Identification of Hub Genes Associated With Progression and Prognosis of Bladder Cancer by Integrated Bioinformatics Analysis.

Daoquan Liu
Wuhan University Zhongnan Hospital https://orcid.org/0000-0002-2628-0229
Jianhong Ma
Wuhan University Zhongnan Hospital
Sheng Wei
Traditional Chinese Medicine Hospital of Xishui
Jianmin Liu
Wuhan University Zhongnan Hospital
Mingzhou Li
Wuhan University Zhongnan Hospital
Weixiang He
Wuhan University Zhongnan Hospital
Qiaofeng Qian
Wuhan University Zhongnan Hospital
Yongying Zhou
Wuhan University Zhongnan Hospital
Yan Li
Wuhan University Zhongnan Hospital
Xun Fu
Wuhan University Zhongnan Hospital
Huan Liu
Wuhan University Zhongnan Hospital
Qian Wang
Wuhan University Zhongnan Hospital
Ping Chen
Wuhan University Zhongnan Hospital
Deqiang Xu
Wuhan University Zhongnan Hospital
Xinhuan Wang
Wuhan University Zhongnan Hospital
Michael E DiSanto

Cooper Medical School of Rowan University: Rowan University Cooper Medical School
Xinhua Zhang (zhangxinhuad@163.com)
Zhongnan Hospital of Wuhan University https://orcid.org/0000-0003-0267-428X

Primary research

Keywords: bladder cancer, differential gene expression analysis, weighted gene co-expression network analysis, differential co-expression genes, biomarkers

Posted Date: January 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-139588/v1

Abstract

Background: Bladder cancer (BLCA) is the most popular malignant carcinomas in genitourinary system which has a high incidence and is prone to relapse. However, the molecular mechanism of BLCA remains to be unclear. Moreover, there is still a shortage of effective biomarkers that can predict progression and prognosis of BLCA. The objective of current study is to screen significant genes as biomarkers to forecast the progression and prognosis of BLCA patients.

Methods: Gene expression profile downloaded from TCGA database and GEO database was used. Differential gene expression analysis and WGCNA were conducted to identify differential co-expression genes. In addition, GO enrichment analysis and KEGG pathway analysis were used to explore the functions of these genes. Moreover, PPI network, OS and DFS were used to identify survival-related hub genes. Finally, the expression levels of these genes were validated by qRT-PCR and HPA database.

Results: About 124 differential co-expression genes were identified. And these genes were mainly enriched in muscle system process and muscle contraction (BP), contractile fiber, myofibril, sarcomere, focal adhesion and cell-substrate junction (CC) and actin binding (MF) in GO enrichment analysis, while enriched in vascular smooth muscle contraction, focal adhesion, cardiac muscle contraction, hypertrophic cardiomyopathy, dilated cardiomyopathy and regulation of actin cytoskeleton in KEGG analysis. Furthermore, five survival-related hub genes (MYH11, ACTA2, CALD1, TPM1, MYLK) were identified via overall OS and DFS. In addition, the expression levels of the five survival-related genes were upregulated with the procession of BLCA, such as grade, stage and TNM stage. Finally, all survival-related hub genes were found to be down-regulated in BLCA via qRT-PCR and HPA database.

Conclusions: Our current study verified five new key genes in BLCA, which could help us better understand the pathogenesis of BLCA. And these five hub genes may be involved in the development and progression of BLCA and served as potential biomarkers.

Background

Bladder cancer (BLCA) which has a high incidence and is prone to relapse, is the most popular malignant carcinomas in genitourinary system(1) with 549,393 new cases and 199,922 deaths in 2018(2), and greatly threatens people's lives and health. In the meantime, it also imposes substantial economic burdens on patients who suffer from this illness. It is known that BLCA is caused by genetic abnormalities and external risk factors including occupational exposure to carcinogen, smoking, coffee, intake of analgesics and artificial sweeteners, infection, inflammation, bladder calculus, chemotherapy and radiation(3). The vast majority of BLCA is of urothelial histology and could be divided into non-muscle invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC) according to tumor stage(4). Among newly diagnosed BLCA, approximately 75% are NMIBC(5, 6). Transurethral resection of bladder tumor (TURBT) and intravesical chemotherapy are the standard for initial plan of action when treating NMIBC(7, 8). The remaining 25% are MIBC and need either radical cystectomy or chemotherapy.

However, some patients had poor outcomes(5, 9). In addition, approximately 70% of NMIBC patients have a high rate of recurrence and about 30% of NMIBC patients may even progress to MIBC after local therapy(10). Therefore, long-term follow-up with cystoscopy and computed tomography (CT) scan was required(11), which greatly increased patient's economic burdens. Currently, effective biomarkers that can accurately forecast the progression and prognosis of BLCA are still scarce. Therefore, it is essential to screen out new biomarkers to improve outcomes of BLCA patients.

With the quick development of high-throughput sequencing technology and gene microarray, thousands of gene data associated with diseases could be identified rapidly, thus bioinformatics has obvious advantages in exploring the pathogenesis of diseases and identifying potential biomarkers. Weighted gene co-expression network analysis (WGCNA) can demonstrate gene function and gene association(12). And it has been widely used to analyze gene expression data. Genes with similar expression patterns can be incorporated into the same modules and analyzed whether they are corelated to clinical trait information(13). On the other hand, differential gene expression analysis can determine quantitative changes of mRNA levels between tumors and normal tissues to screen potential biomarkers(14).

With regarding to BLCA, many potential biomarkers were identified with bioinformatics, for example, CDC20, ASPM, TAGLN, ACTA2, MYH11, CALD1, MYLK, GEM, PRELP, TPM2, and OGN(15-19). In addition, several signaling pathways involved in BLCA, including vascular smooth muscle contraction, hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), focal adhesion, cGMP-PKG, cAMP, MAPK and PI3K-AKT, were screened out(15),(16). However, most previous studies only used the cancer genome atlas (TCGA) database or Gene Expression Omnibus (GEO) database when WGCNA was applied (13, 18). Moreover, although there were studies based on TCGA and GEO database, only differential gene expression analysis was performed(16, 17). It was the first time to combine WGCNA with differential gene expression analysis to identify biomarkers of BLCA based on TCGA database and GEO database.

In present study, we used gene expression profiles downloaded both from TCGA database and GEO database to uncover differential co-expression genes. Then GO functional enrichment analysis, KEGG pathway analysis, PPI network analysis, OS analysis and DFS analysis were performed to identify survival-related hub genes. After that we validated expression level of survival-related hub genes by qRT-PCR and the Human Protein Atlas (HPA) database. Finally, we explored the association between clinical information, including age, gender, grade, stage and TNM stage, and hub genes' expression.

Methods

Workflow

A flow diagram of current study was shown in **Figure 1**. First, we downloaded gene expression profiles from TCGA database and GEO database to screen DEGs. Next, we constructed a co-expression network by WGCNA to identify differential co-expression genes. After that, we perform GO enrichment analysis and KEGG pathway analysis for differential co-expression genes. Additionally, we constructed a PPI

network to screen hub genes and identified survival-related hub genes via OS and DFS analysis. Then, we validated the expressions of survival-related hub genes by HPA database, GEPIA database and qRT-PCR. Finally, we analyzed the association between the expressions of survival-related hub genes and clinical information of BLCA patients.

Datasets from TCGA and GEO database

We downloaded the gene expression profiles of BLCA and matched clinical information from TCGA (https://portal.gdc.cancer.gov/) database. There were 430 samples that included 411 bladder cancer and 19 normal tissues and 19,645 genes. Then we downloaded the GTF configuration file from the Ensembl database (http://asia.ensembl.org/index.html) to match probes with corresponding genes. Probes with more than one gene were eliminated and the average value was calculated out for these genes corresponding to more than one probes. Genes of low read counts are usually excluded for further analysis, so we set the genes with a cpm (count per million) ≥ 1 in our current study. Ultimately, 14,716 genes were used for further analysis.

In addition, we downloaded the normalized expression profiles of GSE13507 from GEO database (https://www.ncbi.nlm.nih.gov/geo/), another gene expression profiles of BLCA which were studied with the GPL6102 platform Illumina human-6 v2.0 expression beadchip. GSE13507 consisted of 188 BLCA samples including 165 primary bladder cancer samples and 23 recurrent non-muscle invasive tumor tissues, and 68 normal samples containing 58 normal looking bladder mucosae surrounding cancer which were histologically confirmed normal and 10 normal bladder mucosae. Probes were converted to the gene symbol based on an annotation file which was provided by the manufacturer and repeated probes for the same gene were removed. Therefore, about 24,323 genes were selected for the subsequent analysis.

Differentially expressed genes (DEGs) analysis

The R package “limma”(20) was used to screen DEGs in TCGA-BLCA and GSE13507 dataset between BLCA and normal tissues. We set the adj.P < 0.05 and $|logFC| \geq 1$ as the threshold for screening DEGs. Ultimately, the DEGs selected from TCGA-BLCA and GSE13507 dataset were visualized as a volcano plot by using the R package “ggplot2”, while as a heatmap plot by using “pheatmap”.

Construction of weighted co-expression network and identification of key modules

We kept the expression profiles of TCGA-BLCA and GSE13507 qualified and then constructed gene co-expression networks by “WGCNA” package in R(12) for these data. To build standard scale-free networks using the pickSoftThreshold function, soft powers $\beta = 5$ and 11 were selected. Next, we used Pearson’s correlation matrices performed for all pairs of genes to construct adjacency matrixes. Then the adjacency matrixes were turned into topological overlap matrixes (TOM) as well as the corresponding dissimilarity (1-TOM). At the same time, average linkage hierarchical clustering was conducted to classify genes with similar expression patterns into gene modules. To identify functional modules in co-expression networks,

we firstly quantized the correlation between module eigengenes and clinical trait information, and further quantified the relationship by calculating Gene Significance (GS). We got the average value of GS of all the genes in a module, called Module Significance (MS) via the data processing. Module with an absolute MS ranking first in all modules was considered candidate relevant to clinical traits. Finally, the module highly correlated with clinical trait information of bladder cancer was selected for further analysis.

Interaction between the modules of interest and DEGs

After DEGs analysis and WGCNA analysis, we obtained several differential co-expression genes between DEGs and co-expression genes that were extracted from the co-expression network to look for potential prognosis genes via the R package “VennDiagram”(21). And the result was present as a Venn diagram.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis

GO enrichment analysis and KEGG pathway analysis were performed for differential co-expression genes via R package “clusterProfiler”(22). Gene sets with $p < 0.05$ were referred to be significantly enriched. GO analysis consist of biological process (BP), cellular component (CC) and molecular function (MF), only the top ten terms with $p < 0.05$ were selected. While in KEGG analysis, we selected a maximum of 30 terms with $p < 0.05$.

Construction of PPI and screening of hub genes

In current study, we used Search Tool for the Retrieval of Interacting Genes (STRING) database (http://www.string-db.org/)(23) which is designed for predicting protein–protein interactions (PPI) to construct a PPI network of differential co-expression genes. We defined these genes with minimum required interaction score greater than 0.7 as hub genes, and the network diagram was visualized by Cytoscape (v3.8.0)(24). Maximal Clique Centrality (MCC) algorithm was efficient to find hub nodes. We calculated MCC of each node by CytoHubba(25), a plugin in Cytoscape, and finally we chose the top 10 MCC values as hub genes.

Verification of the expression patterns and the prognostic values of hub genes

We confirmed the hub genes’ expression pattern in BLCA and normal tissues based on GEPIA database (http://gepia.cancer-pku.cn/)(26). The expression level of each hub gene was shown in a box plot graph. In our study, patients whose clinical data of survival time and survival status were incomplete were excluded. According to the median expression value of hub genes, we divided patients into two groups. After that, we performed a Kaplan-Meier survival analysis to explore the relationship between OS and hub genes in patients using the R package “survival”. Furthermore, the association between DFS and hub genes was conducted in GEPIA database. Finally, it should be emphasized that genes with significant p value ($p < 0.05$) in OS analysis would be thought as survival associated genes. In addition, we divided the bladder cancer patients obtained from TCGA database into two groups according to age, gender, grade, stage and TNM stage using dichotomy, then we explored the association between these clinical information and hub genes’ expression via R package “ggpubr”.

Validation of protein expressions of survival-related hub genes by HPA database

We downloaded immunohistochemistry (IHC) pictures from HPA (https://www.proteinatlas.org/) database to validate the protein expressions of hub genes between BLCA and normal tissues.

Validation of mRNA expressions of survival-related hub genes by qRT-PCR

We further collected bladder cancer and matched paracancerous tissues from Zhongnan Hospital, Wuhan University, to verify the expression of hub genes. Total RNA from tissues was isolated using Hipure Total RNA Mini Kit (Cat. R4111-03, Magen) according to the manufacturer's instruction. After that, the quantity of isolated RNA was measured by a NanoDrop® ND-1000 UV-Vis spectrophotometer (Thermo Scientific). The cDNA was synthesized using 1 μg total RNA by ABScript II RT Master Mix for qPCR (Cat.RK20402, Abclonal). Each qPCR reaction was conducted with 10 μL 2X Universal SYBR Green Fast qPCR Mix (Cat.RK21203, Abclonal), 7 μL ddH_2O, 1 μL cDNA, 1 μL forward primer and 1 μL reverse primer. Values were normalized for amplified GAPDH alleles. Primer sequences were listed as follows:

TPM1: 5'-GCCGACGTAGCTTCTCTGAAC-3', 5'-TTTGGGCTCGACTCTCAATGA-3'; MYLK: 5'-CCCGAGGTTGTCTGGTTCAAA-3', 5'-GCAGGTGTACTTGGCATCGT-3'; CALD1: 5'-TCGACCCAACAATAACAGATGC-3', 5'-TCTCGTATCTTTCTTGGCGACT-3'; MYH11: 5'-CGCCAAGAGACTCGTCTGG-3', 5'-TCTTTCCCAACCGTGACCTTC-3'; ACTA2: 5'-GTGTTGCCCCTGAAGAGCAT-3', 5'-GCTGGGACATTGAAAGTCTCA-3'; GAPDH: 5'-ATGGAGAAGGCTGGGGCTC-3', 5'-AAGTTGTCATGGATGACCTTG-3'.

Statistical analysis

R software 4.0.2 was used for all statistical analyses. Two-tailed Student's t-tests were used to assess the statistical significance of differences between the groups. Statistical significance was considered as $p < 0.05$.

Results

Differentially expressed genes (DEGs) analysis

According to the cut-off criteria of $|logFC| \geq 1.0$ and adj. $P < 0.05$, a total of 875 DEGs including 236 upregulated genes and 639 down-regulated genes in the TCGA-BLCA dataset (**Figure 2B**) and 464 DEGs which consisted of 77 upregulated genes and 387 down-regulated genes in the GSE13507 dataset (**Figure 2D**) were identified to be dysregulated in bladder cancer tissues by R package “limma”. And heatmaps were shown in **Figure 2A, 2C**.

Construction of weighted co-expression network and identification of key modules

The gene co-expression networks were constructed in TCGA-BLCA and GSE13507 via R package “WGCNA” to search the key modules. Four hundred and nine BLCA samples with clinical information were

included for the co-expression analysis in TCGA-BLCA dataset while 256 samples in GSE13507 dataset. Each module was assigned different colors, a sum of 13 modules in TCGA-BLCA dataset and 10 modules in GSE13507 dataset were identified in current study. Genes without being classified into any other module formed the grey module, which was abandoned for further analysis (**Figure 3A, 3C**). Then we evaluated the correlations between modules and clinical traits which were present in the form of heatmap named module-trait relationships (**Figure 3B, 3D**). Among these modules, the brown module including 1,018 co-expression genes in TCGA-BLCA dataset and the turquoise module consisting of 2,102 co-expression genes were found to have highest correlation coefficients with normal tissues (brown module: r = 0.59, p = 1e-41; turquoise module: r = 0.52, p = 6e-19).

Interaction between the interested modules and DEGs

As shown in **Figure 4**, one hundred and twenty-four differential co-expression genes were identified between DEGs and co-expression genes via the R package “VennDiagram” for further analysis.

GO enrichment analysis and KEGG pathway analysis

In order to figure out the potential functions of the 124 differential co-expression genes, GO enrichment analysis and KEGG pathway analysis were conducted by R package “clusterProfiler”. The 124 differential co-expression genes can be categorized into 3 functional subgroups: biological process (BP), cellular component (CC) and molecular function (MF). As shown in **Figure 5A**, genes in the BP group were mainly enriched in muscle system process and muscle contraction, while genes in the CC group were mainly involved in contractile fiber and myofibril. In addition, actin binding was suggested to be related to the 124 differential co-expression genes in the MF group. Moreover, it demonstrated that these genes were mainly involved in vascular smooth muscle contraction and focal adhesion according to KEGG pathway analysis (**Figure 5B**).

PPI network construction and hub genes identification

PPI network of 124 differential co-expression genes, including 124 nodes and 94 edges, was further established by STRING database (**Figure 6A**). Hub genes were defined as minimum interaction score greater than 0.7, then these hub genes were visualized with Cytoscape software (v3.8.0) (**Figure 6B**). Eventually, the top 10 genes with the highest degree of connectivity were chosen as hub genes of BLCA. These hub genes included myosin heavy chain 11 (MYH11), alpha-smooth muscle actin (ACTA2), tropomyosin 2 (TPM2), actin gamma 2 (ACTG2), caldesmon 1 (CALD1), myosin light chain 9 (MYL9), tropomyosin 1 (TPM1), myosin light chain kinase (MYLK), sorbin and SH3 domain containing 1 (SORBS1) and leiomodin 1 (LMOD1), which might play a critical role in the progression of BLCA.

Verification of the expression patterns, the prognostic values and mRNA expression of hub genes

Subsequently, the expression levels of the ten hub genes were substantiated in the GEPIA database, and all hub genes were downregulated significantly in BLCA compared with normal tissues (**Figure 7**). Additionally, overall survival (OS) analysis was performed via the R package “survival” based on the

mRNA sequence data and matched clinical information of BLCA patients. As shown in **Figure 8**, OS analysis demonstrated that higher expression levels of MYH11, ACTA2, CALD1, TPM1 and MYLK were associated with worse prognosis ($p < 0.05$), therefore we defined these five genes as survival-related hub genes. Meanwhile, we performed disease free survival (DFS) analysis using the GEPIA database. However, among these five survival-related genes, only the expression levels of CALD1 and MYLK were found to be associated with the prognosis (**Figure 9**). In addition to OS and DFS analysis, we used dichotomy to classify the bladder cancer patients obtained from TCGA database into two groups according to their age, gender, grade, stage and TNM stage. Then the associations between these clinical information and hub genes' expression were explored via R package "ggpubr". And it was revealed that the expression levels of the five survival-related genes were upregulated with the procession of BLCA, such as grade, stage and TNM stage (**Figure 10**). Finally, the protein levels of MYH11, ACTA2, CALD1, TPM1 and MYLK were determined using HPA database. Also, they were verified with qRT-PCR using our own tissues. Consistent with GEPIA database, the protein and mRNA expression levels of MYH11, ACTA2, CALD1, TPM1 and MYLK were significantly downregulated in the BLCA samples (**Figure 11**).

Discussion

In current study, a total of 124 differential co-expression genes were identified. These genes were mainly enriched in muscle system process, related to the movement, invasion and metastasis of bladder cancer cells. We further identified 10 hub genes (MYH11, ACTA2, TPM2, ACTG2, CALD1, MYL9, TPM1, MYLK, SORBS1, LMOD1) with constructing a PPI network. All 10 genes were downregulated in BLCA tissues. In addition, OS analysis showed 5 of the identified hub genes (MYH11, ACTA2, CALD1, TPM1, MYLK) were associated with worse prognosis of BLCA. Moreover, among these five genes, only CALD1 and MYLK were associated with disease-free survival (DFS). Our novel data indicated the 5 hub genes could be used as prognosis biomarkers of BLCA and vascular smooth muscle signaling may play critical roles in the progression of BLCA.

Previous studies either used single database or single method which lacked sufficiently accuracy and effectiveness. The present study, for the first time, combined WGCNA with differential gene expression analysis to identify biomarkers for the prognosis and progression of BLCA. We identified DEGs and co-expression genes both from TCGA and GEO database and interaction between DEGs and co-expression genes was further analyzed, with 124 differential co-expression genes found. Based on GO enrichment analysis, these genes were mainly enriched in biological process (BP) of muscle system process and muscle contraction, cellular component (CC) of contractile fiber, myofibril, sarcomere, focal adhesion and cell-substrate junction, and molecular function (MF) of actin binding. Similarly, KEGG pathway enrichment analysis showed vascular smooth muscle contraction, focal adhesion, cardiac muscle contraction, hypertrophic cardiomyopathy, dilated cardiomyopathy and regulation of actin cytoskeleton, were mainly involved. Both enrichment analysis for these dysregulated genes revealed that biomechanics, which were involved in cell deformability, adherence, migration, invasion and tumor metastasis, could play an essential role in the progression of BLCA. Indeed, there were several studies in breast cancer and pancreatic cancer, demonstrating biomechanics contributed greatly to tumor

progression and prognosis(27). However, there is no relevant studies about the biomechanics in BLCA. Therefore, it will be helpful for better understanding the pathogenesis of BLCA and better predicting the progression of BLCA when determine these dysregulated genes and related pathways.

Subsequently, we constructed a PPI network of the 124 differential co-expression genes and the top 10 genes which had highest degree of connectivity were chosen as hub genes, including MYH11, ACTA2, TPM2, ACTG2, CALD1, MYL9, TPM1, MYLK, SORBS1 and LMOD1. Interestingly, most of these hub genes were again associated with muscle system process. Indeed, OS analysis showed five (MYH11, ACTA2, CALD1, TPM1 and MYLK) of the identified hub genes were negatively corelated with survival. Thus, these five genes could contribute to the prognosis of BLCA. Furthermore, the expression levels of the five survival-related genes were upregulated with the procession of BLCA, such as grade, stage and TNM stage. However, all five survival-related genes were less expressed in BLCA tissues and higher expression levels of MYH11, ACTA2, CALD1, TPM1 and MYLK showed worse OS, which were consistent with a previous study(17). Different roles of tumor-suppressive genes in different biological processes may explain this phenomenon. It's well known that tumor-suppressive genes play a major role in the normal growth and differentiation of the cell and block the development of cancer(28). However, mutations, deletions or inactivation of tumor-suppressive genes lead to cancers(28). We speculate that the decreased expression of these genes in tumors is associated with the occurrence of tumors, while their expression increased with tumor progression is associated with tumor progression. Maybe these gene's alternations in BLCA tissues need to be examined in further experiments and it could help us better explore the pathogenesis of BLCA.

MYH11 (myosin heavy chain 11) which shows the highest degree of connectivity among hub genes encodes MYH11, a smooth muscle myosin belonging to the myosin heavy chain family(29). We know MYH11 is a contractile protein and is involved in muscle contraction via adenosine triphosphate hydrolysis(30). Furthermore, MYH11 was also participated in cell adhesion and migration(31, 32). Several previous studies have revealed that MYH11 is a biomarker of BLCA(15, 17), which is consistent with our conclusion. Additionally, the expression level of MYH11 has been reported to be dysregulated in colorectal cancer(33, 34), lung cancer(35), laryngeal cancer(36) and acute myeloid leukemia(37). Based on these researches above mentioned, we speculate that myosin-dependent contractile activity in non-muscle cells may be participated in cell adhesion and migration to promote tumor invasion. ACTA2 (alpha-smooth muscle actin) was reported to be downregulated in BLCA in the previous study, which meets our results(15, 17). ACTA2 participates in the maintenance of mechanical tension and cell shape, and is of great importance for tumor cell migration and invasion(38). Huang et al found that PGAM1 interacts with ACTA2 to facilitate cell motility and cancer cell migration(39). Lee et al revealed that the amplification of ACTA2 was significantly associated with early brain metastasis of lung adenocarcinoma(40), and they also uncovered that lung adenocarcinomas patients with high ACTA2 expression have increased distant metastasis and poor prognosis(41). Additionally, ACTA2 has been reported to accelerate the invasiveness and metastasis of breast cancer cells(42). As cell motility is critically dependent on the actin cytoskeleton, so ACTA2 maybe be important for migration, invasion and metastasis. CALD1 (caldesmon 1) encodes a calmodulin- and actin-binding protein and is involved in the

regulation of smooth muscle and nonmuscle contraction. It has been reported that CALD1 can regulate cell migration, invasion and proliferation and enhance stress fiber formation(43). Moreover, high expression of CALD1 in colon cancer showed a poor prognosis(44, 45). CALD1 was identified as a tumor-specific splicing variant in colorectal cancer, bladder cancer, prostate cancer and glioma in particular(46-48). Additionally, it has been suggested that *CALD1* may indicate general cancer-related splicing events(48). TPM1 (tropomyosin 1) is involved in the formation of stress fibers, and reduction in cell motility and migration. It has been reported to be downregulated in various tumors, including gastric cancer(49), bladder cancer(17), intrahepatic cholangiocarcinoma(50), renal cell carcinoma(51) and so on. Above mentioned studies demonstrated that TPM1 inhibited cell proliferation, migration, invasion, angiogenesis and metastasis as a tumor suppressor. Like CALD1, TPM1 also is a tumor-specific splicing variant, which has been verified in head and neck cancer, colon cancer, bladder cancer and prostate cancer(48, 52). MYLK (myosin light chain kinase) encodes myosin light chain kinase which is a calcium/calmodulin dependent enzyme. Previous studies have demonstrated that malignant transformation of normal cells and affect migratory and invasive properties of tumor cells are associated with altered MYLK expression in multiple cancers(53-57). Moreover, we also found that the expression levels of the five survival-related genes were upregulated with the process of BLCA, such as grade, stage and TNM stage, which further suggested the values of these survival-related genes for predicting the progression and prognosis of BLCA patients.

In summary, we discovered that the MYH11, ACTA2, CALD1, TPM1 and MYLK may be used as potential prognosis biomarkers for BLCA patients. And we also revealed vascular smooth muscle contraction may be considered as new therapeutic targets for treatment of BLCA. However, the above conclusions are obtained via bioinformatics analysis merely, further biological experiments are required to confirm the function and molecular mechanisms of these survival-related genes and signal pathways in BLCA.

Abbreviations

BLCA Bladder cancer

NMIBC Non-muscle invasive bladder cancer

MIBC Muscle-invasive bladder cancer

TURBT Transurethral resection of bladder tumor

CT Computed tomography

WGCNA Weighted gene co-expression network analysis

HCM Hypertrophic cardiomyopathy

DCM Dilated cardiomyopathy

TCGA The cancer genome atlas

GEO Gene expression omnibus

GO Gene ontology

KEGG Kyoto encyclopedia of genes and genomes

PPI Protein–protein interaction

OS Overall survival

DFS Disease free survival

qRT-PCR Quantitative real-time polymerase chain reaction

HPA Human protein atlas

cpm Count per million

DEGs Differentially expressed genes

TOM Topological overlap matrixes

GS Gene significance

MS Module significance

BP Biological process

CC Cellular component

MF Molecular function

STRING Search tool for the retrieval of interacting genes

MCC Maximal clique centrality

IHC Immunohistochemistry

MYH11 Myosin heavy chain 11

ACTA2 alpha-smooth muscle actin

TPM2 Tropomyosin 2

ACTG2 actin gamma 2

CALD1 caldesmon 1

MYL9 Myosin light chain 9

TPM1 Tropomyosin 1

MYLK Myosin light chain kinase

SORBS1 Sorbin and SH3 domain containing 1

LMOD1 Leiomodin 1

Declarations

Acknowledgments

We would like to express our appreciation for the pre-dominant technical support of Miss Danni Shan.

Author contributions

DQL conceived and designed the study. DQL, JHM, SW, JML and MZL performed the analysis procedures. DQL, JHM, SW, JML, MZL, WXH, QFQ, YYZ, YL, XF, HL and QW analyzed the results. PC, DQX and XHW contributed analysis tools. DQL, JHM, SW, JML and MZL contributed to the writing of the manuscript. All authors reviewed the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China under Grant Nos. 81160086, 81270843 and 81770757. Study funders have no responsibility for manuscript preparation, publishing decision, data collection, data analysis, or study design.

Availability of data

All available data were analyzed in this study. These can be found here: TCGA database (https://portal.gdc.cancer.gov/), Ensembl database (http://asia.ensembl.org/index.html), GEO database (https://www.ncbi.nlm.nih.gov/geo/), STRING database (http://www.string-db.org/), GEPIA database (http://gepia.cancer-pku.cn/) and HPA database (https://www.proteinatlas.org/).

Ethics approval

The use of these BLCA specimens was approved by the Ethics Committee at Zhongnan Hospital of Wuhan University, and informed consent was obtained from all patients.

Consent for publication

The informed consent obtained from study participants.

Competing interests

The authors declare that they have no competing interests.

References

1. Antoni S, Ferlay J, Soerjomataram I, Znaor A, Jemal A, Bray F. Bladder Cancer Incidence and Mortality: A Global Overview and Recent Trends. Eur Urol. 2017;71(1):96-108.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.
3. CG R. Benign and malignant bladder disorders: tumors of the Bladder: urothelial cancer: epidemiology, mortality, etiology and pathology. In Campbell-Walsh Urology, Wein A, KavoussiL R, Novick A, Partin AW, Peters CA, Eds Saunders, Philadelphia PA. 2016;4:2185-201.
4. Humphrey PA, Moch H, Cubilla AL, Ulbright TM, Reuter VE. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part B: Prostate and Bladder Tumours. Eur Urol. 2016;70(1):106-19.
5. Burger M, Catto JW, Dalbagni G, Grossman HB, Herr H, Karakiewicz P, et al. Epidemiology and risk factors of urothelial bladder cancer. Eur Urol. 2013;63(2):234-41.
6. Lodewijk I, Dueñas M, Rubio C, Munera-Maravilla E, Segovia C, Bernardini A, et al. Liquid Biopsy Biomarkers in Bladder Cancer: A Current Need for Patient Diagnosis and Monitoring. Int J Mol Sci. 2018;19(9).
7. Woldu SL, Bagrodia A, Lotan Y. Guideline of guidelines: non-muscle-invasive bladder cancer. BJU Int. 2017;119(3):371-80.
8. Babjuk M, Böhle A, Burger M, Capoun O, Cohen D, Compérat EM, et al. EAU Guidelines on Non-Muscle-invasive Urothelial Carcinoma of the Bladder: Update 2016. Eur Urol. 2017;71(3):447-61.
9. Alfred Witjes J, Lebret T, Compérat EM, Cowan NC, De Santis M, Bruins HM, et al. Updated 2016 EAU Guidelines on Muscle-invasive and Metastatic Bladder Cancer. Eur Urol. 2017;71(3):462-75.
10. Chamie K, Litwin MS, Bassett JC, Daskivich TJ, Lai J, Hanley JM, et al. Recurrence of high-risk bladder cancer: a population-based analysis. Cancer. 2013;119(17):3219-27.
11. Flaig TW, Spiess PE, Agarwal N, Bangs R, Boorjian SA, Buyyounouski MK, et al. NCCN Guidelines Insights: Bladder Cancer, Version 5.2018. J Natl Compr Canc Netw. 2018;16(9):1041-53.
12. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
13. Wang Y, Chen L, Ju L, Qian K, Liu X, Wang X, et al. Novel Biomarkers Associated With Progression and Prognosis of Bladder Cancer Identified by Co-expression Analysis. Front Oncol. 2019;9:1030.
14. Segundo-Val IS, Sanz-Lozano CS. Introduction to the Gene Expression Analysis. Methods Mol Biol. 2016;1434:29-43.

15. Ning X, Deng Y. Identification of key pathways and genes influencing prognosis in bladder urothelial carcinoma. Onco Targets Ther. 2017;10:1673-86.
16. Xu Y, Wu G, Li J, Li J, Ruan N, Ma L, et al. Screening and Identification of Key Biomarkers for Bladder Cancer: A Study Based on TCGA and GEO Data. Biomed Res Int. 2020;2020:8283401.
17. Hu J, Zhou L, Song Z, Xiong M, Zhang Y, Yang Y, et al. The identification of new biomarkers for bladder cancer: A study based on TCGA and GEO datasets. J Cell Physiol. 2019.
18. Xiong Y, Yuan L, Xiong J, Xu H, Luo Y, Wang G, et al. An outcome model for human bladder cancer: A comprehensive study based on weighted gene co-expression network analysis. J Cell Mol Med. 2020;24(3):2342-55.
19. Chen Q, Hu J, Deng J, Fu B, Guo J. Bioinformatics Analysis Identified Key Molecular Changes in Bladder Cancer Development and Recurrence. Biomed Res Int. 2019;2019:3917982.
20. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
21. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 2011;12:35.
22. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16(5):284-7.
23. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607-d13.
24. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.
25. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014;8 Suppl 4(Suppl 4):S11.
26. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017;45(W1):W98-w102.
27. Suresh S. Biomechanics and biophysics of cancer cells. Acta Biomater. 2007;3(4):413-38.
28. Kontomanolis EN, Koutras A, Syllaios A, Schizas D, Mastoraki A, Garmpis N, et al. Role of Oncogenes and Tumor-suppressor Genes in Carcinogenesis: A Review. Anticancer Res. 2020;40(11):6009-15.
29. Matsuoka R, Yoshida MC, Furutani Y, Imamura S, Kanda N, Yanagisawa M, et al. Human smooth muscle myosin heavy chain gene mapped to chromosomal region 16q12. Am J Med Genet. 1993;46(1):61-7.
30. Adelstein RS, Eisenberg E. Regulation and kinetics of the actin-myosin-ATP interaction. Annu Rev Biochem. 1980;49:921-56.
31. Krendel M, Mooseker MS. Myosins: tails (and heads) of functional diversity. Physiology (Bethesda). 2005;20:239-51.

32. Pollard TD, Weihing RR. Actin and myosin and cell movement. CRC Crit Rev Biochem. 1974;2(1):1-65.
33. Zhao B, Baloch Z, Ma Y, Wan Z, Huo Y, Li F, et al. Identification of Potential Key Genes and Pathways in Early-Onset Colorectal Cancer Through Bioinformatics Analysis. Cancer Control. 2019;26(1):1073274819831260.
34. Xi WD, Liu YJ, Sun XB, Shan J, Yi L, Zhang TT. Bioinformatics analysis of RNA-seq data revealed critical genes in colon adenocarcinoma. Eur Rev Med Pharmacol Sci. 2017;21(13):3012-20.
35. Nie MJ, Pan XT, Tao HY, Xu MJ, Liu SL, Sun W, et al. Clinical and prognostic significance of MYH11 in lung cancer. Oncol Lett. 2020;19(6):3899-906.
36. Su J, Zhang Y, Su H, Zhang C, Li W. A recurrence model for laryngeal cancer based on SVM and gene function clustering. Acta Otolaryngol. 2017;137(5):557-62.
37. Liu P, Tarlé SA, Hajra A, Claxton DF, Marlton P, Freedman M, et al. Fusion between transcription factor CBF beta/PEBP2 beta and a myosin heavy chain in acute myeloid leukemia. Science. 1993;261(5124):1041-4.
38. Lambrechts A, Van Troys M, Ampe C. The actin cytoskeleton in normal and pathological cell motility. Int J Biochem Cell Biol. 2004;36(10):1890-909.
39. Huang K, Liang Q, Zhou Y, Jiang LL, Gu WM, Luo MY, et al. A Novel Allosteric Inhibitor of Phosphoglycerate Mutase 1 Suppresses Growth and Metastasis of Non-Small-Cell Lung Cancer. Cell Metab. 2019;30(6):1107-19.e8.
40. Lee HW, Seol HJ, Choi YL, Ju HJ, Joo KM, Ko YH, et al. Genomic copy number alterations associated with the early brain metastasis of non-small cell lung cancer. Int J Oncol. 2012;41(6):2013-20.
41. Lee HW, Park YM, Lee SJ, Cho HJ, Kim DH, Lee JI, et al. Alpha-smooth muscle actin (ACTA2) is required for metastatic potential of human lung adenocarcinoma. Clin Cancer Res. 2013;19(21):5879-89.
42. Jeon M, You D, Bae SY, Kim SW, Nam SJ, Kim HH, et al. Dimerization of EGFR and HER2 induces breast cancer cell motility through STAT1-dependent ACTA2 induction. Oncotarget. 2017;8(31):50570-81.
43. Mayanagi T, Sobue K. Diversification of caldesmon-linked actin cytoskeleton in cell motility. Cell Adh Migr. 2011;5(2):150-9.
44. Yokota M, Kojima M, Higuchi Y, Nishizawa Y, Kobayashi A, Ito M, et al. Gene expression profile in the activation of subperitoneal fibroblasts reflects prognosis of patients with colon cancer. Int J Cancer. 2016;138(6):1422-31.
45. Cao R, Wang G, Qian K, Chen L, Ju L, Qian G, et al. TM4SF1 regulates apoptosis, cell cycle and ROS metabolism via the PPARgamma-SIRT1 feedback loop in human bladder cancer cells. Cancer Lett. 2018;414:278-93.
46. Bisognin A, Pizzini S, Perilli L, Esposito G, Mocellin S, Nitti D, et al. An integrative framework identifies alternative splicing events in colorectal cancer development. Mol Oncol. 2014;8(1):129-41.

47. Liu J, Li H, Shen S, Sun L, Yuan Y, Xing C. Alternative splicing events implicated in carcinogenesis and prognosis of colorectal cancer. J Cancer. 2018;9(10):1754-64.
48. Thorsen K, Sørensen KD, Brems-Eskildsen AS, Modin C, Gaustadnes M, Hein AM, et al. Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. Mol Cell Proteomics. 2008;7(7):1214-24.
49. Lin J, Shen J, Yue H, Cao Z. miRNA-183-5p.1 promotes the migration and invasion of gastric cancer AGS cells by targeting TPM1. Oncol Rep. 2019;42(6):2371-81.
50. Yang W, Wang X, Zheng W, Li K, Liu H, Sun Y. Genetic and epigenetic alterations are involved in the regulation of TPM1 in cholangiocarcinoma. Int J Oncol. 2013;42(2):690-8.
51. Wang J, Tang C, Yang C, Zheng Q, Hou Y. Tropomyosin-1 Functions as a Tumor Suppressor with Respect to Cell Proliferation, Angiogenesis and Metastasis in Renal Cell Carcinoma. J Cancer. 2019;10(10):2220-8.
52. Liang Y, Song J, He D, Xia Y, Wu Y, Yin X, et al. Systematic analysis of survival-associated alternative splicing signatures uncovers prognostic predictors for head and neck cancer. J Cell Physiol. 2019;234(9):15836-46.
53. Kim DY, Helfman DM. Loss of MLCK leads to disruption of cell-cell adhesion and invasive behavior of breast epithelial cells via increased expression of EGFR and ERK/JNK signaling. Oncogene. 2016;35(34):4495-508.
54. Yu HJ, Serebryannyy LA, Fry M, Greene M, Chernaya O, Hu WY, et al. Tumor stiffness is unrelated to myosin light chain phosphorylation in cancer cells. PLoS One. 2013;8(11):e79776.
55. Wang B, Yan Y, Zhou J, Zhou Q, Gui S, Wang Y. A novel all-trans retinoid acid derivatives inhibits the migration of breast cancer cell lines MDA-MB-231 via myosin light chain kinase involving p38-MAPK pathway. Biomed Pharmacother. 2013;67(5):357-62.
56. Cui WJ, Liu Y, Zhou XL, Wang FZ, Zhang XD, Ye LH. Myosin light chain kinase is responsible for high proliferative ability of breast cancer cells via anti-apoptosis involving p38 pathway. Acta Pharmacol Sin. 2010;31(6):725-32.
57. Léveillé N, Fournier A, Labrie C. Androgens down-regulate myosin light chain kinase in human prostate cancer cells. J Steroid Biochem Mol Biol. 2009;114(3-5):174-9.

Figures

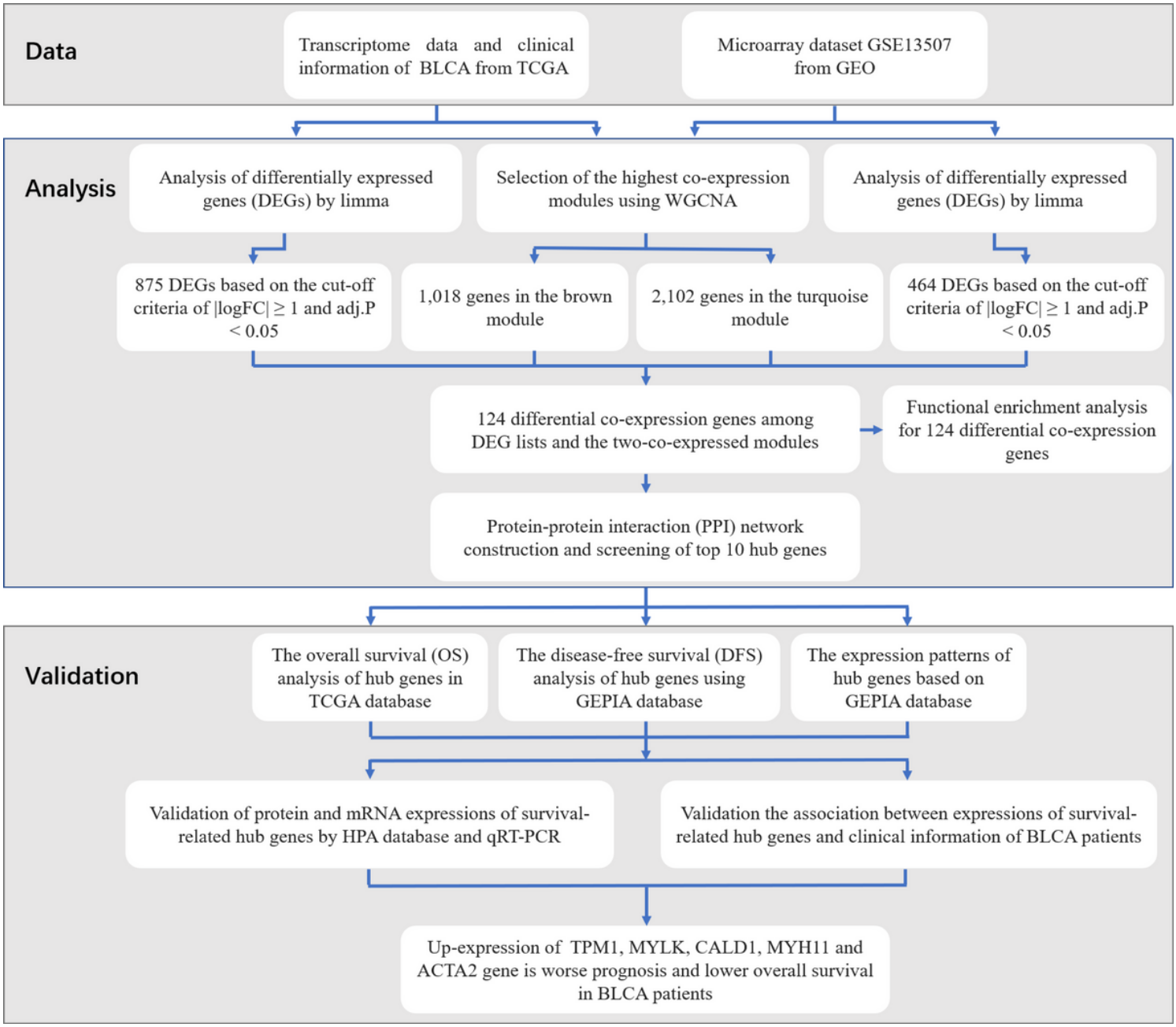

Figure 1

Flow diagram of this study.

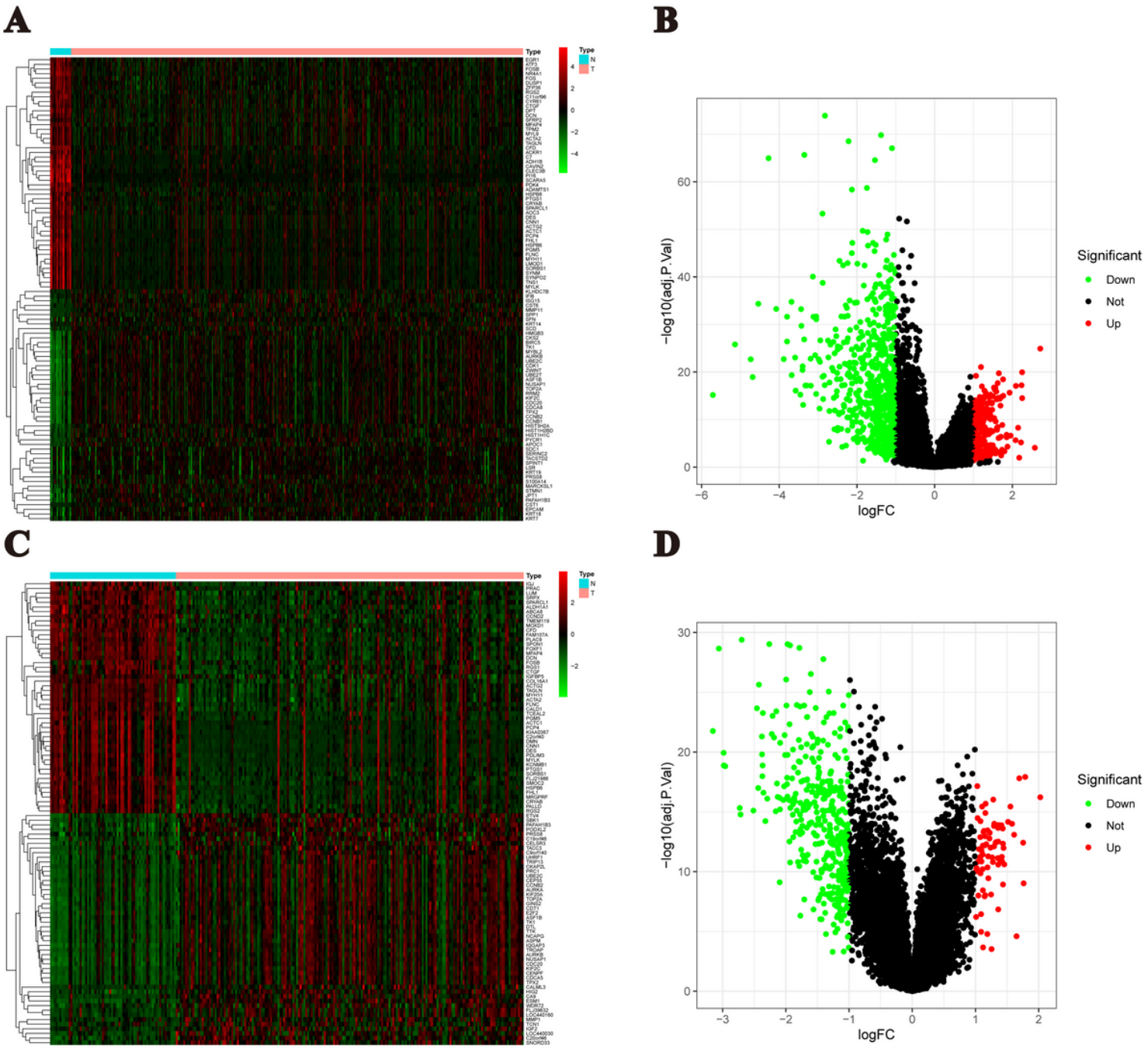

Figure 2

Identification of differentially expressed genes (DEGs) in the TCGA and GSE13507 datasets of BLCA. (A), (C) Volcano plot of DEGs in the TCGA and GSE13507 datasets. (B), (D) Heatmap plot of DEGs in the TCGA and GSE13507 datasets. The dotted line corresponds to logFC = 0. The red nodes represent upregulated genes with a logFC ≥ 1. The green nodes represent downregulated genes with a logFC ≤ -1. with the cut-off criteria of adj.P < 0.05 and |logFC| ≥ 1

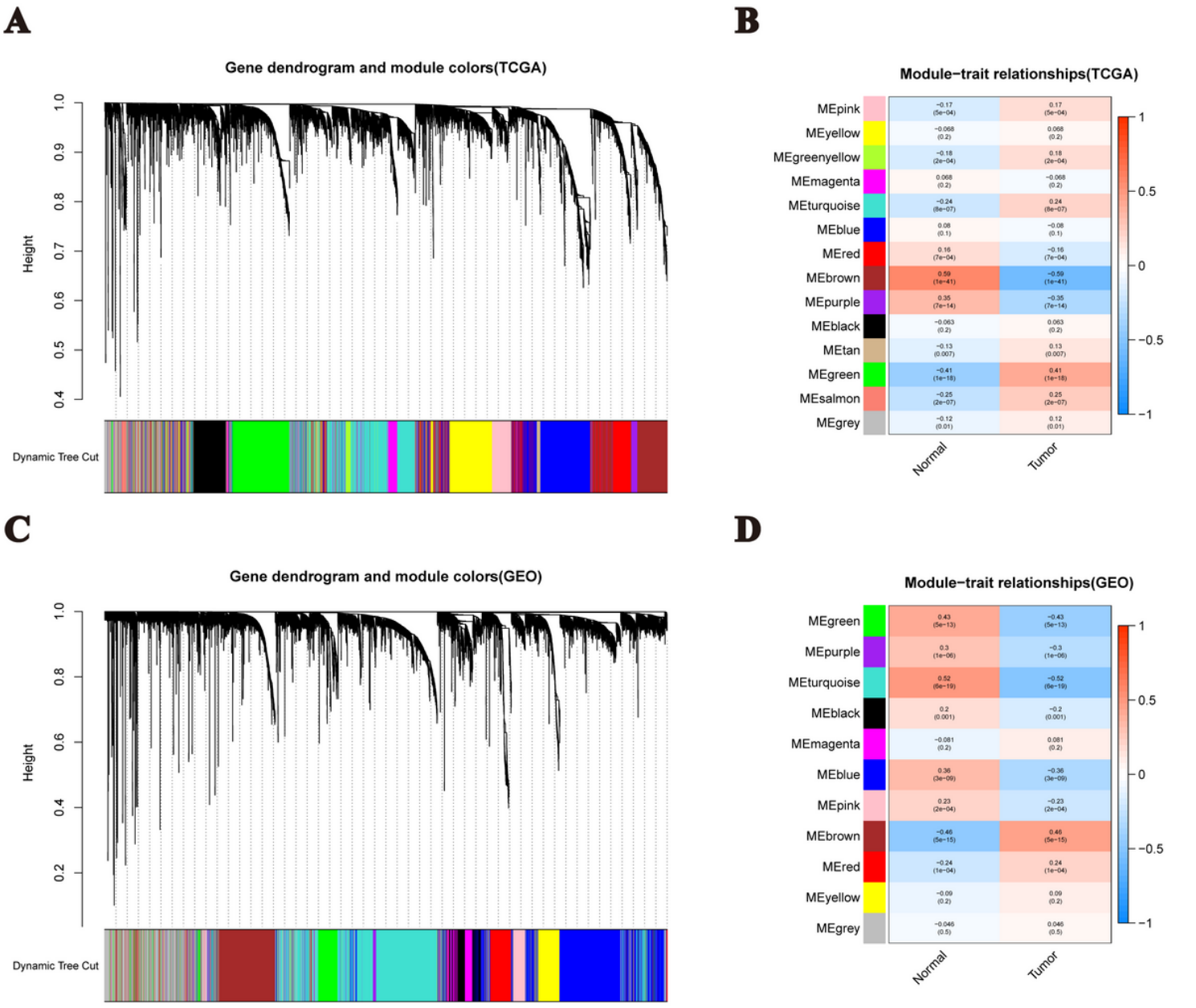

Figure 3

Identification of modules associated with the clinical information in the TCGA and GSE13507 datasets of BLCA. (A), (C) Gene dendrogram and module colors in the TCGA and GSE13507 datasets. Each module was assigned different colors. (B), (D) Module-trait relationships in the TCGA and GSE13507 datasets. Each row corresponds to a color module and column corresponds to a clinical trait (cancer and normal). Each cell contains the corresponding correlation and P-value.

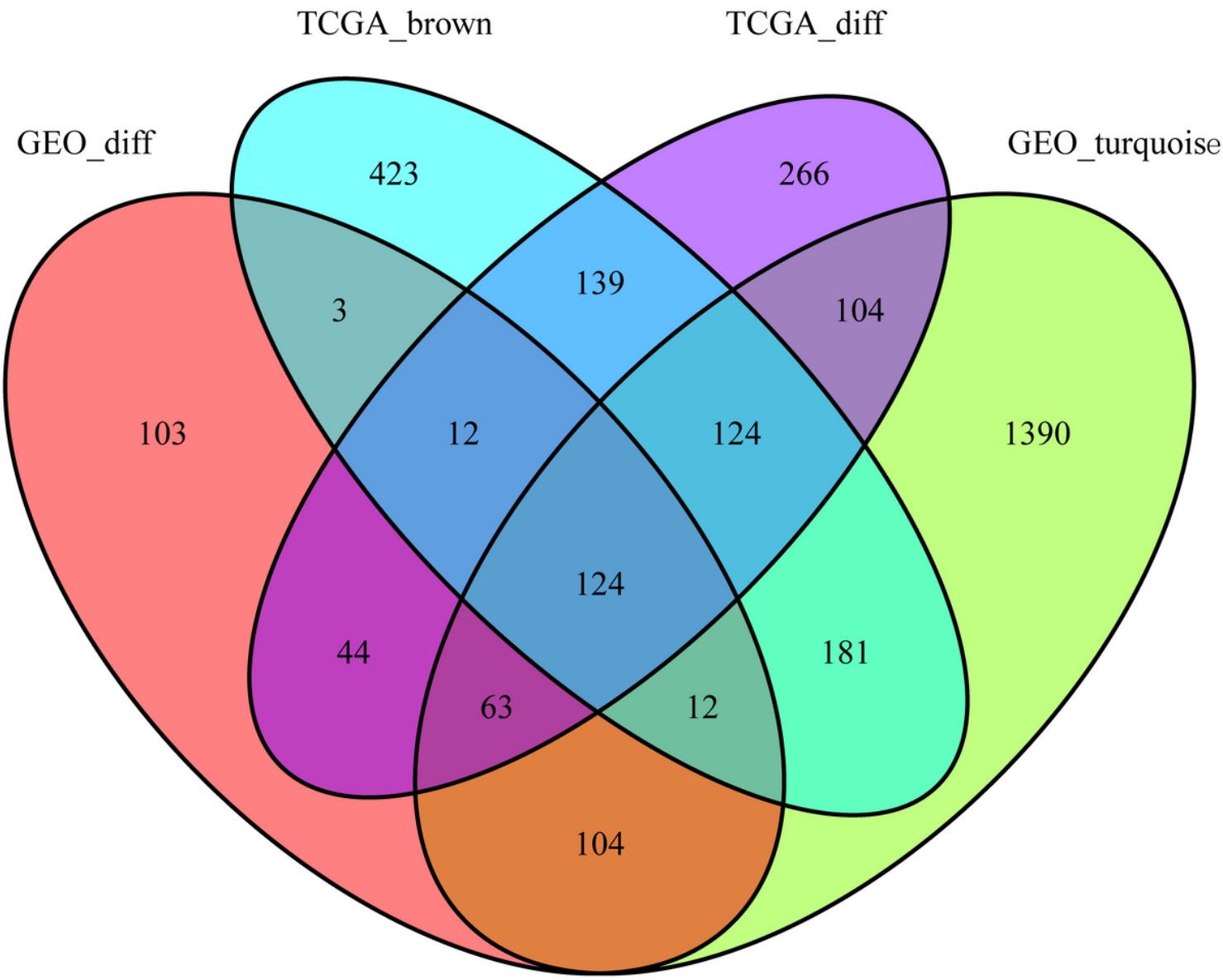

Figure 4

The Venn diagram of genes among DEG lists and co-expression module. In total, 124 differential co-expression genes in the intersection of DEG lists and two co-expression modules.

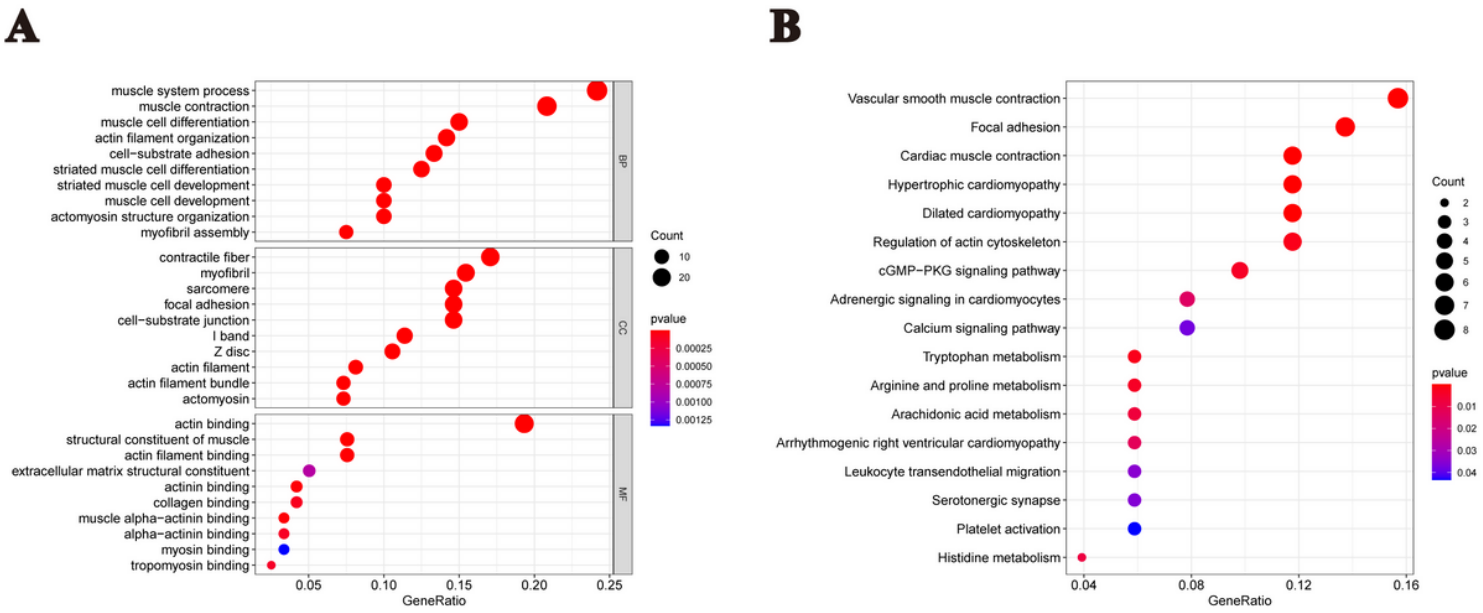

Figure 5

GO and KEGG pathway enrichment analysis for 124 differential co-expression genes. (A) GO enrichment analysis for 124 differential co-expression genes. (B) KEGG pathway enrichment analysis for 124 differential co-expression genes. The color represents the adjusted p-value, and the size of the spots represents the gene number.

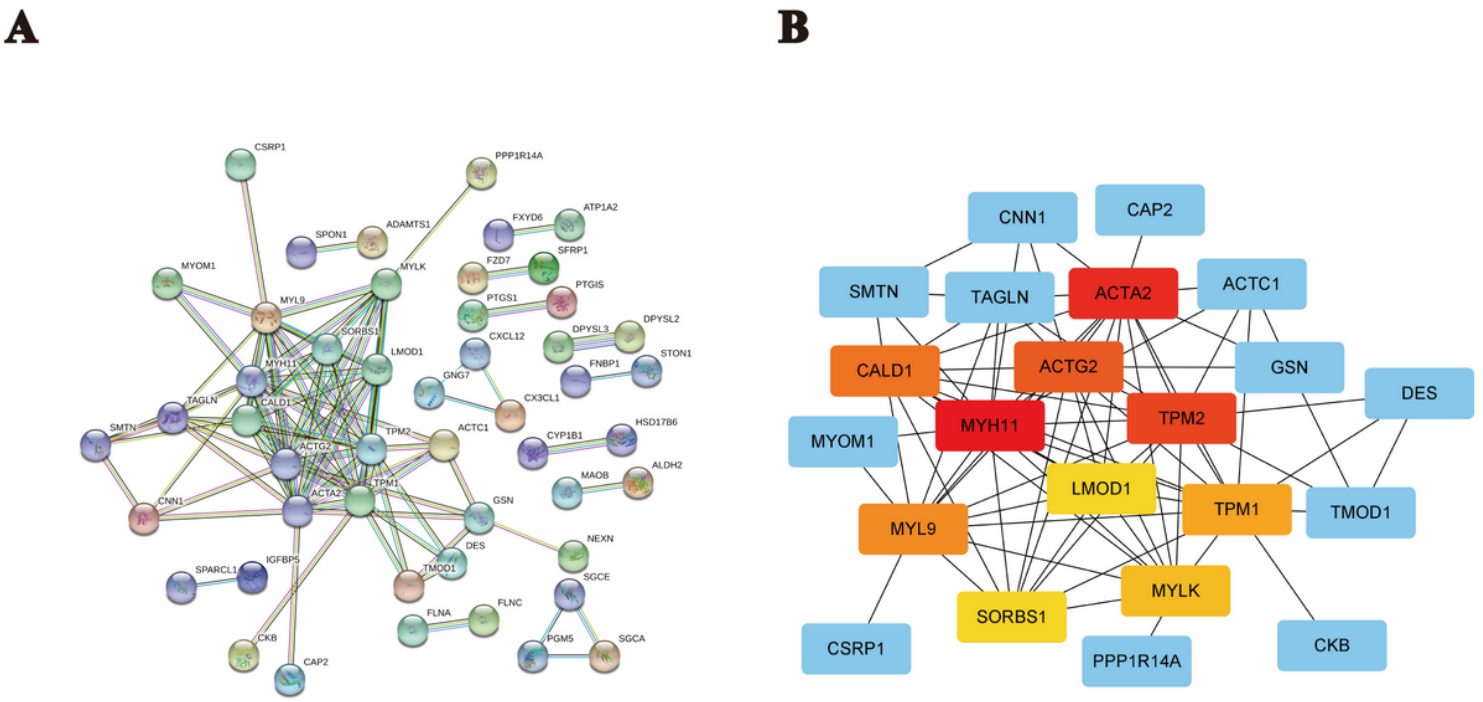

Figure 6

Visualization of the protein-protein interaction (PPI) network and hub genes. (A) PPI network of 124 differential co-expression genes. The round nodes represent the genes. Edges indicate interaction associations between nodes. (B) Identification of the hub genes from the PPI network using maximal clique centrality (MCC) algorithm. Edges represent the protein-protein associations. The red nodes represent genes with a high MCC score, while the yellow nodes represent genes with a low MCC score.

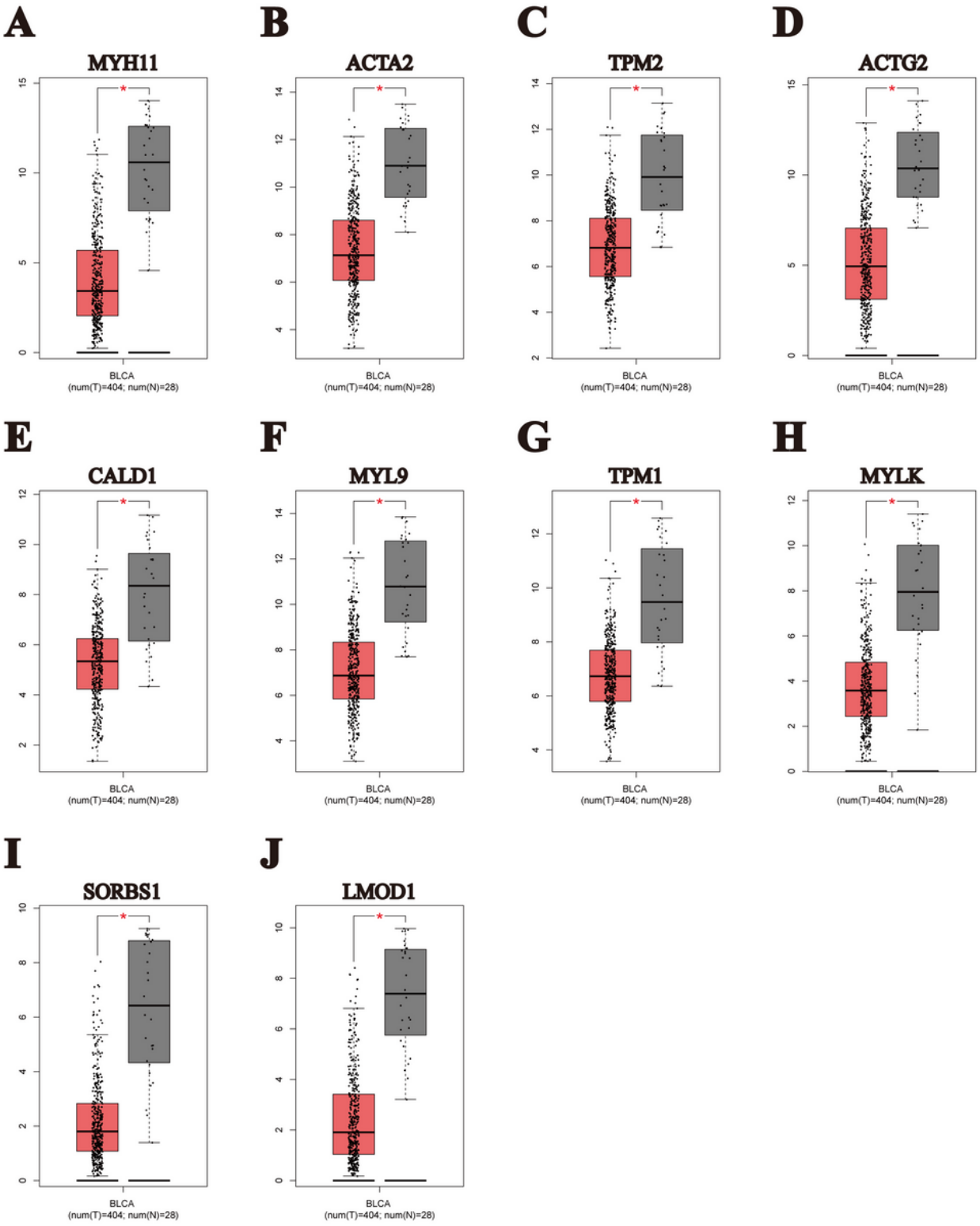

Figure 7

Validation of expression levels of these ten hub gens among BLCA and normal tissues from the TCGA database. (A) Gene expression value MYH11 among samples of TCGA. (B) Gene expression value ACTA2 among samples of TCGA. (C) Gene expression value TPM2 among samples of TCGA. (D) Gene expression value ACTG2 among samples of TCGA. (E) Gene expression value CALD1 among samples of TCGA. (F) Gene expression value MYL9 among samples of TCGA. (G) Gene expression value TPM1 among samples of TCGA. (H) Gene expression value MYLK among samples of TCGA. (I) Gene expression value SORBS1 among samples of TCGA. (J) Gene expression value LMOD1 among samples of TCGA.

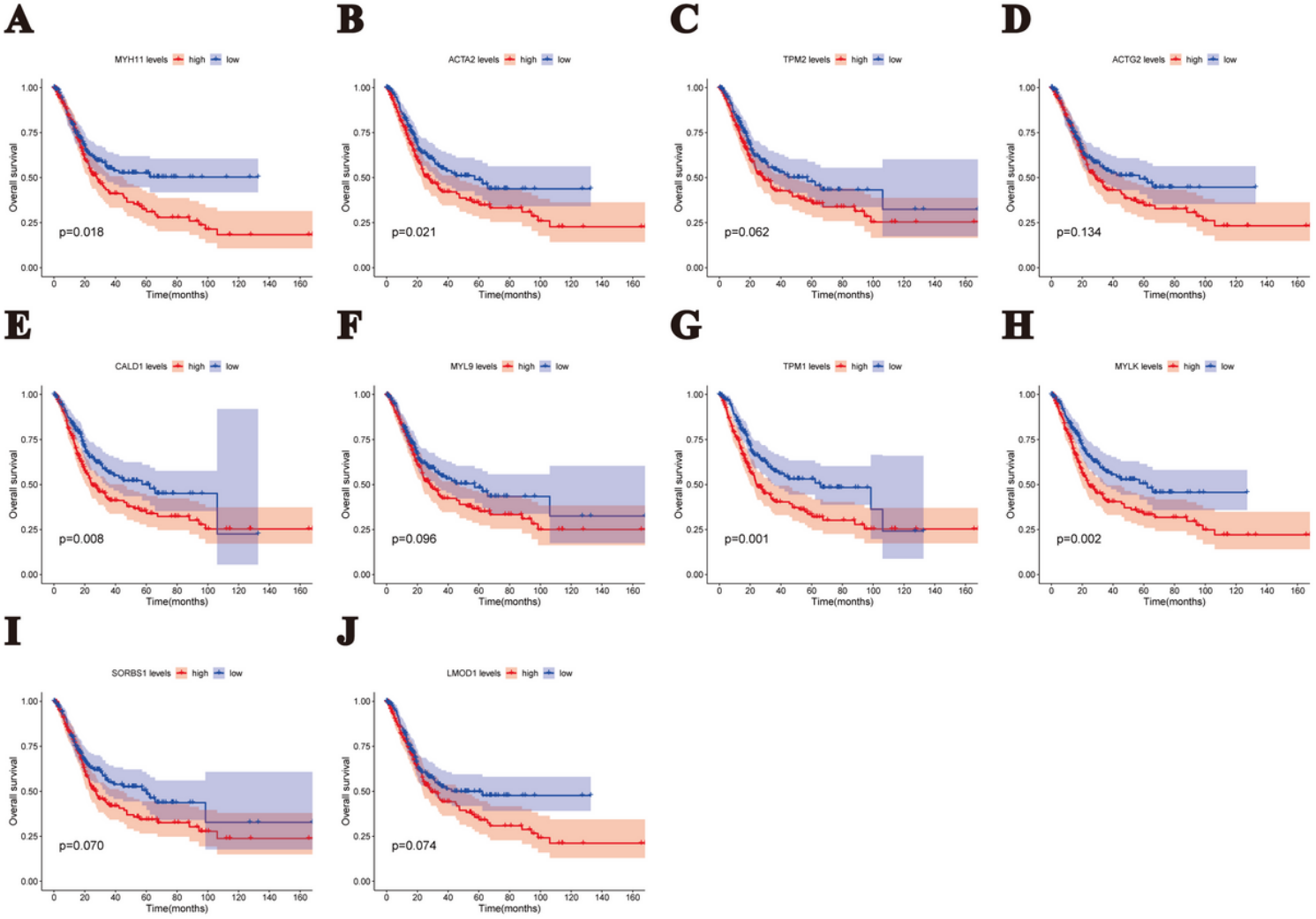

Figure 8

Overall survival (OS) analysis of 10 hub genes in BLCA patients based the TCGA database. (A) Survival analysis for MYH11 in BLCA. (B) Survival analysis for ACTA2 in BLCA. (C) Survival analysis for TPM2 in BLCA. (D) Survival analysis for ACTG2 in BLCA. (E) Survival analysis for CALD1 in BLCA. (F) Survival analysis for MYL9 in BLCA. (G) Survival analysis for TPM1 in BLCA. (H) Survival analysis for MYLK in BLCA. (I) Survival analysis for SORBS1 in BLCA. (J) Survival analysis for LMOD1 in BLCA. The patients were stratified into high-level group (red) and low-level group (green) according to median expression of the gene. Log-rank $p < 0.05$ was considered as statistical significance.

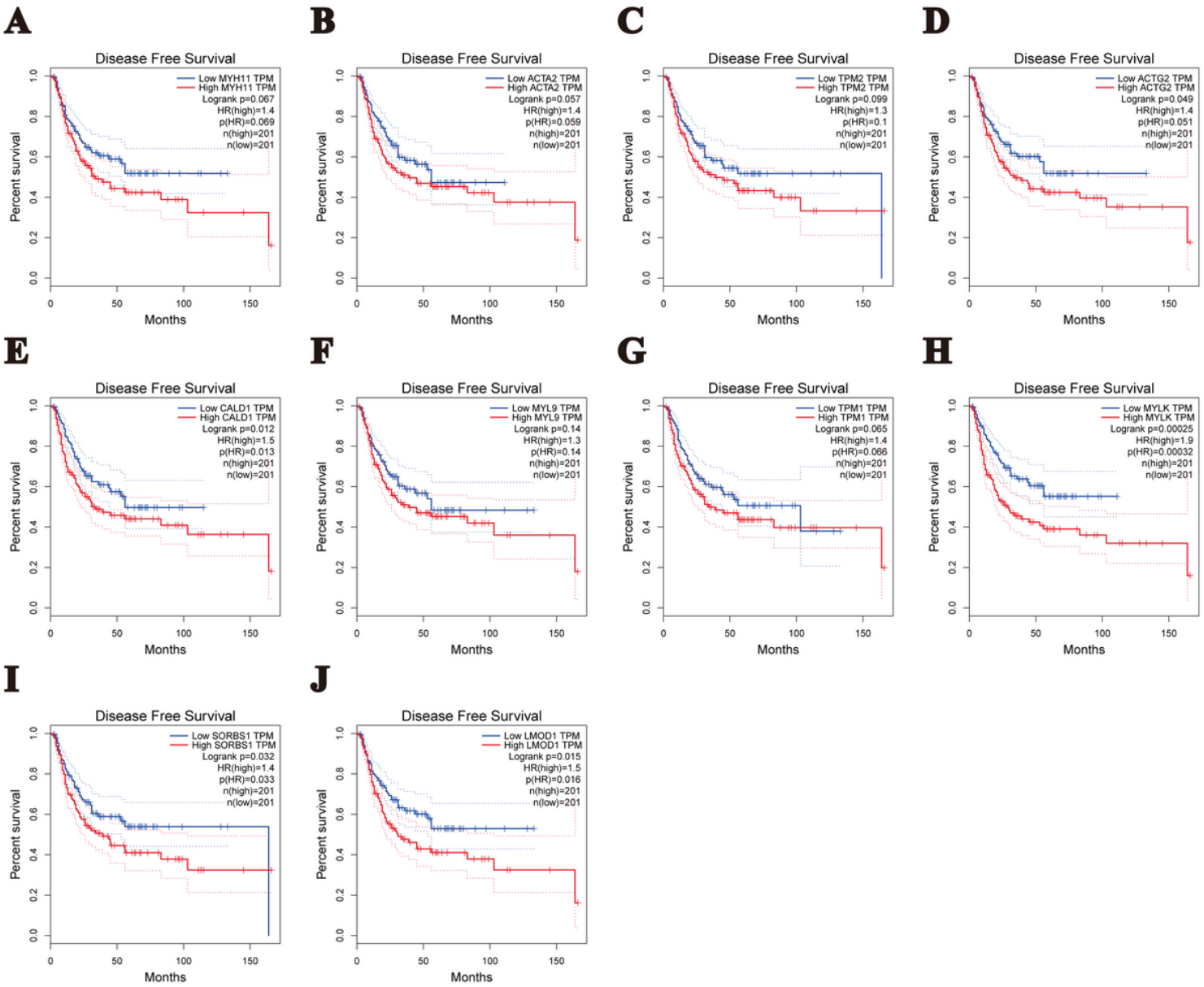

Figure 9

Disease free survival (DFS) analysis of 10 hub genes in BLCA patients from the GEPIA database. (A) Survival analysis for MYH11 in BLCA. (B) Survival analysis for ACTA2 in BLCA. (C) Survival analysis for TPM2 in BLCA. (D) Survival analysis for ACTG2 in BLCA. (E) Survival analysis for CALD1 in BLCA. (F) Survival analysis for MYL9 in BLCA. (G) Survival analysis for TPM1 in BLCA. (H) Survival analysis for MYLK in BLCA. (I) Survival analysis for SORBS1 in BLCA. (J) Survival analysis for LMOD1 in BLCA. The patients were stratified into high-level group (red) and low-level group (green) according to median expression of the gene. Log-rank $p < 0.05$ was considered as statistical significance.

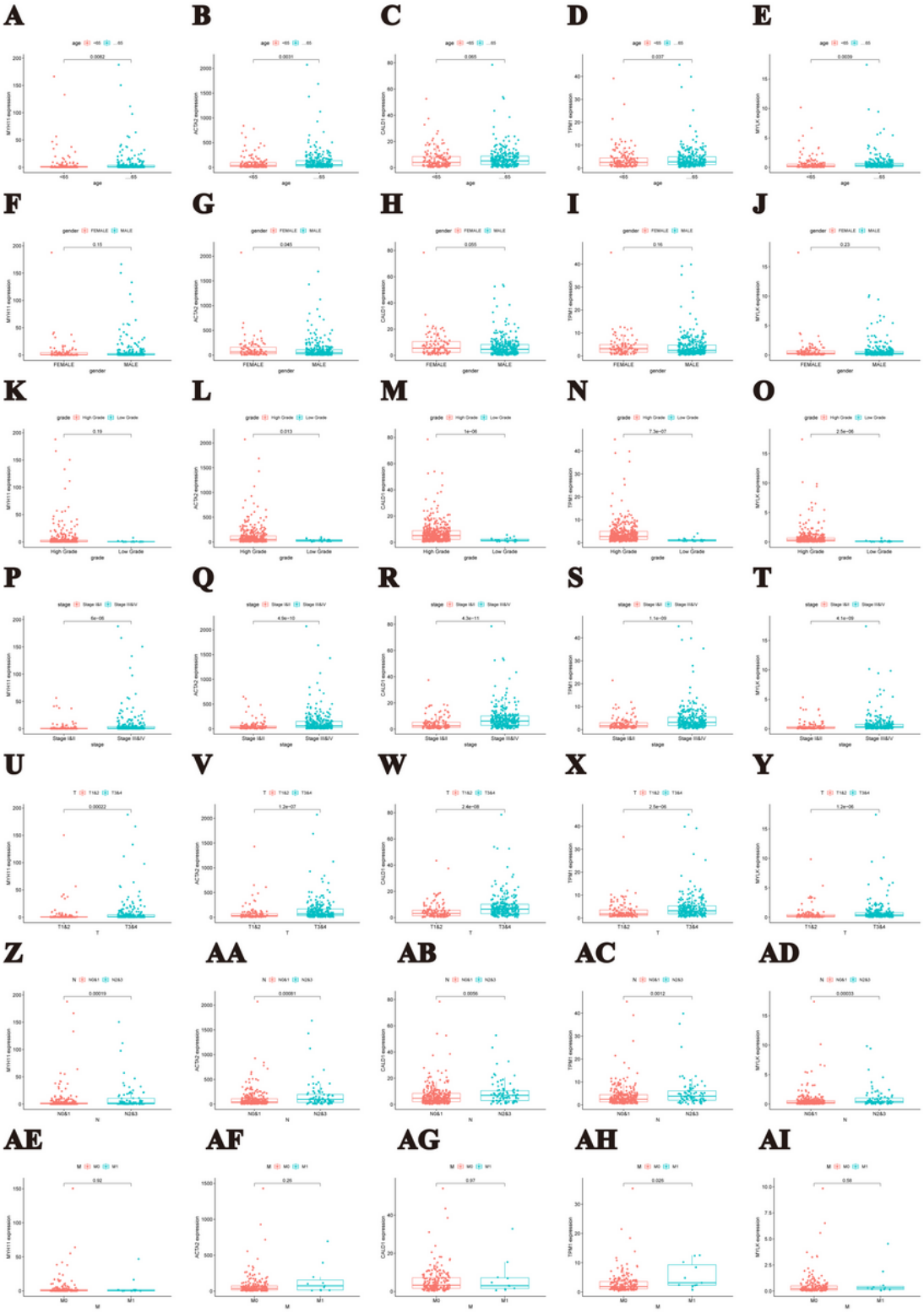

Figure 10

Association between five survival-related hub genes' expression and age, gender, stage, grade, TNM stage. (A-E) Association between five survival-related hub genes' expression (MYH11, ACTA2, CALD1, TPM1, MYLK) and age. The patients were stratified into < 65 group (red) and ≥ 65 group (green). (F-J) Association between five survival-related hub genes' expression (MYH11, ACTA2, CALD1, TPM1, MYLK) and gender. The patients were stratified into female group (red) and male group (green). (K-O)

Association between five survival-related hub genes' expression (MYH11, ACTA2, CALD1, TPM1, MYLK) and grade. The patients were stratified into high grade group (red) and low grade group (green). (P-T) Association between five survival-related hub genes' expression (MYH11, ACTA2, CALD1, TPM1, MYLK) and stage. The patients were stratified into stage □&□ group (red) and stage □&□ group (green). (U-Y) Association between five survival-related hub genes' expression (MYH11, ACTA2, CALD1, TPM1, MYLK) and T stage. The patients were stratified into T1&2 group (red) and stage T3&4 group (green). (Z-AD) Association between five survival-related hub genes' expression (MYH11, ACTA2, CALD1, TPM1, MYLK) and N stage. The patients were stratified into N0&1 group (red) and stage N2&3 group (green). (AE-AH) Association between five survival-related hub genes' expression (MYH11, ACTA2, CALD1, TPM1, MYLK) and M stage. The patients were stratified into M0 group (red) and stage M1 group (green).

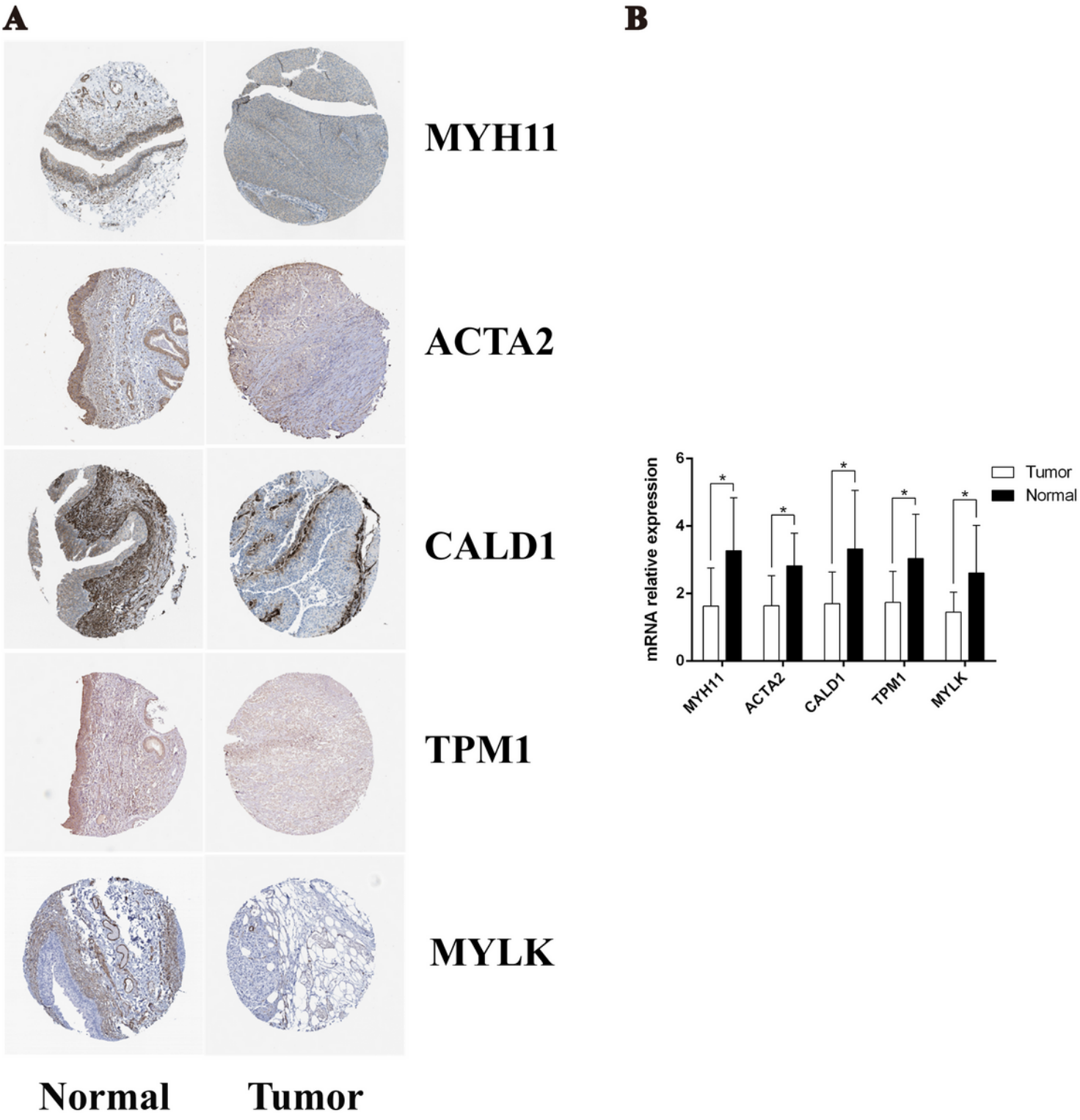

Figure 11

Verification of the protein and mRNA expression level of five survival-related hub genes from the human protein atlas (HPA) database and qRT-PCR. (A) Protein levels of MYH11, ACTA2, CALD1, TPM1, MYLK in BLCA and normal tissues from HPA database. (B) mRNA levels of MYH11, ACTA2, CALD1, TPM1, MYLK in BLCA and normal tissues from qRT-PCR.