



# Identifying Potential Prognostic Markers for Muscle-Invasive Bladder Urothelial Carcinoma by Weighted Gene Co-Expression Network Analysis

Yueyi Feng<sup>1</sup> · Yiqing Jiang<sup>2</sup> · Tao Wen<sup>3</sup> · Fang Meng<sup>4,5</sup> · Xiaochen Shu<sup>1</sup>

Received: 15 November 2018 / Accepted: 1 April 2019 / Published online: 22 April 2019  
© Arányi Lajos Foundation 2019

## Abstract

Muscle-invasive bladder urothelial carcinoma (MIBC) is characterized as a genetic heterogeneous cancer with a high percentage of recurrence and worse prognosis. Identify the prognostic potentials of novel genes for muscle-invasive urothelial bladder cancer could at least provide important information for early detection and clinical treatment. Weighted gene co-expression network analysis (WGCNA) algorithm, a powerful systems biology approach, was utilized to extract co-expressed gene networks from mRNA expression dataset to construct transcriptional modules in MIBC samples, which was associated with demographic and clinical traits of MIBC patients. The potential prognostic markers of MIBC were screened out in the discovery dataset and verified in an independent external validation dataset. A total of 8 co-expression modules were detected through the WGCNA algorithm in the discovery datasets based on 401 MIBC samples. One transcriptional module enriched in cell development was observed to be correlated with the MIBC prognosis in the discovery datasets (HR = 1.48, 95%CI = 1.04–2.11) and independently verified in an external dataset (HR = 3.59, 95%CI = 1.09–11.79). High expression of hub genes including discoidin domain receptor tyrosine kinase 2 (DDR2), PDZ and LIM domain 3 (PDLIM3), zinc finger protein 521 (ZNF521), methionine sulf-oxide reductase B3 (MSRB3) were significantly associated with the unfavorable survival of MIBC patients. We identified and validated four novel potential biomarkers associated with prognosis of MIBC patients by constructing genes co-expression networks. The discovery of these genetic markers may provide a new target for the development of MIBC chemotherapeutic drugs.

**Keywords** Muscle-invasive bladder urothelial carcinoma (MIBC) · WGCNA · Prognostic marker · Hub gene

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s12253-019-00657-6>) contains supplementary material, which is available to authorized users.

✉ Fang Meng  
mengf@ism.cams.cn

✉ Xiaochen Shu  
xcshu@suda.edu.cn

<sup>1</sup> Department of Epidemiology, School of Public Health, Medical College of Soochow University, Suzhou 215123, China

<sup>2</sup> Department of General Surgery, Harrison International Peace Hospital, Hengshui 053000, China

<sup>3</sup> Medical Research Centre, Beijing Chao-Yang Hospital, Capital Medical University, Beijing 100020, China

<sup>4</sup> Centre of Systems Medicine, Chinese Academy of Medical Sciences, Beijing 100730, China

<sup>5</sup> Suzhou Institute of Systems Medicine, Suzhou 215123, China

## Introduction

Bladder cancer is a highly heterogeneous cancer of the urinary system, 90% of which is diagnosed as epithelial bladder cancer. On average, 430,000 incident bladder carcinoma cases and 165,000 deaths are reported approximately each year worldwide [1]. Among them, about 1 in 3 are first diagnosed as or later progress to muscle-invasive bladder carcinoma if not detected early or treated properly by underlying major changes at the molecular and genetic level [2]. Pathological grade of MIBC is reported to be usually high, and epithelial–mesenchymal transition (EMT) can activate cancer stem cells and mesenchymal cells in tumor tissue, which makes MIBC more invasive and poor prognosis [3, 4]. Usually, treatment for bladder cancer depends on how deeply the tumor invades into the bladder wall. The standard clinical treatment for MIBC

patients after transurethral radical cystectomy is neoadjuvant chemotherapy, such as cisplatin chemotherapy; however, about 50% of these patients developed drug resistance and metastatic recurrence and eventually died of the disease [5–7]. Although some studies have revealed molecular mechanisms of MIBC at different perspectives [8–10], knowledge on genetic markers and the prognosis of MIBC patients remains scanty. In consideration of the highly invasive nature and the genetic variability of MIBC, exploring the gene-based prognostic markers for MIBC are thus of great importance in clinics.

Weighted gene co-expression network analysis (WGCNA) which emerged as a method for multigene analysis to discover the relationships between gene-gene, and gene-traits has been utilized to study a wide range of biological systems. Different from the conventional clustering rule, WGCNA establishes a scale-free topological network which is similar to protein-protein interaction (PPI) network for weighted Pearson correlation coefficient matrix. WGCNA further clusters unsupervised hierarchical clustering to classify functionally similar genes into a single module, which is more in line with the biological relevance. As a powerful technique for multigene analysis, WGCNA is widely used in finding potential genotypic and phenotypic markers for various types of cancer [11–13].

In this study, we used the WGCNA algorithm to analyze the messenger RNA Seq expression data of 401 MIBC samples from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>). Among the constructed co-expression modules, we selected the modules which were associated with clinical manifestations such as overall survival (OS) and further screened out potential prognostic biomarkers. Furthermore, the selected markers were validated in an independent external dataset of 80 MIBC samples from The European Bioinformatics Institute (EMBL-EBI) database (<https://www.ebi.ac.uk/>). In addition, we explored the potential relationship between identified prognostic markers and EMT.

## Methods

### Data Collection

The raw expression profiles of messenger RNA as well as related demographic and clinical data of patients with muscle-invasive bladder urothelial carcinoma (MIBC) were downloaded from the data repository of TCGA (The Cancer Genome Atlas: <https://portal.gdc.cancer.gov/>) and EMBL-EBI (The European Bioinformatics Institute: <https://www.ebi.ac.uk/>). The TCGA dataset, from which a total of 420 valid pieces of individual data were obtained, including 401 cancer tissues and 19 normal tissues, was

used as discovery dataset to construct co-expression network and to identify potential prognostic biomarkers as well [14]. While dataset E-MTAB-1803 from EMBL-EBI was utilized to validate the findings from TCGA, which included 80 valid pieces of data [15].

### Data Preprocessing

BioMart online tools (<http://asia.ensembl.org/biomart/martview/>) were applied to annotate the mRNA expression data and to calculate the average value of multiple probes corresponding to a single gene. Not annotation and low abundant mRNAs were filtered out accordingly.

### EMT Scores

We used EMT (epithelial mesenchymal transition) signature genes to calculate the EMT score, which was applied to evaluate the mesenchymal expression pattern of each patient [16]. The algorithm was defined as the difference between mean expression values of mesenchymal markers and the mean expression values of epithelial markers. A higher score represents a more inclination to mesenchymal expression.

### Screening Differentially Expressed Genes

The “DESeq2” R package was employed to screen the differentially expressed genes (DEGs) between normal bladder and muscle-invasive bladder urothelial carcinoma samples from TCGA dataset [17]. Expression data was normalized by performing the “VST” function provided by “DESeq2” R package. DEGs were selected according to the following cutoff values: adjusted  $P < 0.05$  and  $|\log_{2}FC$  (log<sub>2</sub> fold change)  $> 1$ .

### Co-Expression Module Construction

After differential expression analysis, DEGs were selected for the construction of co-expression network. The “WGCNA” R package was performed to construct gene network [18]. First, zero expression genes were eliminated by “goodSamplesGenes” function, and the samples were clustered hierarchically to eliminate the samples with obvious outliers. Second, Pearson correlation coefficient (PCC) was calculated for all genes to obtain correlation matrix, and the adjacency matrix was constructed by calculating power function ( $a_{i,j} = |\text{Cor}_{i,j}|^{\beta}$ ) ( $\text{Cor}_{i,j}$  denotes the PCC between gene  $i$  and gene  $j$ ;  $\beta$  was a soft-threshold, which strengthens the strong correlation, punishes the weak correlation and avoids the selection of an arbitrary cut-off) [11, 19]. Third, the adjacency matrix was transformed into a scale-free topologically overlapping matrix (TOM) which had more biological relevance and could be used to

measure the network connectivity of a gene defined as the sum of its adjacency with all other genes [20]. Similarly, the intramodular connectivity was defined as the connectivity between a gene and other genes within a module. Fourth, the similarity distance between genes was calculated based on topological overlap matrix, and the average linkage clustering was carried out accordingly. Based on the criterion of at least 30 genes in each module, the cluster tree was dynamically cut and the co-expressed genes were grouped into a module which was represented by a designated color. Finally, the MEs (module eigengene), the first principal component of each module, were calculated as the representation of the module. Module membership was consequently defined as coefficients of correlation between genes and module MEs.

### Survival Analysis

Survival analysis was carried out using “survival” R packages (<http://cran.r-project.org/web/packages/survival/index.html>). The hazard ratios (HRs) and the corresponding 95% confidence intervals (CIs) were calculated using a Cox PH regression model and survival curves were plotted from Kaplan–Meier estimates. For single gene survival analysis, expression level of genes were categorized into high and low by median value and modules were accordingly divided into high and low expressed in all samples. Overall survival was used as the ending points of the study.

### Screening and Verification of Hub Genes

Hub gene was defined as having high intramodular connectivity ( $k_{in}$ ) and highly associated with clinical traits as well. In this study, the selection of hub genes was guided by the following criteria: (i) internal connectivity ranked as the top 20 of all genes in the module, and (ii) GS (gene significance) which was defined as negative  $\log_{10}$  for  $P$  value of gene cox regression was greater than 1.5. In the validation dataset, that is, E-MTAB-1803, cox regression analysis was performed to verify the effect of hub gene on prognosis of MIBC patients.

### Gene Function Analysis

Gene ontology (GO) annotation, enrichment analysis and pathway analysis were carried out with “clusterProfiler” R package to explore the most likely gene function [21]. This R package crawls the latest annotation information from Gene Ontology Consortium site online (<http://www.geneontology.org/>) and Kyoto Encyclopedia of Genes and Genomes (<https://www.kegg.jp/>), which together reduced background noise information and made the results more reliable.

## Results

### Screening of Differentially Expressed Genes (DEGs)

In this study, 2287 up-regulated mRNAs and 1938 down-regulated mRNAs total 4225 DEGs were detected from a total of 20,075 genes according to the criteria of “P (Benjamini adjusted P value) < 0.05” and “ $|\log_{2}FC$  (log<sub>2</sub> Fold Change) | > 1”, based on differential expression analysis from 420 MIBC messenger RNA Count data.

### Identification of Co-Expressed Gene Clusters

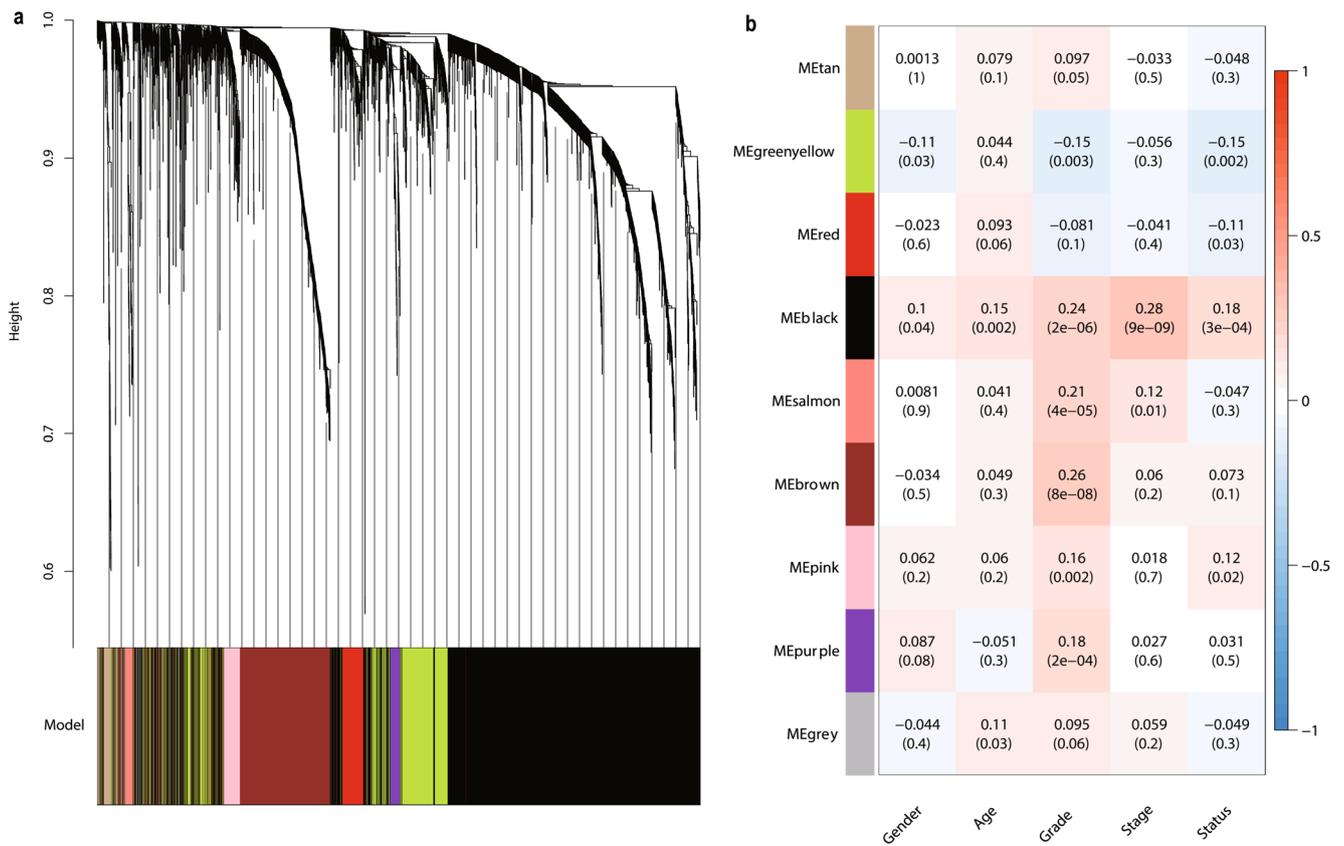
Weighted gene co-expression network analysis (WGCNA) algorithm was performed to analyze the expression of 4225 genes in 401 MIBC samples. We chose  $\beta = 4$  as the soft threshold to weight the correlation matrix and checked scale-free topology (Supplementary file 1). Totally, 8 co-expression modules and 1 non-co-expression module (marked as grey) were identified (Fig. 1a). Number of genes contained in the constructed modules ranged from 53 to 2386 (Table 1). MEs matrix which was obtained by principal component analysis provided the information of overall expression level of each module. For any single module, Pearson correlation coefficient was calculated between each gene within the module and its MEs, a measure of module membership (Supplementary file 2).

### Correlation between Co-Expression Modules and Clinical Traits

Information on clinic traits was publically accessible in TCGA database. To determine whether the co-expression module is associated with clinical traits, we calculated Spearman correlation coefficient (SCC) matrix between MEs (undichotomized) and clinical features such as age, gender, pathological grade, tumor TNM stage and so on. We observed positive associations of the black (SCC = 0.24,  $P = 2 \times 10^{-6}$ ), brown (SCC = 0.26,  $P = 8 \times 10^{-8}$ ), and purple (SCC = 0.18,  $P = 2 \times 10^{-4}$ ) module MEs with tumor pathological type (differentiated vs undifferentiated). Black module MEs (SCC = 0.28,  $P = 9 \times 10^{-9}$ ) was found to be correlated with tumor TNM stage (Fig. 1b).

### Identifying Associations between Co-Expression Modules with Overall Survival

In order to identify the relationship between co-expression modules and overall survival (OS), cox regression model was used to calculate the hazard ratio (HR), its 95% confidence interval and  $p$  value of dichotomized MEs for each module (Table 1). In unadjusted survival analyses, greenyellow module (HR = 0.71,  $P = 2.4 \times 10^{-2}$ , 95%CI = 0.52–0.96) and black



**Fig. 1** Identification of MIBC co-expression modules by WGCNA, and correlation of modules with clinical traits. **a** The unsupervised hierarchical cluster dendrogram was used to identify co-expression modules and assign colors to them. A total of 9 modules were identified, ranging in size from 53 to 2386 genes, and 53 non-co-expressed genes were classified as

grey modules. **b** The Spearman Correlation matrix Heatmap of co-expression module MEs and clinical traits (age, gender, TNM stage, pathological grade, survival status). The size of SCC represents the strength of the relationship between MEs and clinical traits. The larger the absolute value of SCC, the darker the color

module (HR = 1.75,  $P < 1.0 \times 10^{-3}$ , 95%CI = 1.29–2.37) were found to be associated with the OS. After multiple adjustment,

the significant association in the black module was still observed in the discovery dataset (HR = 1.48,  $P = 3.02 \times 10^{-2}$ ,

**Table 1** Survival analysis for gene co-expression modules with OS as endpoints in the discovery dataset and validation dataset

Module	Gene counts	Discovery dataset (N = 401)						Validation dataset (N = 80)					
		Unadjusted			Adjusted*			Unadjusted			Adjusted*		
		P	HR	95% CI	P	HR	95% CI	P	HR	95% CI	P	HR	95% CI
Tan	54	$1.37 \times 10^{-1}$	0.80	0.59–1.07	$2.64 \times 10^{-1}$	0.83	0.59–1.15	$9.50 \times 10^{-1}$	0.98	0.53–1.81	$6.17 \times 10^{-1}$	0.76	0.27–2.19
Greenyellow	596	$2.40 \times 10^{-2}$	0.71	0.52–0.96	$7.94 \times 10^{-1}$	0.95	0.64–1.41	$3.11 \times 10^{-1}$	0.73	0.40–1.34	$7.59 \times 10^{-1}$	1.15	0.48–2.73
Red	160	$2.88 \times 10^{-1}$	0.85	0.63–1.15	$2.84 \times 10^{-1}$	0.83	0.60–1.15	$5.01 \times 10^{-1}$	0.81	0.44–1.49	$2.84 \times 10^{-1}$	2.55	0.46–14.04
Black	2386	<b><math>&lt;1.00 \times 10^{-3}</math></b>	<b>1.75</b>	<b>1.29–2.37</b>	<b><math>3.02 \times 10^{-2}</math></b>	<b>1.48</b>	<b>1.04–2.11</b>	<b><math>2.90 \times 10^{-2}</math></b>	<b>2.01</b>	<b>1.07–3.75</b>	<b><math>3.55 \times 10^{-2}</math></b>	<b>3.59</b>	<b>1.09–11.79</b>
Salmon	54	$4.28 \times 10^{-1}$	0.89	0.66–1.19	$1.57 \times 10^{-1}$	0.79	0.57–1.10	$7.35 \times 10^{-1}$	0.90	0.49–1.66	$2.33 \times 10^{-1}$	0.42	0.10–1.74
Blown	718	$1.76 \times 10^{-1}$	1.23	0.91–1.65	$9.30 \times 10^{-2}$	1.36	0.95–1.94	$5.10 \times 10^{-1}$	0.81	0.44–1.50	$3.64 \times 10^{-1}$	2.05	0.43–9.73
Pink	117	$9.80 \times 10^{-2}$	1.29	0.95–1.73	$1.50 \times 10^{-1}$	1.27	0.92–1.75	$9.48 \times 10^{-1}$	0.98	0.53–1.80	$7.93 \times 10^{-1}$	0.86	0.28–2.64
Purple	87	$3.55 \times 10^{-1}$	1.15	0.85–1.55	$4.69 \times 10^{-1}$	0.87	0.60–1.27	$4.27 \times 10^{-1}$	1.28	0.69–2.37	$5.20 \times 10^{-1}$	0.75	0.31–1.81
Grey	53	$3.00 \times 10^{-1}$	0.85	0.63–1.15	$6.70 \times 10^{-1}$	0.93	0.66–1.31	$5.76 \times 10^{-1}$	1.19	0.64–2.21	$7.52 \times 10^{-2}$	2.70	0.90–8.08

Entries in bold indicates results significant at the 0.05 level

\*Adjusted for age, tumor TNM stage, gender, and pathological grade

95% CI = 1.04–2.11) and validated in an independent external dataset (HR = 3.586,  $P = 3.55 \times 10^{-2}$ , 95% CI = 1.09–11.79). In addition, this discovery was consistent with our finding that gene expression was correlated with higher tumor grade (SCC = 0.24) and higher tumor TNM stage (SCC = 0.28) in black module. Finally, we chose the black module for further analysis for searching for hub genes.

### Screening Hub Genes and Association with OS

First, we identified 517 genes significantly associated with poor prognosis from 2386 genes contained in the identified black co-expression module after univariate survival analysis. HR values and corresponding 95% CIs with  $P$  values for all genes were listed in supplementary file 2.

We, next, calculated the connectivity within the module and the total connectivity (Supplementary file 2). According to the size of GS value ( $> 1.5$ ) and the strength of internal connectivity ( $k_{in}$  ranked as top 20), 8 genes associated with poor outcomes in black module were identified as hub genes (BOC, DDR2, MSRB3, MYLK, PDLIM3, ZNF521, RASL12, CALD1) (Table 2). Furthermore, we conducted validation in an external independent dataset for these 8 hub genes. However, only three of them (DDR2, ZNF521, PDLIM3) have been verified, and one (MSRB3) was marginally significant (Table 2). Figure 2 showed Kaplan-Meier's estimation of gene expression with the time of death for the hub genes, in the discovery dataset and the validation dataset. Additionally, we calculated the correlation between the expression value of hub genes and EMT scores and found that DDR2 ( $\rho = 0.85$ ), MSRB3 ( $\rho = 0.86$ ), PDLIM3 ( $\rho = 0.82$ ), ZNF521 ( $\rho = 0.78$ ) were positively correlated with EMT (Fig. 3).

### Gene Function Analysis

In order to further understand the function of any designated co-expression module, we used "clusterProfiler" R package

for gene function annotation and enrichment analysis. GO enrichment analysis showed that the biological processes enriched by black modules included extracellular matrix organization, extracellular structure organization, and regulation of membrane potential etc. Pathway analysis showed the pathways included neuroactive ligand-receptor interaction, MAPK signaling pathway, cGMP-PKG signaling pathway and calcium signaling pathway etc. (Supplementary file 3).

### Discussion

The diversity in genetic background of muscle-invasive bladder urothelial carcinoma (MIBC) leads to a great variability in cancer prognosis and response to treatment although new forms of adjuvant chemotherapy have been proposed in clinics [5, 22, 23]. Few studies had attempted to explore the potentials of biological markers which could be relevant to biochemical and clinical manifestations of MIBC at the level of cellular and molecular origins or biological pathways [24]; however, there are no evidences that markedly linked genetic markers with prognosis of MIBC so far. Based on the publicly accessible data repository, we performed a powerful coexpression-based analysis, WGCNA (weighted gene co-expression network analysis), to analyze a messenger RNA expression dataset to identify the relationships between gene-gene, and gene-clinical manifestations, with focus on the prognosis of MIBC. Overall, a total of 4225 genes expression data coming from 401 MIBC patients had been analyzed in this study. WGCNA uses the method of soft threshold to power the correlation coefficient between genes, transforms the relationship between genes into scale-free topological overlapped networks in accordance with biological significance, and utilizes the internal connectivity to measure the relationship between genes. It can effectively reduce error caused by manmade cut-off correlation coefficient. In addition, the unsupervised learning method used by WGCNA effectively

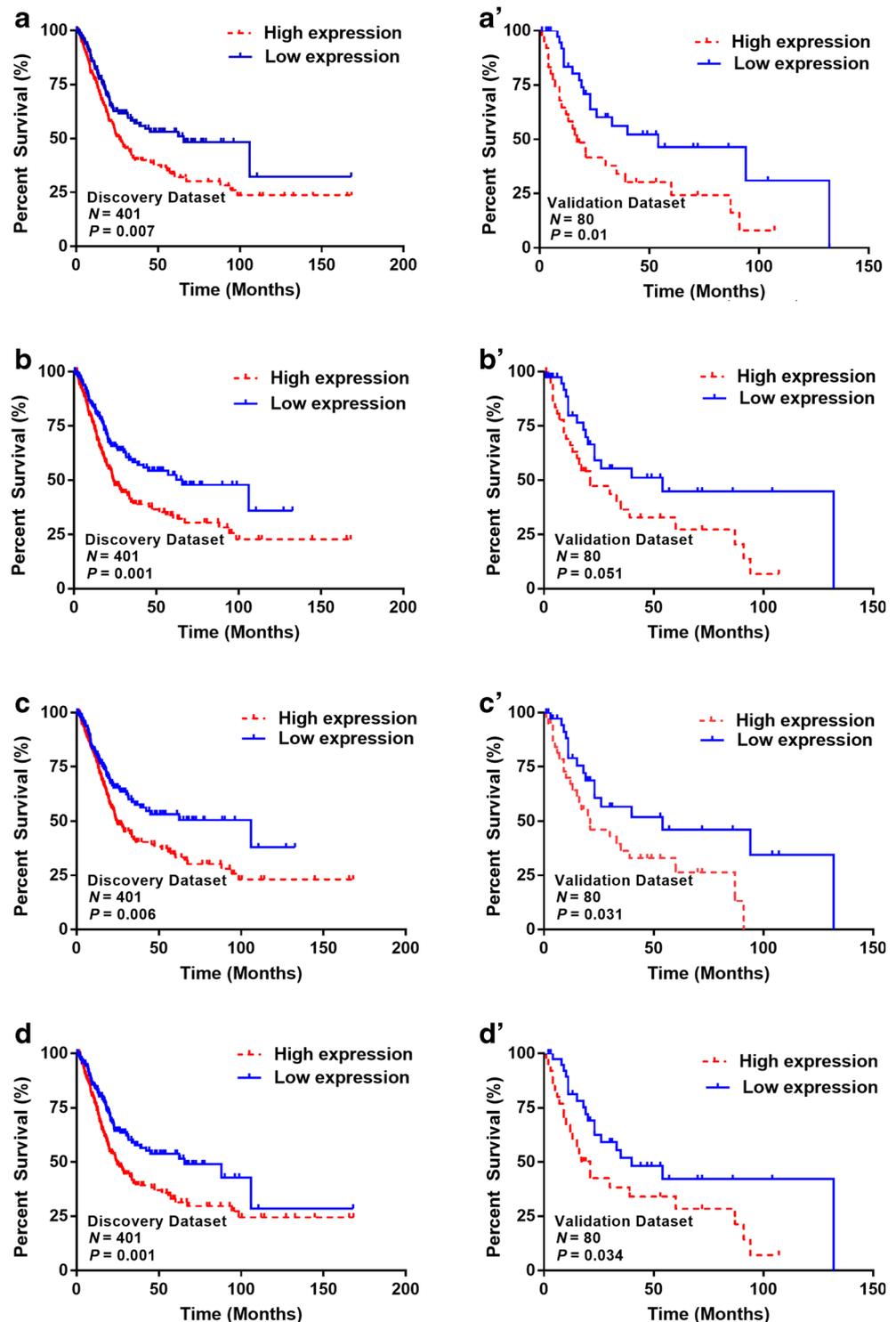
**Table 2** Survival analysis for hub genes with OS as endpoints in the discovery dataset and validation dataset

Hub gene	k.in rank	Discovery dataset			Validation dataset		
		$P$	HR	95% CI	$P$	HR	95% CI
MSRB3	1	<b><math>1.00 \times 10^{-3}</math></b>	<b>1.66</b>	<b>1.22–2.25</b>	<b><math>5.10 \times 10^{-2}</math></b>	<b>1.87</b>	<b>1.00–3.49</b>
BOC	2	$<1.00 \times 10^{-3}$	1.76	1.29–2.38	$6.60 \times 10^{-2}$	1.79	0.96–3.32
DDR2	3	<b><math>8.00 \times 10^{-3}</math></b>	<b>1.50</b>	<b>1.11–2.04</b>	<b><math>1.20 \times 10^{-2}</math></b>	<b>2.22</b>	<b>1.19–4.11</b>
PDLIM3	5	<b><math>7.00 \times 10^{-3}</math></b>	<b>1.52</b>	<b>1.12–2.06</b>	<b><math>3.20 \times 10^{-2}</math></b>	<b>2.01</b>	<b>1.06–3.82</b>
RASL12	8	$2.70 \times 10^{-2}$	1.40	1.04–1.89	$3.56 \times 10^{-1}$	1.33	0.72–2.45
ZNF521	14	<b><math>2.00 \times 10^{-3}</math></b>	<b>1.63</b>	<b>1.20–2.21</b>	<b><math>3.70 \times 10^{-2}</math></b>	<b>1.92</b>	<b>1.04–3.56</b>
CALD1	18	$1.10 \times 10^{-2}$	1.48	1.09–2.01	$2.50 \times 10^{-1}$	1.43	0.78–2.62
MYLK	20	$4.0 \times 10^{-3}$	1.56	1.15–2.11	$1.41 \times 10^{-1}$	1.59	0.86–2.93

Dataset and validation dataset. Entries in bold indicates results significant at the 0.05 level

\* $k_{in}$  is the degree of connectivity between a designated gene and other genes in a module

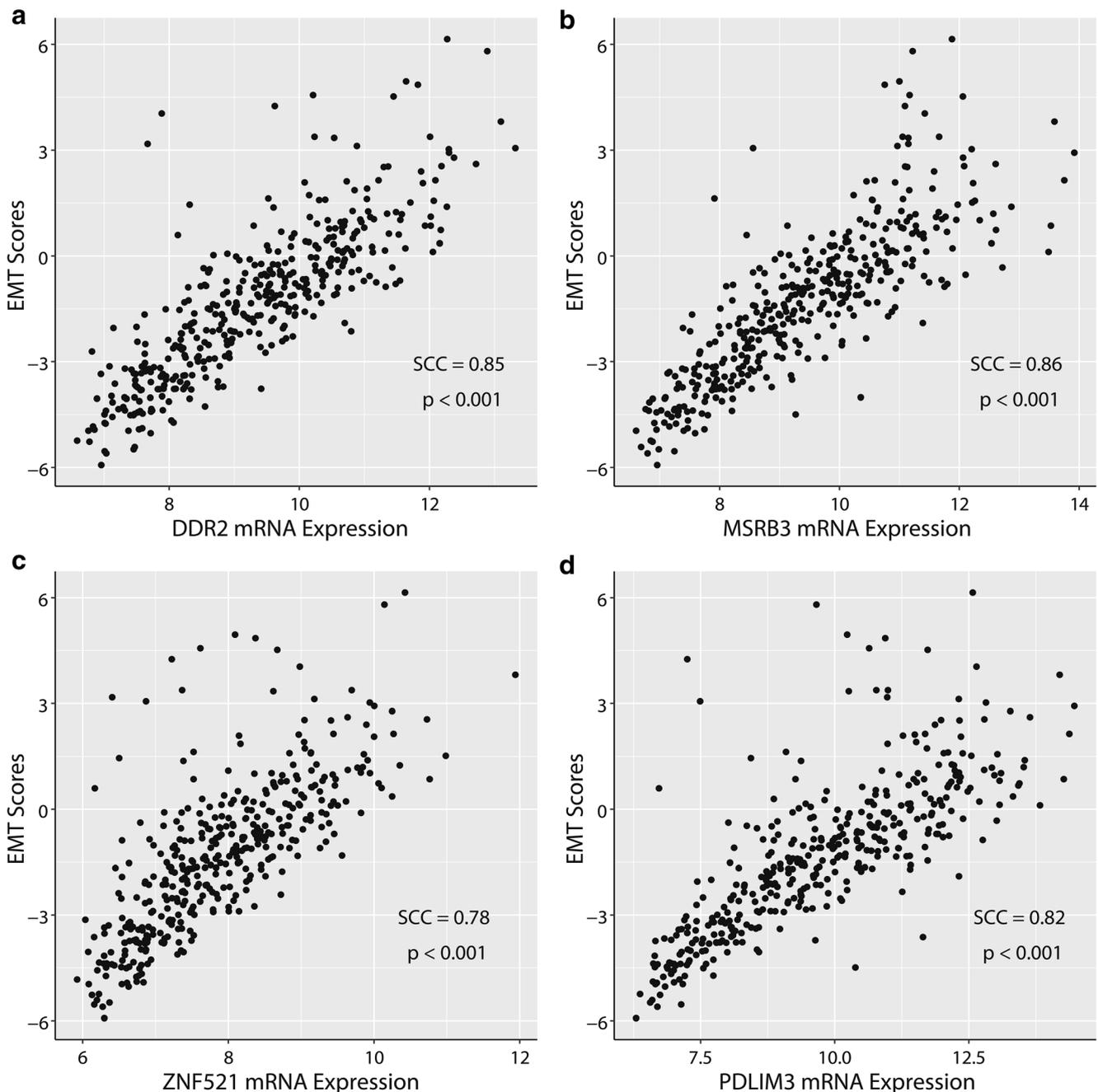
**Fig. 2** Kaplan–Meier estimates of probability of overall survival among different expression level of hub genes. **a** *DDR2* Kaplan–Meier survival plot for OS in discovery dataset. **a'** *DDR2* Kaplan–Meier survival plot for OS in validation dataset. **b** *MSRB3* Kaplan–Meier survival plot for OS in discovery dataset. **b'** *MSRB3* Kaplan–Meier survival plot for OS in validation dataset. **c** *PDLIM3* Kaplan–Meier survival plot for OS in discovery dataset. **c'** *PDLIM3* Kaplan–Meier survival plot for OS in validation dataset. **d** *ZNF521* Kaplan–Meier survival plot for OS in discovery dataset. **d'** *ZNF521* Kaplan–Meier survival plot for OS in validation dataset. The increased expression of these hub genes indicates poor prognosis



avoids the subjective decision error caused by the influence of previous research reports.

In the present study, we constructed eight co-expression modules from 4225 genes, calculated the expression level of modules based on principal component analysis, and performed correlation analysis and Cox PH regression analysis. We observed that high expression of

black module containing 2386 genes was associated with worse overall survival of MIBC patients in the discovery dataset. After multiple adjustment, high expression level in black module still significantly predicted an unfavorable prognosis in discovery dataset (HR = 1.48, 95% CI = 1.04–2.11). This remarkable finding was verified in an independent external validation dataset. Furthermore, we



**Fig. 3** Plot of spearman correlation between (a) EMT Score and *DDR2* expression value, b EMT Score and *MSRB3* expression value, c EMT Score and *ZNF521* expression value, d EMT Score and *PDLIM3*

expression value. The x axis is the normalized mRNA expression value, and the y axis is the EMT score

screened out eight hub genes in the discovery dataset, three of them being verified to be significant (*DDR2*, *PDLIM3*, *ZNF521*) and one being marginal significant (*MSMR3*) in the validation dataset. Moreover, biological relevances were mainly suggestive of extracellular matrix organization, extracellular structure organization, regulation of membrane potential etc. and the biological pathways could be the neuroactive ligand-receptor interaction,

MAPK signaling pathway, cGMP-PKG signaling pathway, calcium signaling pathway and so on.

*DDR2*, one of the RTKs (receptor tyrosine kinases) members in collagen receptor family, is mainly involved in extracellular matrix process and ERK1 / ERK2 regulation. It plays an important role in various cellular functions and disease processes such as malignant progression of ovarian, breast, prostate, lung and kidney. *DDR2* was reported to influence

the invasiveness of tumor cells by regulating the activity of matrix remodeling enzyme in the malignant progression of cancer [25–27]. A Japanese study found that high expression level of DDR2 was relevant to a worse prognosis of colorectal cancer [28]. A cohort study conducted in Taiwan, for the first time to our knowledge, associated overexpression of DDR2 with unfavorable prognosis of MIBC, which was consistent with the main findings in our study [24].

PDLIM3 encoding a protein that contains a PDZ domain and a LIM domain, had been reported to be associated with the development of hypertrophic cardiomyopathy, but not carcinogenesis [29, 30]. As we know, prognosis and PDLIM3 overexpression in MIBC patients had not been associated so far. The underlying mechanism of unfavorable prognosis for the high expression of PDLIM3 in MIBC patients observed in our study remains unknown. The GO annotations for this gene in our study had already linked it with cytoskeletal protein binding and metal ion binding. The mechanism underlying the carcinogenesis, malignant prognosis and pathological differentiation warrants further study.

Zinc finger protein 521 (ZNF521) is a polyzinc finger transcription factor which is mainly involved in cell differentiation and transcription regulation processes. Overexpression of ZNF521 was positively correlated with MLL-rearranged acute myeloid leukemia as a strong regulator of hematopoietic stem cell homeostasis in a recent study [31]. Based on a human medulloblastoma study, ZNF521 may play an indirect role in the process of cell cycle and apoptosis through the interaction with NuRD complex and thus affect the occurrence and development of tumor [32]. Translation elongation factor 1 subunit beta (EFB1) had been found to play an inhibitory role in the development of cancer. Inhibition of EFB1 expression by binding the carboxyl terminal of ZNF521 to EFB1 had been observed in a study of B lymphocytes although the mechanism by which EFB1 inhibits the development and metastasis of cancer has not yet been clearly established [33, 34]. Further study focusing on the relationship between ZNF521 and EFB1 may shed light on the underlying mechanism of ZNF521 expression with prognosis of MIBC.

The methionine sulfoxide reductase MSRB3 mainly located in mitochondria and endoplasmic reticulum (ER) belongs to one of the three types of MRSB family and encodes an enzyme that catalyzes methionine R sulfoxide reduction to methionine [35, 36]. Role of MSRB3 in malignant procession has not been clearly elucidated. Studies reported that MSRB3 was involved in the regulation of cell cycle and cell proliferation in breast cancer and hepatocellular carcinoma and the overexpression of MSRB3 may promote the proliferation of cancer cells. The biological mechanism could due to the inhibition of ER stress response and the imbalance of cellular calcium homeostasis [37–40]. To date, expression of MSRB3 had not been

associated with cancer prognosis. The possible biological relevances suggested by the annotation analysis in our study had been directed to the calcium homeostasis, the signal exchange between cells, or excessive proliferation of cancer cells, which, all together, may promote the unfavorable prognosis of MIBC patients.

Epithelial–mesenchymal transition induces polar epithelial cells to obtain interstitial phenotype and promotes the production of cancer stem cells, which plays an important role in cancer metastasis, drug resistance and escape apoptosis [41, 42]. Studies had shown that EMT process was associated with poor prognosis of MIBC [4]. Our main finding that the correlation of the expression level of DDR2, MSRB3, PDLIM3 and ZNF521 positively with EMT status may suggest that the high expression of four identified prognostic marker genes could promote EMT process. The existing mechanism underlying this association was explained as that DDR2 may promote EMT through ERK2/SNAIL1 pathway, mTORC2 activation and ATK phosphorylation [43, 44]. More interpretations need to be discovered in further studies.

WGCNA, a powerful systems biology approach, was used to analyze mRNA expression dataset composed of 401 MIBC samples to identify genes which could be potentially associated with cancer prognosis. WGCNA was further used to annotate the underlying biological mechanism responsive for the postulated association. The distinguishing advantage of WGCNA include that the unsupervised nature of hierarchical clustering method could avoid potential biases and subjective decisions due to previous reports of related studies or the supervised selection of genes. Notwithstanding, the present study do have limits. First, sampling bias may exist considering the relatively small size of the validation dataset, which may hinder the concordance of the significant results between both datasets to some extent. In addition, absence of relevant clinical manifestations in TCGA database impeded further exploration of their effects on the observed association.

## Conclusions

In summary, we first constructed eight gene co-expression clusters based on a total of 20,075 genes from TCGA dataset by using WGCNA algorithm, from which we further identified and validated four novel genes which at their high level of expression were observed to be correlated with unfavorable prognosis of MIBC patients. Moreover, the constructed modules were found to be associated with tumor TNM stage and pathological grade in the same direction with prognosis. In addition, biological function of these four potential prognostic biomarkers by using the annotation analysis could be directed to cancer cell proliferation by regulation of membrane potential and ligand-receptor interaction for a distinct possibility.

**Acknowledgements** The corresponding author would like to thank all the co-workers for collecting, managing, maintaining and interpreting the data used in this analysis. We also appreciate Soochow University for providing the financial support to conduct the present study.

**Author's Contributions** Conception and design: Yueyi Feng, Xiaochen Shu.

Financial support: Xiaochen Shu.

Collection and assembly of data: Yueyi Feng.

Data analysis and interpretation: Yueyi Feng, Yiqing Jiang, Tao Wen, Fang Meng, Xiaochen Shu.

Manuscript writing: Yueyi Feng, Xiaochen Shu.

Final approval of manuscript: Yueyi Feng, Yiqing Jiang, Tao Wen, Fang Meng, and Xiaochen Shu.

Xiaochen Shu will be responsible for the overall content as guarantor of this work.

**Funding** This work was supported by Scientific Research Foundation for Talented Scholars in Soochow University, China (Q413900215).

## Compliance with Ethical Standards

**Conflict of Interest** No potential conflict of interests was reported by the authors.

## References

1. Antoni S, Ferlay J, Soerjomataram I, Znaor A, Jemal A, Bray F (2017) Bladder Cancer incidence and mortality: a global overview and recent trends. *Eur Urol* 71(1):96–108. <https://doi.org/10.1016/j.eururo.2016.06.010>
2. Kamat AM, Hahn NM, Efsthathiou JA, Lerner SP, Malmström P-U, Choi W, Guo CC, Lotan Y, Kassouf W (2016) Bladder cancer. *Lancet* 388(10061):2796–2810. [https://doi.org/10.1016/s0140-6736\(16\)30512-8](https://doi.org/10.1016/s0140-6736(16)30512-8)
3. Migita T, Ueda A, Ohishi T, Hatano M, Seimiya H, Horiguchi SI, Koga F, Shibasaki F (2017) Epithelial-mesenchymal transition promotes SOX2 and NANOG expression in bladder cancer. *Lab Invest* 97:567–576. <https://doi.org/10.1038/labinvest.2017.17>
4. Baumgart E, Cohen MS, Silva Neto B, Jacobs MA, Wotkowicz C, Rieger-Christ KM, Biolo A, Zeheb R, Loda M, Libertino JA, Summerhayes IC (2007) Identification and prognostic significance of an epithelial-mesenchymal transition expression profile in human bladder tumors. *Clinical cancer research : an official journal of the American association for. Cancer Res* 13(6):1685–1694. <https://doi.org/10.1158/1078-0432.CCR-06-2330>
5. Grossman HB, Natale RB, Tangen CM, Speights VO, Vogelzang NJ, Trump DL, deVere White RW, Sarosdy MF, Wood DP Jr, Raghavan D, Crawford ED (2003) Neoadjuvant chemotherapy plus cystectomy compared with cystectomy alone for locally advanced bladder cancer. *N Engl J Med* 349(9):859–866. <https://doi.org/10.1056/NEJMoa022148>
6. Amir S, Mabjeesh NJ (2017) microRNA expression profiles as decision-making biomarkers in the management of bladder cancer. *Histol Histopathol* 32(2):107–119. <https://doi.org/10.14670/HH-11-814>
7. Dudek AM, van Kampen JGM, Witjes JA, Kiemeny L, Verhaegh GW (2018) LINC00857 expression predicts and mediates the response to platinum-based chemotherapy in muscle-invasive bladder cancer. *Cancer Med* 7:3342–3350. <https://doi.org/10.1002/cam4.1570>
8. Siegel RL, Miller KD, Jemal A (2018) Cancer statistics, 2018. *CA Cancer J Clin* 68(1):7–30. <https://doi.org/10.3322/caac.21442>
9. Kiselyov A, Bunimovich-Mendrazitsky S, Startsev V (2016) Key signaling pathways in the muscle-invasive bladder carcinoma: clinical markers for disease modeling and optimized treatment. *Int J Cancer* 138(11):2562–2569. <https://doi.org/10.1002/ijc.29918>
10. Masson-Lecomte A, Rava M, Real FX, Hartmann A, Allory Y, Malats N (2014) Inflammatory biomarkers and bladder cancer prognosis: a systematic review. *Eur Urol* 66(6):1078–1091. <https://doi.org/10.1016/j.eururo.2014.07.033>
11. Clarke C, Madden SF, Doolan P, Aherne ST, Joyce H, O'Driscoll L, Gallagher WM, Hennessy BT, Moriarty M, Crown J, Kennedy S, Clynes M (2013) Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. *Carcinogenesis* 34(10):2300–2308. <https://doi.org/10.1093/carcin/bgt208>
12. Oros Klein K, Oualkacha K, Lafond MH, Bhatnagar S, Tonin PN, Greenwood CM (2016) Gene Coexpression analyses differentiate networks associated with diverse cancers harboring TP53 missense or null mutations. *Front Genet* 7:137. <https://doi.org/10.3389/fgene.2016.00137>
13. Guo L, Zhang K, Bing Z (2017) Application of a coexpression network for the analysis of aggressive and nonaggressive breast cancer cell lines to predict the clinical outcome of patients. *Mol Med Rep* 16(6):7967–7978. <https://doi.org/10.3892/mmr.2017.7608>
14. Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, Hinoue T, Laird PW, Hoadley KA, Akbani R, Castro MAA, Gibb EA, Kanchi RS, Gordenin DA, Shukla SA, Sanchez-Vega F, Hansel DE, Czerniak BA, Reuter VE, Su X, de Sa Carvalho B, Chagas VS, Mungall KL, Sadeghi S, Pedamallu CS, Lu Y, Klimczak LJ, Zhang J, Choo C, Ojesina AI, Bullman S, Leraas KM, Lichtenberg TM, Wu CJ, Schultz N, Getz G, Meyerson M, Mills GB, McConkey DJ, Network TR, Weinstein JN, Kwiatkowski DJ, Lerner SP (2017) Comprehensive molecular characterization of muscle-invasive bladder Cancer. *Cell* 171(3):540–556 e525. <https://doi.org/10.1016/j.cell.2017.09.007>
15. Rebouissou S, Bernard-Pierrot I, de Reynies A, Lepage ML, Krucker C, Chapeaublanc E, Herault A, Kamoun A, Caillaud A, Letouze E, Elarouci N, Neuzillet Y, Denoux Y, Molinie V, Vordos D, Laplanche A, Maille P, Soyeux P, Ofualuka K, Reyat F, Biton A, Sibony M, Paoletti X, Southgate J, Benhamou S, Lebret T, Allory Y, Radvanyi F (2014) EGFR as a potential therapeutic target for a subset of muscle-invasive bladder cancers presenting a basal-like phenotype. *Sci Transl Med* 6(244):244ra291. <https://doi.org/10.1126/scitranslmed.3008970>
16. Mak MP, Tong P, Diao L, Cardnell RJ, Gibbons DL, William WN, Skoulidis F, Parra ER, Rodriguez-Canales J, Wistuba II, Heymach JV, Weinstein JN, Coombes KR, Wang J, Byers LA (2016) A patient-derived, pan-Cancer EMT signature identifies global molecular alterations and immune target enrichment following epithelial-to-mesenchymal transition. *Clin Cancer Res* 22(3):609–620. <https://doi.org/10.1158/1078-0432.CCR-15-0876>
17. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550. <https://doi.org/10.1186/s13059-014-0550-8>
18. 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>
19. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 4: Article17. <https://doi.org/10.2202/1544-6115.1128>
20. Yuan L, Chen L, Qian K, Qian G, Wu CL, Wang X, Xiao Y (2017) Co-expression network analysis identified six hub genes in association with progression and prognosis in human clear cell renal cell carcinoma (ccRCC). *Genom Data* 14:132–140. <https://doi.org/10.1016/j.gdata.2017.10.006>

21. Yu G, Wang LG, Han Y, He QY (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *Omic* : a Journal Of Integrative Biology 16(5):284–287. <https://doi.org/10.1089/omi.2011.0118>
22. Achkar T, Parikh RA (2018) Adjuvant therapy in muscle-invasive bladder Cancer and upper tract urothelial carcinoma. *Urol Clin N Am* 45(2):257–266. <https://doi.org/10.1016/j.ucl.2017.12.010>
23. Hermans TJN, Voskuilen CS, van der Heijden MS, Schmitz-Drager BJ, Kassouf W, Seiler R, Kamat AM, Grivas P, Kiltie AE, Black PC, van Rhijn BWG (2017) Neoadjuvant treatment for muscle-invasive bladder cancer: the past, the present, and the future. *Urol Oncol* 36:413–422. <https://doi.org/10.1016/j.urolonc.2017.10.014>
24. Tsai MC, Li WM, Huang CN, Ke HL, Li CC, Yeh HC, Chan TC, Liang PI, Yeh BW, Wu WJ, Lim SW, Li CF (2016) DDR2 overexpression in urothelial carcinoma indicates an unfavorable prognosis: a large cohort study. *Oncotarget* 7(48):78918–78931. <https://doi.org/10.18632/oncotarget.12912>
25. Corsa CA, Brenot A, Grither WR, Van Hove S, Loza AJ, Zhang K, Ponik SM, Liu Y, DeNardo DG, Eliceiri KW, Keely PJ, Longmore GD (2016) The action of Discoidin domain receptor 2 in basal tumor cells and stromal Cancer-associated fibroblasts is critical for breast Cancer metastasis. *Cell Rep* 15(11):2510–2523. <https://doi.org/10.1016/j.celrep.2016.05.033>
26. Grither WR, Divine LM, Meller EH, Wilke DJ, Desai RA, Loza AJ, Zhao P, Lohrey A, Longmore GD, Fuh KC (2018) TWIST1 induces expression of discoidin domain receptor 2 to promote ovarian cancer metastasis. *Oncogene* 37(13):1714–1729. <https://doi.org/10.1038/s41388-017-0043-9>
27. Kim D, Ko P, You E, Rhee S (2014) The intracellular juxtamembrane domain of discoidin domain receptor 2 (DDR2) is essential for receptor activation and DDR2-mediated cancer progression. *Int J Cancer* 135(11):2547–2557. <https://doi.org/10.1002/ijc.28901>
28. Sasaki S, Ueda M, Iguchi T, Kaneko M, Nakayama H, Watanabe T, Sakamoto A, Mimori K (2017) DDR2 expression is associated with a high frequency of peritoneal dissemination and poor prognosis in colorectal Cancer. *Anticancer Res* 37(5):2587–2591. <https://doi.org/10.21873/anticancer.11603>
29. Bagnall RD, Yeates L, Semsarian C (2010) Analysis of the Z-disc genes PDLIM3 and MYPN in patients with hypertrophic cardiomyopathy. *Int J Cardiol* 145(3):601–602. <https://doi.org/10.1016/j.ijcard.2010.08.004>
30. Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S, Danielsson A, Edlund K, Asplund A, Sjostedt E, Lundberg E, Szgyarto CA, Skogs M, Takanen JO, Berling H, Tegel H, Mulder J, Nilsson P, Schwenk JM, Lindskog C, Danielsson F, Mardinoglu A, Sivertsson A, von Feilitzen K, Forsberg M, Zwaalen M, Olsson I, Navani S, Huss M, Nielsen J, Ponten F, Uhlen M (2014) Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* 13(2):397–406. <https://doi.org/10.1074/mcp.M113.035600>
31. Germano G, Morello G, Aveic S, Pinazza M, Minuzzo S, Frasson C, Persano L, Bonvini P, Viola G, Bresolin S, Tregnago C, Paganin M, Pigazzi M, Indraccolo S, Basso G (2017) ZNF521 sustains the differentiation block in MLL-rearranged acute myeloid leukemia. *Oncotarget* 8(16):26129–26141. <https://doi.org/10.18632/oncotarget.15387>
32. Spina R, Filocamo G, Iaccino E, Scicchitano S, Lupia M, Chiarella E, Mega T, Bernaudo F, Pelaggi D, Mesuraca M, Pazzaglia S, Semenkov S, Bar EE, Kool M, Pfister S, Bond HM, Eberhart CG, Steinkuhler C, Morrone G (2013) Critical role of zinc finger protein 521 in the control of growth, clonogenicity and tumorigenic potential of medulloblastoma cells. *Oncotarget* 4(8):1280–1292. <https://doi.org/10.18632/oncotarget.1176>
33. Harder L, Eschenburg G, Zech A, Kriebitzsch N, Otto B, Streichert T, Behlich AS, Dierck K, Klingler B, Hansen A, Stanulla M, Zimmermann M, Kremmer E, Stocking C, Horstmann MA (2013) Aberrant ZNF423 impedes B cell differentiation and is linked to adverse outcome of ETV6-RUNX1 negative B precursor acute lymphoblastic leukemia. *J Exp Med* 210(11):2289–2304. <https://doi.org/10.1084/jem.20130497>
34. Mega T, Lupia M, Amodio N, Horton SJ, Mesuraca M, Pelaggi D, Agosti V, Grieco M, Chiarella E, Spina R, Moore MA, Schuringa JJ, Bond HM, Morrone G (2011) Zinc finger protein 521 antagonizes early B-cell factor 1 and modulates the B-lymphoid differentiation of primary hematopoietic progenitors. *Cell Cycle* 10(13):2129–2139. <https://doi.org/10.4161/cc.10.13.16045>
35. Brot N, Weissbach L, Werth J, Weissbach H (1981) Enzymatic reduction of protein-bound methionine sulfoxide. *Proc Natl Acad Sci U S A* 78(4):2155–2158
36. Kim HY, Gladyshev VN (2004) Methionine sulfoxide reduction in mammals: characterization of methionine-R-sulfoxide reductases. *Mol Biol Cell* 15(3):1055–1064. <https://doi.org/10.1091/mbc.E03-08-0629>
37. Kwak GH, Kim HY (2017) MsrB3 deficiency induces cancer cell apoptosis through p53-independent and ER stress-dependent pathways. *Arch Biochem Biophys* 621:1–5. <https://doi.org/10.1016/j.abb.2017.04.001>
38. Kwak GH, Kim TH, Kim HY (2017) Down-regulation of MsrB3 induces cancer cell apoptosis through reactive oxygen species production and intrinsic mitochondrial pathway activation. *Biochem Biophys Res Commun* 483(1):468–474. <https://doi.org/10.1016/j.bbrc.2016.12.120>
39. Lee E, Kwak GH, Kamble K, Kim HY (2014) Methionine sulfoxide reductase B3 deficiency inhibits cell growth through the activation of p53-p21 and p27 pathways. *Arch Biochem Biophys* 547:1–5. <https://doi.org/10.1016/j.abb.2014.02.008>
40. Morel AP, Ginestier C, Pommier RM, Cabaud O, Ruiz E, Wicinski J, Devouassoux-Shisheboran M, Combaret V, Finetti P, Chassot C, Pinatel C, Fauvet F, Saintigny P, Thomas E, Moyret-Lalle C, Lachuer J, Despras E, Jauffret JL, Bertucci F, Guitton J, Wierinckx A, Wang Q, Radosevich-Robin N, Penault-Llorca F, Cox DG, Hollande F, Ansieau S, Caramel J, Birnbaum D, Vigneron AM, Tissier A, Charafe-Jauffret E, Puisieux A (2017) A stemness-related ZEB1-MSRB3 axis governs cellular pliancy and breast cancer genome stability. *Nat Med* 23(5):568–578. <https://doi.org/10.1038/nm.4323>
41. Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M, Campbell LL, Polyak K, Brisken C, Yang J, Weinberg RA (2008) The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 133(4):704–715. <https://doi.org/10.1016/j.cell.2008.03.027>
42. Polyak K, Weinberg RA (2009) Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer* 9(4):265–273. <https://doi.org/10.1038/nrc2620>
43. Manzotti G, Torricelli F, Benedetta D, Lococo F, Sancisi V, Rossi G, Piana S, Ciarrocchi A (2018) An epithelial-to-mesenchymal transcriptional switch triggers evolution of pulmonary Sarcomatoid carcinoma (PSC) and identifies Dasatinib as new therapeutic option. *Clinical cancer research : an official journal of the American Association for Cancer Research*. <https://doi.org/10.1158/1078-0432.CCR-18-2364>
44. Xie B, Lin W, Ye J, Wang X, Zhang B, Xiong S, Li H, Tan G (2015) DDR2 facilitates hepatocellular carcinoma invasion and metastasis via activating ERK signaling and stabilizing SNAIL1. *J Exp Clin Cancer Res : CR* 34:101. <https://doi.org/10.1186/s13046-015-0218-6>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.