb, A. M., Hogue, D., Dowd, P., Kljavin, N., Yi, S., … Davis, D. P. (2005). Maternal embryonic leucine zipper kinase/murine protein serine-threonine kinase 38 is a promising therapeutic target for multiple cancers. *Cancer Research*, *65*(21), 9751–9761. https://doi.org/10.1158/0008-5472.CAN-04-4531

Hayama, S., Daigo, Y., Yamabuki, T., Hirata, D., Kato, T., Miyamoto, M., … Nakamura, Y. (2007). Phosphorylation and activation of cell division cycle associated 8 by aurora kinase B plays a significant role in human lung carcinogenesis. *Cancer Research*, *67*(9), 4113–4122. https://doi.org/10.1158/0008-5472.CAN-06-4705

Hedegaard, J., Lamy, P., Nordentoft, I., Algaba, F., Høyer, S., Ulhøi, B. P., … Dyrskjøt, L. (2016). Comprehensive transcriptional analysis of early-stage urothelial carcinoma. *Cancer Cell*, *30*(1), 27–42. https://doi.org/10.1016/j.ccell.2016.05.004

Kanehira, M., Katagiri, T., Shimo, A., Takata, R., Shuin, T., Miki, T., … Nakamura, Y. (2007). Oncogenic role of MPHOSPH1, a cancer-testis antigen specific to human bladder cancer. *Cancer Research*, *67*(7), 3276–3285. https://doi.org/10.1158/0008-5472.CAN-06-3748

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. https://doi.org/10.1186/1476-4598-9-3

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*, 559. https://doi.org/10.1186/1471-2105-9-559

Langfelder, P., Luo, R., Oldham, M. C., & Horvath, S. (2011). Is my network module preserved and reproducible? *PLoS Computational Biology*, *7*(1), e1001057. https://doi.org/10.1371/journal.pcbi.1001057

Li, S., Liu, X., Liu, T., Meng, X., Yin, X., Fang, C., … Wang, X. (2017). Identification of biomarkers correlated with the TNM staging and overall survival of patients with bladder cancer. *Frontiers in Physiology*, *8*, 947. https://doi.org/10.3389/fphys.2017.00947

Lin, S.-C., Kao, C.-Y., Lee, H.-J., Creighton, C. J., Ittmann, M. M., Tsai, S.-J., … Tsai, M.-J. (2016). Dysregulation of miRNAs-COUP-TFII-FOXM1-CENPF axis contributes to the metastasis of prostate cancer. *Nature Communications*, *7*, 11418. https://doi.org/10.1038/ncomms11418

Liu, M., Hu, Q., Tu, M., Wang, X., Yang, Z., Yang, G., & Luo, R. (2018). MCM6 promotes metastasis of hepatocellular carcinoma via MEK/ERK pathway and serves as a novel serum biomarker for early recurrence. *Journal of Experimental and Clinical Cancer Research*, *37*(1), 10. https://doi.org/10.1186/s13046-017-0669-z

Mahapatra, S., Mandal, B., & Swarnkar, T. (2018). Biological networks integration based on dense module identification for gene prioritization from microarray data. *Gene Reports*, *12*, 276–288. https://doi.org/10.1016/j.genrep.2018.07.008

Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., … Groop, L. C. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics*, *34*(3), 267–273. https://doi.org/10.1038/ng1180

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. https://doi.org/10.1007/s00018-014-1582-7

O'Brien, S. L., Fagan, A., Fox, E. J. P., Millikan, R. C., Culhane, A. C., Brennan, D. J., … Gallagher, W. M. (2007). CENP-F expression is associated with poor prognosis and chromosomal instability in patients with primary breast cancer. *International Journal of Cancer*, *120*(7), 1434–1443. https://doi.org/10.1002/ijc.22413

Rattner, J. B., Rao, A., Fritzler, M. J., Valencia, D. W., & Yen, T. J. (1993). CENP-F is a.ca 400 kDa kinetochore protein that exhibits a cell-cycle dependent localization. *Cell Motility and the Cytoskeleton*, *26*(3), 214–226. https://doi.org/10.1002/cm.970260305

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabasi, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science*, *297*(5586), 1551–1555. https://doi.org/10.1126/science.1073374

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. https://doi.org/10.1016/s1476-5586(04)80047-2

Sanchez-Carbayo, M., Socci, N. D., Lozano, J., Saint, F., & Cordon-Cardo, C. (2006). Defining molecular profiles of poor outcome in patients with invasive bladder cancer using oligonucleotide microarrays. *Journal of Clinical Oncology*, *24*(5), 778–789. https://doi.org/10.1200/JCO.2005.03.2375

Schafer-Hales, K., Iaconelli, J., Snyder, J. P., Prussia, A., Nettles, J. H., El-Naggar, A., … Marcus, A. I. (2007). Farnesyl transferase inhibitors impair chromosomal maintenance in cell lines and human tumors by compromising CENP-E and CENP-F function.

Molecular Cancer Therapeutics, *6*(4), 1317–1328. https://doi.org/10.1158/1535-7163.MCT-06-0703

Sebti, S. M. (2005). Protein farnesylation: Implications for normal physiology, malignant transformation, and cancer therapy. *Cancer Cell*, *7*(4), 297–300. https://doi.org/10.1016/j.ccr.2005.04.005

Sjodahl, G., Lauss, M., Lovgren, K., Chebil, G., Gudjonsson, S., Veerla, S., … Hoglund, M. (2012). A molecular taxonomy for urothelial carcinoma. *Clinical Cancer Research*, *18*(12), 3377–3386. https://doi.org/10.1158/1078-0432.CCR-12-0077-T

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 of the United States of America*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

van den Bosch, S., & Alfred Witjes, J. (2011). Long-term cancer-specific survival in patients with high-risk, non-muscle-invasive bladder cancer and tumour progression: A systematic review. *European Urology*, *60*(3), 493–500. https://doi.org/10.1016/j.eururo.2011.05.045

van Kessel, K. E. M., van der Keur, K. A., Dyrskjøt, L., Algaba, F., Welvaart, N. Y. C., Beukers, W., … Zwarthoff, E. C. (2018). Molecular markers increase precision of the European Association of Urology Non-Muscle-Invasive Bladder Cancer Progression Risk Groups. *Clinical Cancer Research*, *24*(7), 1586–1593. https://doi.org/10.1158/1078-0432.CCR-17-2719

Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., … Geschwind, D. H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*, *474*(7351), 380–384. https://doi.org/10.1038/nature10110

Yip, A. M., & Horvath, S. (2007). Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics*, *8*, 22. https://doi.org/10.1186/1471-2105-8-22

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology*, *16*(5), 284–287. https://doi.org/10.1089/omi.2011.0118

Yurakh, A. O., Ramos, D., Calabuig-Farinas, S., Lopez-Guerrero, J. A., Rubio, J., Solsona, E., Llombart-Bosch, A.. (2006). Molecular and immunohistochemical analysis of the prognostic value of cell-cycle regulators in urothelial neoplasms of the bladder. *European Urology*, *50*(3), 506–515. https://doi.org/10.1016/j.eururo.2006.03.027

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*, Article 17. https://doi.org/10.2202/1544-6115.1128

Zhang, B., Shi, X., Xu, G., Kang, W., Zhang, W., Zhang, S., … Zou, X. (2017). Elevated PRC1 in gastric carcinoma exerts oncogenic function and is targeted by piperlongumine in a p53-dependent manner. *Journal of Cellular and Molecular Medicine*, *21*(7), 1329–1341. https://doi.org/10.1111/jcmm.13063

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: Shi J, Zhang P, Liu L, Min X, Xiao Y. Weighted gene coexpression network analysis identifies a new biomarker of CENPF for prediction disease prognosis and progression in nonmuscle invasive bladder cancer. *Mol Genet Genomic Med*. 2019;7:e982. https://doi.org/10.1002/mgg3.982