Identification of Prognostic Biomarkers of Potential Hub Genes in Urothelial Carcinoma and Function in Microenvironment

Wei Chu
Qilu Hospital of Shandong University Qingdao https://orcid.org/0000-0001-6416-2226
Bing Zhang
Qilu Hospital of Shandong University Qingdao
Haifeng Gong
Qilu Hospital of Shandong University Qingdao
Qianqian Zhao
Qilu Hospital of Shandong University Qingdao
Jun Chen
Qilu Hospital of Shandong University Qingdao
Hong Ji (hezejihong@sina.com)
Qilu Hospital of Shandong University Qingdao https://orcid.org/0000-0003-1996-8989

Research

Keywords: Urothelial carcinoma, gene expression profile, bioinformatics, tumor microenvironment, immune infiltration

Posted Date: May 14th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-502631/v1

Abstract

Background Urothelial carcinoma (UC) is the most common histological type of urinary system. In the past decades, despite the advances in UC diagnosis and therapy, there are still challenges to improve the overall survival (OS) of UC patients. PD-L1 inhibitor and PD-1 inhibitor have been approved for treating invasive UC, however, only about 20% of patients with metastatic UC show clinical benefits from immune checkpoint inhibitors. Therefor, bioinformatics tools were utilized to screen prognostic-related biomarkers, and analyze their relationship with immunocyte in UC, hoping to provide new ideas for the clinical treatment of UC patients.

Methods Three gene expression profiles (i.e. GSE32548, GSE32894 and GSE48075) were selected from GEO, and divide them into invasive and superficial UC group for study. NetworkAnalyst tool was used to construct gene regulatory network of DEGs, while DAVID and Metascape were utilized to perform gene ontology (GO) and kyoto encyclopedia of genes and genomes (KEGG) enrichment analysis of DEGs. The hub genes were screened by STRING and cytoscape, and the ONCOMINE, GEPIA, UALCAN, cBioPortal and HPA databases were used to analyze the expression differences and survival curves of UC at the DNA, RNA, protein levels and protein levels. TIMER was used to analyze the relationship between hub genes and immunocyte infiltration.

Results In total, 63 DEGs were identified from the GEO database of UC, of which 31 and 32 were up-and down-regulated. GO/KEGG pathway analysis identified DEGs were mainly enriched in the collagen catabolic process, extracellular matrix (ECM) organization, ECM structural constituent and ECM-receptor interaction. Nine hub genes (i.e. COL1A1, COL1A2, COL3A1, COL5A2, MMP9, POSTN, SPP1, VCAN and THBS2) upregulated in invasive UC compared with superficial UC were identified. cBioportal database analysis showed that 35% of UC patients presented genetic variants in the hub genes, of which amplification and deletion mutations were the most common. ONCOMINE and UALCAN database analysis showed that the mRNA expression of all hub genes in invasive UC was significantly higher than that in superficial UC and normal tissues. HPA database analysis showed that there was up-regulation of COL3A1, SPP1, POSTN and VCAN protein in UC tissues than in normal tissues. GEPIA showed that COL1A2, COL3A1, THBS2, and VCAN were positively correlated with the OS rate among patients with UC ($P < 0.05$). UALCAN showed that UC patients with high expression of COL1A1, COL1A2, COL5A2 and POSTN had a poorer prognosis ($P < 0.05$). TRRUST database analysis indicated that there was a significant correlation between the expression of the hub genes and the infiltration of CD4 + T cells, CD8 + T cells, macrophages, neutrophils and dendritic cells.

Conclusion Hub genes played important roles in pathogenesis and treatment prognosis of UC. Hub genes analysis provides new predictive biomolecules for UC immunotherapy and prognosis judgment.

Background

Urothelial carcinoma (UC), the most common histological type of urinary system [1], representing 2.1% of all cancer deaths [2]. There are an estimated 429,000 cases of UC diagnosed and 150,000 deaths worldwide each year [3]. UC derives from the transitional cells of the urinary tract and it is classified into the upper and lower tract UC [1, 4]. According to the anatomical location, among which up to 90% is located in bladder and merely 5% is located in upper urinary tract [4]. According to 2016 WHO classification of tumors of urinary system and male genital organs, UC can be further classified into invasive and superficial types depending on depth of tumor infiltration [5]. To date, despite the advances in UC diagnosis and therapy in the past decades, there are still challenges to improve the overall survival (OS) of UC patients [6]. According to a previous study, 30-70% of UC patients would present recurrence and 30% of cases with superficial UC would show invasive lesions [7]. Therefore, it is necessary to futher analysis UC and develop novel prognostic biomarkers for UC patients.

Nowadays, more studies focus on the roles of immune checkpoint inhibitors (ICIs) and immunocytes in pathogenesis, immunotherapy of cancer, as well as disease progression [7]. Interaction between tumor cells and immunocytes suppressed immune response and promoted tumor progression [8-10]. In 1976, Bacillus Calmette-Guerin was first used as an immune agent for treating UC [11] as it showed anti-tumor effects by activating the immune system and inducing inflammatory reaction [12]. However, no further meaningful progress has been achieved in treating invasive UC, local advanced or metastatic UC [13]. Up to now, several ICIs (e.g. PD-L1 inhibitor and PD-1 inhibitor) have been approved for treating invasive UC, which significantly prolong the survival of patients with invasive UC [14, 15]. However, only about 20% of patients with metastatic UC show clinical benefits from ICIs [1]. Moreover, immunotherapies are expensive. Therefore, it is necessary to select suitable patients for immunotherapy.

In this study, bioinformatics tools were utilized to conduct detailed molecular analysis of UC, we selected three gene expression profiles (i.e. GSE32548, GSE32894 and GSE48075) from Gene Expression Omnibus (GEO). NetworkAnalyst tool was used to construct gene regulatory network of DEGs, DAVID and Metascape were used to perform gene ontology (GO) and kyoto encyclopedia of genes and genomes (KEGG) enrichment analysis of DEGs. Hub genes were screened by Search Tool for Retrieval of Interacting Genes (STRING) and cytoscape software, and the ONCOMINE, Gene Expression profiling Interactive Analysis (GEPIA), UALCAN, cBioPortal and Human Protein Atlas (HPA) databases were used to analyze the expression differences and survival curves of UC at the DNA, RNA, and protein levels. TIMER was used to analyze the relationship between hub genes and immunocyte infiltration. We hope to further explore its sensitive and specific treatment methods for UC patients with immunotherapy.

Materials And Methods

Data source

GEO (www.ncbi.nlm.nih.gov/geo) serving as an open functional genomics database of the high-throughput resources [16]. Three gene expression profiles (i.e. GSE32548, GSE32894 and GSE48075) with adequate samples of invasive and superficial UC were selected .[17-19] Ta and T1 tumors were classified

as superficial UC, while tumors of T2 or higher stage were classified as infiltrating [20]. Samples with unknown stages (e.g. Tx) were excluded from analysis. All of the samples were based on Agilent GPL6947 platform (Illumina HumanHT-12 V3.0), and all data were freely available online. This study did not involve any experiments on humans or animals.

Data processing of DEGs

GEO2R online analysis tool (https://www.ncbi.nlm.nih.gov/geo/geo2r/) was used to detect the DEGs between superficial and infiltrating UC samples. NetworkAnalyst (https://www.networkanalyst.ca/) is an analytical platform integrating tissue-or cell-type-specific protein-protein interaction (PPI) network, gene co-expression network, and gene regulatory network [21]. In this study, we constructed gene regulatory network of DEGs using NetworkAnalyst tool. Genes met the cut-off criteria, with adjusted $P < 0.05$ and $|\log FC| \geq 1.0$, were considered as DEGs. Statistical analysis was carried out for each dataset in NetworkAnalyst, and Venn plot (http://bioinformatics.psb.ugent.be/webtools/Venn/) were used to intersect three datasets to obtain common DEGs.

GO/KEGG pathway analysis of DEGs

GO is an ontology widely used in bioinformatics and a common useful method for large-scale functional enrichment research, covering biological process (BP), molecular function (MF), and cellular component (CC) [22]. KEGG contains abundant data on genomes, biological pathways, diseases, chemical substances, and drugs [23]. Initially, DAVID (https://david.ncifcrf.gov) website is designed to investigate the function of a large number of genes or proteins, containing a comprehensive set of functional annotation tools to clarify biological functions of target genes [24], and we utilized it for the GO annotation and KEGG pathway enrichment analysis of DEGs. To ensure the credibility of results, we also analyzed the data with online tools from Metascape (http://metascape.org/gp/index.html#/main/step1) software [25]. Cut-off value for significant GO terms and KEGG pathways were a false discovers rate (FDR) of less than 0.05. Meanwhile, GO and KEGG pathway enrichment analyses of DEGs were conducted using DAVID and Metascape, in order to visualize and integrate DEGs enrichment of BP, MF, CC and KEGG pathways.

PPI network construction and hub genes identification

STRING (http://string-db.org/; version 11.0) database is designed to analyze the PPI [26]. To evaluate the potential PPI relationship, DEGs were mapped using STRING database to evaluate PPI with a combined score of > 0.4 setting as the cut-off criterion. Subsequently, the PPI network was visualized by Cytoscape (www.cytoscape.org/; version 3.8.1) software serving as a bioinformatics software for computational analysis of cellular networks and merging of experimental omics datasets [27]. Nodes with a higher degree of connectivity tend to be more essential in maintaining the stability of the entire network. Hub genes were selected using CytoHubba network analyzer plug-in. Furthermore, Molecular Complex Detection (MCODE) plug-in was used to screen modules of hub genes from PPI network.

Hub genes analysis

ONCOMINE (www.oncomine.org) database is an integrated online cancer microarray database for DNA or RNA sequencing analysis [28]. In our study, data were extracted to evaluate the expression of hub genes in invasive and superficial UC tissues samples. Cut-off of *P*-value and fold change were as following: *P*-value: 0.01, fold change: 1.5, gene rank: 10%, data type: mRNA, Analysis type: cancer vs cancer analysis.

cBioPortal (www.cbioportal.org) is an online open-access website resource for exploring, visualizing, and analyzing multi dimensional cancer genomics data [29]. Copy number variation (CNV), mutations, and the summary of the gene types in UC were evaluated according to the online instructions of cBioPortal. In addition, relationship between gene mutation and UC prognosis was analyzed using cBioPortal tool based on TCGA database. In this study, we analyzed CNV of 413 UC samples (TCGA, provisional) and *P*-value of less than 0.05 was considered to be significantly different.

UALCAN (http://ualcan.path.uab.edu) is an interactive web resource based on the level 3 RNA-seq and clinical data of 31 cancer types from TCGA database [30]. In this study, data for hub genes expression was obtained using “Expression Analysis” module of UALCAN dataset. Afterwards, it was utilized to analyze mRNA expression of hub genes in individual UC stages and their association with individual stages. Student’s *t* test was conducted to analyze the differences in transcriptional expression. $P < 0.05$ was considered to be statically significant.

HPA (https://www.proteinatlas.org) is a free online database containing immunohistochemistry-based expression data for near 20 highly common kinds of cancers (n = 12 for each cancer type) [31]. It provides abundant transcriptome and proteome data on human normal tissues or pathological tissues through RNA sequence analysis and immunohistochemical analysis. In the present study, we compared the protein expression and distribution of hub genes in UC tissues with that of normal tissues in HPA. According to the fraction of stained cells, staining quantity was also divided into four levels: none, < 25%, 25-75%, and > 75%. Protein expression levels were based on staining intensity and staining quantity. Classification criteria for protein expression levels were as follows: negative, not detected; weak and < 25%, not detected; weak combined with either 25-75% or 75%, low; moderate and < 25%, low; moderate combined with either 25-75% or 75%, medium; strong and < 25%, medium; and strong combined with either 25-75% or 75%, high.

GEPIA (http://gepia.cancer-pku.cn) is an interactive web server for estimating mRNA expression data based on 9,736 tumors and 8,587 normal samples in TCGA and Genotype-tissue Expression dataset projects [32]. In this study, we examined the expression of hub genes in 402 UC patients from TCGA database based on gene expression by using the Logrank and Mantel-Cox tests. A total of 402 patients were used to evaluate the OS. Hazard ratio (HR), 95% confident interval (CI), and *P*-value were calculated accordingly. Similarly, in order to ensure its accuracy, we also used UALCAN platform to analyze hub genes survival. Survival analysis of hub genes in UC was performed on the GEPIA and UALCAN online database, respectively.

TIMER (https://cistrome.shinyapps.io/timer/) is a reliable, intuitive tool that provides systematic evaluations of the infiltration of different immunocyte and their clinical impact [33]. In this study, Gene module was used to evaluate the correlation between the hub genes level and the infiltration of immunocyte. Survival module was used to evaluate the correlation among clinical outcome and the infiltration of immunocyte and the hub genes expression. SCNA module was used to explore the relationship between copy number variation and immune infiltration.

Results

Identification of DEGs

Three gene expression profiles (i.e. GSE32548, GSE32894 and GSE48075) were selected in this study. GSE32548 gene expression profile contained 130 samples including 38 invasive UC samples and 92 superficial UC samples, respectively. GSE32894 gene expression profile had 306 samples including 208 invasive UC samples and 98 superficial UC samples, respectively. GSE48075 gene expression profile had 142 samples, among which 73 were invasive UC samples and 69 were superficial UC samples (Table 1). Based on the criteria of $P < 0.05$ and $|\log FC| \geq 1$ through NetworkAnalyst platform, we obtained 157 DEGs from GSE32548, including 99 up-regulated genes and 58 down-regulated genes, 97 DEGs from GSE32894, including 47 up-regulated genes and 50 down-regulated genes, and 227 DEGs from GSE48075, including 130 up-regulated genes and 97 down-regulated genes (Supplementary Table 1). In addition, 63 DEGs were differentially expressed among all three groups, including 31 up-regulated genes and 32 down-regulated genes. All DEGs were identified by comparing superficial UC with infiltrating UC samples. Subsequently, DEGs of three datasets and common DEGs were visualized with Venn plot (Fig. 1A-C), and volcano maps (Fig. 1D-1G) were utilized to analyze the intersection of DEG profiles.

GO/KEGG Functional enrichment analyses of DEGs

The biological functions of all identified DEGs were evaluated by GO and KEGG pathway enrichment analyses. In the enrichment analysis of BPs, DEGs were significantly enriched in collagen catabolic process, collagen fibril organization, extracellular matrix (ECM) organization, skeletal system development, transforming growth factor beta receptor signaling pathway and cellular response to amino acid stimulus (Fig. 2A). In CC analysis, DEGs were significantly enriched in ECM, collagen trimer, extracellular space, proteinaceous ECM, extracellular region and extracellular exosome (Fig. 2B). In MF analysis, DEGs were predominantly enriched in ECM structural constituent, platelet-derived growth factor binding, heparin binding oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, monooxygenase activity and calcium ion binding ($P < 0.05$, Fig. 2C and Table 2). KEGG pathway analysis indicated that DEGs were significantly enriched in ECM-receptor interaction, focal adhesion, protein digestion and absorption, amoebiasis, platelet activation and PI3K-Akt signaling pathway with the highest gene numbers ($P < 0.05$, Fig. 2D and Table 3). We further analyzed DEGs

enrichment of GO and KEGG pathways through Metascape (Supplementary Table 2)and results were roughly consistent with DAVID website.

PPI network construction and hub genes identification

Protein interaction among DEGs were predicted with STRING tools. A total of 62 nodes and 109 edges involved in PPI network (Fig. 2E). The results were then transferred to Cytoscape software to analyze the interaction between candidate DEGs in invasive and superficial UC. To investigate the significant modules in this PPI network, the most significant modules were obtained by Cytotype MCODE, with an enrichment score of 10.36. These indicated that module one consisted of 12 nodes and 57 edges (Fig. 2F). The top 9 genes with degree of ≥ 10 were screened based on CytoHubba analysis in the PPI network (Table 4). The results showed that matrix metallopeptidase 9 (MMP9) was the most outstanding gene with a connectivity degree of 16, followed by periostin (POSTN; degree = 13), collagen type III alpha 1 chain (COL3A1; degree = 13), collagen type I alpha 1 chain (COL1A1; degree = 13), collagen type I alpha 2 chain (COL1A2; degree = 13), Versican (VCAN; degree = 12), thrombospondin 2 (THBS2; degree = 11), collagen type V alpha 2 chain (COL5A2; degree = 11), secreted phosphoprotein 1 (SPP1; degree = 10). All of these hub genes were up-regulated in invasive UC compared with superficial UC.

Differential expression analysis of hub genes at DNA, mRNA, protein levels and prognostic of UC

In order to explore the distinct prognostic and potential therapeutic value of hub genes in UC patients, DNA, mRNA and protein expression was analyzed by ONCOMINE database, cBioportal, UALCAN and HPA, respectively.

A comprehensive analysis was given to the molecular characteristics of differentially expressed hub genes by cBioportal database. Provisional datasets of TCGA were utilized to analyze genetic alterations of hub genes that were differentially expressed. As a result, 44 patients (35%) showed genetic alteration of hub genes in cBioportal for Cancer Genomics database. Amplification and deep deletion mutations were the most common genetic alteration (Fig. 3J). VACN, THBS2, POSTN and COL5A2 were the highest four genes with genetic alterations with a mutation rate of 16%, 13%, 12% and 10%, respectively.

In ONCOMINE database, mRNA expression of the hub genes in 20 types of cancers was compared to that of normal tissue (Fig. 4). We found the expression of all hub genes in UC were higher than that in normal tissues. Similarly, we further analyzed mRNA expressions of the hub genes through GEPIA database, the results showed that the expression level of mRNA in UC of MMP9 and SPP1 was higher than that of normal tissues ($P < 0.05$, Supplementary Fig. 1). Furthermore, we also indicated that mRNA expression of hub genes in invasive UC patients was significantly higher than that in superficial UC patients in ONCOMINE database, ($P < 0.05$, Fig. 3A-3I, Supplementary Table 3).

In this section, we tried to explore the protein expression pattern of the hub genes in UC by HPA. There were no data on the expression of COL5A2 in UC and normal humans in the HPA database. MMP9 protein was not detected in normal and UC tissues. COL1A1, COL1A2 and THBS2 showed moderate

expression in normal and UC tissues. Low COL3A1 and SPP1 expression was detected in normal tissues, while moderate and high expression was detected in UC tissues, respective. POSTN and VCAN were not detected in normal tissues, and UC tissues with moderate expression of POSTN and high expression of VCAN (Fig. 5). The hub genes were not significantly different from the normal and UC tissues, which may be related to the small sample size (n = 12), and the regulation of other factors in the process of RNA translation protein.

The effects of these hub genes on OS and disease-free survival (DFS) rates were analyzed by Kaplan-Meier curve using the online GEPIA platform. Our data showed that COL1A2, COL3A1, THBS2 and VCAN were positively correlated with OS rate in UC patients ($P < 0.05$, Fig. 6). Meanwhile, there was no correlation between hub genes and DFS rate in UC (Supplementary Fig 2). Interestingly, the UALCAN websites showed that UC patients with high expression of COL1A1, COL1A2, COL5A2 and POSTN showed a poor prognosis ($P < 0.05$, Supplementary Fig 3). It is worth noting that COL1A2 is meaningful in both databases.

Immunocyte infiltration of the hub genes in patients with UC

TIMER database was utilized to investigate the correlation between differentially expressed hub genes and immunocyte infiltration. Gene module analysis showed that a negative correlation between expression of all hub genes and infiltration of B cell ($P < 0.05$). A positive correlation was noticed between hub genes expression and the infiltration of CD4+T cells, CD8+T cell, macrophages, neutrophils and dendritic cells (all $P < 0.05$, Fig. 7). We also evaluated the correlation of differentially expressed hub genes and immunocytes infiltration. Cox proportional hazard model was used for the correction of the following confounding factors: B cells, CD4+T cells, macrophages, neutrophils, COL1A1, COL1A2, COL3A1, COL5A2, MMP9, POSTN, SPP1, THBS2, VCAN. Expression of macrophages ($P = 0.004$), dendritic cells ($P = 0.016$), MMP9 ($P = 0.043$) was significantly associated with the UC patients (Table 5). Additionally, we further evaluated the impact of immune infiltration on the clinical prognosis of patients with UC. Our findings showed that B cells, T cells, macrophages, neutrophils, and dendritic cells were factors related to the cumulative survival rate of UC over time (Fig. 8). And SCNA module analysis showed that several immune cell infiltration levels seemed to associate with hub genes CNV in UC (Fig. 9).

Discussion

The pathogenesis and progression of UC is complicated with various genetic abnormalities [15]. Therefore, it is essential to identify appropriate biomarkers for UC. We conducted a series of bioinformatics analysis based on three independent gene chip databases.

In the present study, 63 common DEGs were identified including 31 that were up-regulated and 32 that were down-regulated, respectively. We focused on the function of differentially expressed DEGs using GO and KEGG pathway enrichment analysis. Our data demonstrated that DEGs were mainly enriched in collagen catabolic process, ECM organization, ECM structural constituent and ECM-receptor interaction. ECM was a key component exerting an active effect in all the marks of cancer and was associated with

tumor metastasis [34]. Collagen was recognized as the most important component of the ECM, which provided multiple biochemical and biophysical cues to tumor cells [35]. ECM-receptor interaction was shown to be important components of cancer cell proliferation and invasion [36]. These results indicated that these DEGs were mainly involved in the occurrence, development and metastasis of UC. Among the DEGs, nine potential hub genes were obtained using the STRING and MCODE plug-in of Cytoscape. Notably, all our hub genes were overexpressed in UC. Afterwards, we analyzed the characteristics of hub genes through various databases with an aim to investigate the potential biomarkers of UC.

There were many studies reporting the family of collagen (e.g. COL1A1, COL1A2, COL3A1 and COL5A2) that was involved in carcinogenesis. Li et al. [37] indicated that the high COL1A1 expression was associated with shorter survival in UC patients. Brooks et al. [38] found that high COL1A1 expression was associated with shorter survival time in patients with superficial UC. COL1A1 has been considered as an oncogene and promoted cancer migration and invasion by inducing epithelial-mesenchymal transition (EMT) [39]. In addition, increasing evidence confirmed that COL1A1 involved in proliferation, migration, development, and progression of various cancers [37, 39]. This indicated that COL1A1 might be a putative therapeutic target for cancer. Rong et al. [40] reported high expression of COL1A2 in gastric cancer than in normal tissues. Additionally, Tamilzhalagan et al. [41] reported that miR-25 could selectively target the COL1A2 gene. MiR-25 silencing increased the expression of COL1A2 and inhibited the expression of E-cadherin. This is related to EMT, revealing the inhibitory effect of miR-25 on diffuse gastric cancer [15]. Notably, COL1A2 was highly expressed in pancreatic cancer, sarcoma, and gastric cancer tissues and promoted cell proliferation, migration, and invasion, which provided insights for mechanistic studies and clinical trials [42]. It has been reported that mRNA transcription level of bladder cancer is increased and CpG hypermethylation of COL1A2 contributes to proliferation and migration activity of human bladder cancer [43]. It has been reported that mRNA transcription level of bladder cancer is increased and CpG hypermethylation of COL1A2 contributes to proliferation and migration activity of human bladder cancer. Qiu et al. reported that miR-29a/b promoted cell migration and invasion in nasopharyngeal carcinoma progression by regulating COL3A1 gene expression [42]. While Su et al. indicated that let-7d can at least partially inhibit growth, metastasis, and tumor macrophage infiltration in renal cell carcinoma by targeting COL3A1 and CCL7 [42]. Yuan et al. discovered that COL3A1 played certain roles in the progression of human UC and influenced the prognosis probably by regulating MAPK signaling pathway, which contributed to the poor prognosis of UC [42]. These were consistent with our results. In a previous study, the in situ expression of COL5A2 increased in invasive breast cancer and was associated with ECM remodeling, suggesting that COL5A2 is associated with tumor progression in breast cancer [42]. Moreover, the expression of COL5A2 was related to the malignancy of colorectal cancer, suggesting its potential as a biomarker for colorectal cancer. Zeng et al. [44] demonstrated that COL5A2 was related to the poor clinical outcome and survival rate of patients with UC, demonstrating that COL5A2 was associated with poor clinical outcomes and survivals of patients with UC. This implied that it could be regarded as a biomarker of UC. Therefore, we considered that COL1A1, COL1A2, COL3A1 and COL5A2 may serve as promising candidates for the diagnosis, treatment and prognosis of UC.

MMP9 is an important matrix proteinase involving in embryonic development, wound healing, arthritis, angiogenesis, and cancer invasion and metastasis [45]. MMP9 mRNA expression was higher in invasive UC than that in superficial UC [46]. Besides, Lei et al. [8] revealed MMP9 could promote cancer invasion, metastasis and angiogenesis, while its expression was correlated with UC grade, invasiveness and poor prognosis. Wu et al. [47] found that the Wnt signaling pathway would regulate UC metastasis through activating MMP9, and showed that MMP9 may be regulated by miR-3713 in UC cells.

POSTN is an ECM protein expressed in many normal tissues, where it regulated cell adhesion and played an important role in the development and maintenance of mechanical stress-bearing structures (e.g. bones, teeth, and heart valves) [48]. POSTN was overexpressed in many human cancers, and high POSTN expression was correlated with tumor proliferation, cell motility and invasion, metastasis, angiogenesis, evasion of apoptosis, clinical stage, and survival of patients with UC [48], which is consistent with our findings. In addition, at the molecular level, POSTN involved in activation of PI3K/Akt and/or MAP kinase pathways via interacting with integrin receptors (i.e. aVb3 and aVb5), which subsequently promoted cell adhesion, motility, and angiogenesis. Furthermore, it could regulate E-cadherin expression and cell invasion in a cell-type dependent manner via Akt phosphorylation and regulating E-cadherin transcription, as well as negative regulation on Snail and Twist [49]. We identify POSTN expression as a feature of invasive UC associated with poor outcomes.

SPP1, also known as Osteopontin (OPN), is located on chromosome 4 (4q13) including 6 introns and 7 exons [50]. Recent studies have reported that SPP1 was significantly associated with cell growth, adherence and invasion in tumorigenesis and metastasis [50, 51]. It was over-expressed in lung, colon, breast, and prostate cancers. Our study showed that SPP1 expression was related to tumor stage, progression and survival rate. In addition, high SPP1 gene expression was related to poor prognosis. This implied that SPP1 may be a diagnostic and prognostic marker for UC. Meanwhile, SPP1 can enhance cancer cell survival, angiogenesis, and inflammation, and promote metastasis by enhancing EMT. This indicated that SPP1 was a tumor-promoting gene in tumors [51, 52].

THBS2, also known as TSP2, is a disulfide-linked glycoprotein that mediates ECM assembly, cell-to-matrix interactions, degradation of MMP2 and MMP9, and inhibition of angiogenesis [53]. THBS2 mainly participates in tumors by inhibiting angiogenesis and negatively regulating MMP-2 and MMP-9 [51]. THBS2 is generally considered a tumor suppressive gene. However, there are disputes on expression of THBS2 in various tumors. High THBS2 immunoreactivity was associated with the tumor response to neoadjuvant chemoradiotherapy. Besides, it was also an independent and good prognostic factor for DFS in patients with rectal cancer [53]. THBS2 gene is down-regulated in prostate cancer tissues and cell lines [53]. In addition, Chijiwa found that the transcription level of THBS2 in lung adenocarcinoma was significantly higher than that of normal lung tissue ($P < 0.0001$) [53]. In our study, THBS2 was up-regulated in UC, which played important roles in UC progression.

VCAN, also known as chondroitin sulfate proteoglycan 2 (CSPG2), is a highly conserved structural component in ECM expressed in invasive and metastatic cancer [54]. To date, four subtypes (i.e.V0, V1,

V2, and V3) have been identified, and all of which contribute to the proliferation, adhesion, and migration of tumor cells, as well as interaction with tumor microenvironment (TME) [55]. In a recent study, VCAN has been reported to be correlated with low RhoGDI2 expression, high VCAN expression and poor clinical outcomes [54]. Our data showed VCAN expression was related to grade and prognosis of UC patients. The highly inflammatory microenvironment induced high expression of VCAN, and increased macrophage infiltration. In turn, infiltration of macrophage exacerbated VCAN overexpression along with the secretion of other cytokines and inflammatory mediators. In this context, VCAN appeared to mediate a dialogue between inflammatory cells, cytokines and cancer cells in TME [54]. Therefore, we proposed that VCAN may serve as the first step in the amplification of inflammation.

Increasing evidence indicated that immunocyte infiltration may affect tumor progression and recurrence [9]. In recent years, immunotherapy has been preferred in treating aggressive or advanced cancers with promising efficiency [7, 10]. As immune microenvironment affects tumor progression, we intended to explore an immune-related gene signature associated with prognosis of patients with UC. TME is composed of cancer cells and non-cancer cell components such as immunocyte (B cells, CD8+T cells, CD4+T cells, macrophages, neutrophils and dendritic cells), mesenchymal cells, endothelial cells, ECM molecules, and inflammatory mediators [56]. Tumor gene expression profile could quantify the immune activity in TME, such as CD4+T recognizing cancer antigens, and then activated M1 macrophages may inhibit tumor growth [57]. In addition, fibroblasts were associated with exocrine phenotype of T cells in bladder cancer, and the secreting transforming growth factor-β (TGF-β) can lead to the efflux of immunocyte or the resistance of chemotherapy drugs [58]. Therefore, it is necessary to screen biomarkers for hub genes in the immune responses.

Although the effects of CD4+T cells on the clinical outcome were controversial, high infiltration of CD8+ cytotoxic T cells and memory T cells were clearly shown with a longer survival. There was increase in expression of T-helper 1 and cytotoxicity-related genes in UC patients [8]. Mariathasan et al. also shown that invasive UC patient with a high number of CD8+T cells within the tumor showed a longer survival than those with fewer CD8+T cells [8]. Immunocyte has complex biological relationships with tumor cells, which then triggers the progression or repression [15]. Therefore, we speculated that the diversity of immunocyte in TME might be one of the causes leading to different immunotherapeutic responses. In this study, we found that hub genes expressions were negatively correlation with B cell infiltration ($P < 0.05$), and was positive correlation wit the infiltration of CD4+T cells, CD8+T cell, macrophages, neutrophils and dendritic cells. In addition, our research results also showed that B cells, T cells, macrophages, neutrophils, and dendritic cells were factors related to the cumulative survival rate of UC over time and the immunocyte infiltration levels seemed to associate with hub genes CNV in UC. These findings together suggest that these hub genes were prognostic indicators and may involve in activation and recruitment of immunocytes in UC.

Our study focused on the exploration of therapeutic targets and prognostic biomarkers in hub genes of UC microenvironment by the screening of multiple databases. There are some limitations in our study. First, multiple different databases are utilized, and it is difficult to guarantee the unity among different

databases. Second, we did not verify the data from these databases. In dependent cohort and in vitro or in vivo studies are required to validate our results. Meanwhile, functional and molecular mechanisms investigations on the hub genes in UC should be further performed.

Conclusion

In conclusion, 63 differentially expressed genes were identified in the GEO datasets of UC. Nine of them (i.e. COL1A1, COL1A2, COL3A1, COL5A2, MMP9, POSTN, SPP1, THBS2 and VCAN) were identified to be important for the pathogenesis and progression of UC. In the database, the expression levels of hub genes in invasive UC tissues were significantly higher than in superficial UC tissues. The expression of hub genes cannot only be used as a prognostic indicator of UC, but it is also related to the immunocyte infiltration. The present study provided novel insights into the occurrence and progression of UC. However, the diagnostic and prognostic value of these genes required further validation. Therefore, we hope our results can provide novel insights for the design of new immunotherapeutic drugs, help clinicians choose appropriate drugs for their UC patients and prognostic biomarkers, and to more accurately predict the survival of patients with UC.

Abbreviations

UC: Urothelial carcinoma; OS: Overall survival; ICIs: Immune checkpoint inhibitors; GEO: Gene Expression Omnibus. GO: Gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; STRING: Search Tool for Retrieval of Interacting Genes; GEPIA: Gene Expression profiling Interactive Analysis; HPA: Human Protein Atlas; PPI: Protein-protein interaction; MCODE: Molecular complex detection; CNV: Copy number variation; HR: Hazard ratio; CI: Confident interval; ECM; Extracellular matrix; COL1A1: Collagen type I alpha 1 chain; COL1A2: Collagen type I alpha 2 chain; COL3A1: Collagen type III alpha 1 chain; COL5A2: Collagen type V alpha 2 chain; MMP9: Metallopeptidase 9; POSTN: Periostin; SPP1: Secreted phosphoprotein 1; THBS2: Thrombospondin 2; VCAN: Versican; DFS: Disease-free survival; EMT: Epithelial-mesenchymal transition; OPN: Osteopontin; CSPG2: chondroitin sulfate proteoglycan 2; TME: tumor microenvironment; TGF-β: transforming growth factor-β.

Declarations

Acknowledgements

The authors have no acknowledgement to disclose.

Funding

Natural Science Foundation of Shandong Province Project (ZR2020MC068)

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CW wrote and corrected the manuscript. CW and GHF built up all the molecular tools for the research. ZB and ZQQ supervised the whole research. CJ agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of the work were appropriately investigated and resolved. JH critically revised the manuscript to the approval of the final version agreed to be accountable. All authors read and approved the final manuscript.

Author details

1. Department of Pathology, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, 758 Hefei Road, Qingdao, Shandong, 266035, China
2. Department of Pathology, Binzhou Medical University Hospital, Binzhou, Shandong, 256603, China
3. Department of Urology, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, 758 Hefei Road, Qingdao, Shandong, 266035, China

References

[1] Andreev-Drakhlin AY, Egoryan G, Shah AY, Msaouel P, Alhalabi O, Gao J. The evolving treatment landscape of advanced urothelial carcinoma. Curr Opin Oncol. 2021;33:221-30.

[2] van der Weyden L, O'Donnell M, Plog S. Histological Characterization of Feline Bladder Urothelial Carcinoma. J Comp Pathol. 2021;182:9-14.

[3] Viswambaram P, Hayne D. Gender discrepancies in bladder cancer: potential explanations. Expert Rev Anticancer Ther. 2020;20:841-9.

[4] McGregor BA, Sonpavde G. Enfortumab Vedotin, a fully human monoclonal antibody against Nectin 4 conjugated to monomethyl auristatin E for metastatic urothelial Carcinoma. Expert Opin Investig Drugs. 2019;28:821-6.

[5] Reuter V, Algaba F, Amin M.B, Cao D,Cheng L, Compérat E, et al. WHO Classifification of Tumours of Urinary System and Male Genital Organs. 4th ed. IARC. 2017; 99-107.

[6] Luna-Velez MV, Dijkstra JJ, Heuschkel MA, Smit FP, van de Zande G, Smeets D, et al. Androgen receptor signalling confers clonogenic and migratory advantages in urothelial cell carcinoma of the bladder. Mol Oncol. 2021.

[7] 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.

[8] Na L, Bai Y, Sun Y, Wang Z, Wang W, Yuan L, et al. Identification of 9-Core Immune-Related Genes in Bladder Urothelial Carcinoma Prognosis. Front Oncol. 2020;10:1142.

[9] Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity. 2013;39:782-95.

[10] Liu X, Wu S, Yang Y, Zhao M, Zhu G, Hou Z. The prognostic landscape of tumor-infiltrating immune cell and immunomodulators in lung cancer. Biomed Pharmacother. 2017;95:55-61.

[11] Menon S, Shin S, Dy G. Advances in Cancer Immunotherapy in Solid Tumors. Cancers (Basel). 2016;8.

[12] Zhang C, Shen L, Qi F, Wang J, Luo J. Multi-omics analysis of tumor mutation burden combined with immune infiltrates in bladder urothelial carcinoma. J Cell Physiol. 2020;235:3849-63.

[13] Li X, Shang D, Shen H, Song J, Hao G, Tian Y. ZSCAN16 promotes proliferation, migration and invasion of bladder cancer via regulating NF-kB, AKT, mTOR, P38 and other genes. Biomed Pharmacother. 2020;126:110066.

[14] Lopez-Beltran A, Lopez-Rios F, Montironi R, Wildsmith S, Eckstein M. Immune Checkpoint Inhibitors in Urothelial Carcinoma: Recommendations for Practical Approaches to PD-L1 and Other Potential Predictive Biomarker Testing. Cancers (Basel). 2021;13.

[15] Tzanakakis G, Kavasi RM, Voudouri K, Berdiaki A, Spyridaki I, Tsatsakis A, et al. Role of the extracellular matrix in cancer-associated epithelial to mesenchymal transition phenomenon. Dev Dyn. 2018;247:368-81.

[16] Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30:207-10.

[17] Lindgren D, Sjodahl G, Lauss M, Staaf J, Chebil G, Lovgren K, et al. Integrated genomic and gene expression profiling identifies two major genomic circuits in urothelial carcinoma. PLoS One. 2012;7:e38863.

[18] Sjodahl G, Lauss M, Lovgren K, Chebil G, Gudjonsson S, Veerla S, et al. A molecular taxonomy for urothelial carcinoma. Clin Cancer Res. 2012;18:3377-86.

[19] Guo CC, Bondaruk J, Yao H, Wang Z, Zhang L, Lee S, et al. Assessment of Luminal and Basal Phenotypes in Bladder Cancer. Sci Rep. 2020;10:9743.

[20] Shen C, Liu J, Wang L, Liang Z, Niu H, Wang Y. Identification of metabolism-associated genes and construction of a prognostic signature in bladder cancer. Cancer Cell Int. 2020;20:538.

[21] Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res. 2019;47:W234-W41.

[22] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25-9.

[23] Kanehisa M. The KEGG database. Novartis Found Symp. 2002;247:91-101; discussion -3, 19-28, 244-52.

[24] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44-57.

[25] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10:1523.

[26] 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:D607-D13.

[27] Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. J Proteome Res. 2019;18:623-32.

[28] Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia. 2004;6:1-6.

[29] Unberath P, Knell C, Prokosch HU, Christoph J. Developing New Analysis Functions for a Translational Research Platform: Extending the cBioPortal for Cancer Genomics. Stud Health Technol Inform. 2019;258:46-50.

[30] Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. Neoplasia. 2017;19:649-58.

[31] Asplund A, Edqvist PH, Schwenk JM, Ponten F. Antibodies for profiling the human proteome-The Human Protein Atlas as a resource for cancer research. Proteomics. 2012;12:2067-77.

[32] 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:W98-W102.

[33] Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer Res. 2017;77:e108-e10.

[34] Li M, Wang X, Liu J, Mao X, Li D, Wang Z, et al. Identification of Core Prognosis-Related Candidate Genes in Chinese Gastric Cancer Population Based on Integrated Bioinformatics. Biomed Res Int. 2020;2020:8859826.

[35] Yang B, Zhao Y, Wang L, Zhao Y, Wei L, Chen D, et al. Identification of PLOD Family Genes as Novel Prognostic Biomarkers for Hepatocellular Carcinoma. Front Oncol. 2020;10:1695.

[36] Zeng Q, Sun S, Li Y, Li X, Li Z, Liang H. Identification of Therapeutic Targets and Prognostic Biomarkers Among CXC Chemokines in the Renal Cell Carcinoma Microenvironment. Front Oncol. 2019;9:1555.

[37] Li F, Teng H, Liu M, Liu B, Zhang D, Xu Z, et al. Prognostic Value of Immune-Related Genes in the Tumor Microenvironment of Bladder Cancer. Front Oncol. 2020;10:1302.

[38] Brooks M, Mo Q, Krasnow R, Ho PL, Lee YC, Xiao J, et al. Positive association of collagen type I with non-muscle invasive bladder cancer progression. Oncotarget. 2016;7:82609-19.

[39] Zhu H, Chen H, Wang J, Zhou L, Liu S. Collagen stiffness promoted non-muscle-invasive bladder cancer progression to muscle-invasive bladder cancer. Onco Targets Ther. 2019;12:3441-57.

[40] Rong L, Huang W, Tian S, Chi X, Zhao P, Liu F. COL1A2 is a Novel Biomarker to Improve Clinical Prediction in Human Gastric Cancer: Integrating Bioinformatics and Meta-Analysis. Pathol Oncol Res. 2018;24:129-34.

[41] Tamilzhalagan S, Rathinam D, Ganesan K. Amplified 7q21-22 gene MCM7 and its intronic miR-25 suppress COL1A2 associated genes to sustain intestinal gastric cancer features. Mol Carcinog. 2017;56:1590-602.

[42] Guo C, Shao T, Wei D, Li C, Liu F, Li M, et al. Bioinformatic Identification of Potential Hub Genes in Muscle-Invasive Bladder Urothelial Carcinoma. Cell Transplant. 2020;29:963689720965178.

[43] Mori K, Enokida H, Kagara I, Kawakami K, Chiyomaru T, Tatarano S, et al. CpG hypermethylation of collagen type I alpha 2 contributes to proliferation and migration activity of human bladder cancer. Int J Oncol. 2009;34:1593-602.

[44] Zeng XT, Liu XP, Liu TZ, Wang XH. The clinical significance of COL5A2 in patients with bladder cancer: A retrospective analysis of bladder cancer gene expression data. Medicine (Baltimore). 2018;97:e0091.

[45] Gilkes DM, Semenza GL, Wirtz D. Hypoxia and the extracellular matrix: drivers of tumour metastasis. Nat Rev Cancer. 2014;14:430-9.

[46] Zhang HJ, Tao J, Sheng L, Hu X, Rong RM, Xu M, et al. Twist2 promotes kidney cancer cell proliferation and invasion by regulating ITGA6 and CD44 expression in the ECM-receptor interaction pathway. Onco Targets Ther. 2016;9:1801-12.

[47] Wu WB, Wang W, Du YH, Li H, Xia SJ, Liu HT. MicroRNA-3713 regulates bladder cell invasion via MMP9. Sci Rep. 2016;6:32374.

[48] Silvers CR, Liu YR, Wu CH, Miyamoto H, Messing EM, Lee YF. Identification of extracellular vesicle-borne periostin as a feature of muscle-invasive bladder cancer. Oncotarget. 2016;7:23335-45.

[49] Kim CJ, Sakamoto K, Tambe Y, Inoue H. Opposite regulation of epithelial-to-mesenchymal transition and cell invasiveness by periostin between prostate and bladder cancer cells. Int J Oncol. 2011;38:1759-66.

[50] Li S, Yang R, Sun X, Miao S, Lu T, Wang Y, et al. Identification of SPP1 as a promising biomarker to predict clinical outcome of lung adenocarcinoma individuals. Gene. 2018;679:398-404.

[51] Hubbard NE, Chen QJ, Sickafoose LK, Wood MB, Gregg JP, Abrahamsson NM, et al. Transgenic mammary epithelial osteopontin (spp1) expression induces proliferation and alveologenesis. Genes Cancer. 2013;4:201-12.

[52] Xu C, Sun L, Jiang C, Zhou H, Gu L, Liu Y, et al. SPP1, analyzed by bioinformatics methods, promotes the metastasis in colorectal cancer by activating EMT pathway. Biomed Pharmacother. 2017;91:1167-77.

[53] Chang IW, Li CF, Lin VC, He HL, Liang PI, Wu WJ, et al. Prognostic Impact of Thrombospodin-2 (THBS2) Overexpression on Patients with Urothelial Carcinomas of Upper Urinary Tracts and Bladders. J Cancer. 2016;7:1541-9.

[54] Said N, Theodorescu D. RhoGDI2 suppresses bladder cancer metastasis via reduction of inflammation in the tumor microenvironment. Oncoimmunology. 2012;1:1175-7.

[55] Luo Y, Zeng G, Wu S. Identification of Microenvironment-Related Prognostic Genes in Bladder Cancer Based on Gene Expression Profile. Front Genet. 2019;10:1187.

[56] Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. Cancer Cell. 2012;21:309-22.

[57] Lin P, Guo YN, Shi L, Li XJ, Yang H, He Y, et al. Development of a prognostic index based on an immunogenomic landscape analysis of papillary thyroid cancer. Aging (Albany NY). 2019;11:480-500.

[58] Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature. 2018;554:544-8.

Tables

Table 1: Statistics of the three microarray databases derived from the GEO database

Dataset ID	Superficial urothelial carcinoma	Invasive urothelial carcinoma	Total number
GSE32548	92	38	130
GSE32894	98	208	306
GSE48075	69	73	142

Table 2: GO analysis of differentially expressed genes in UC.

Category	Term	Count	*P*-Value
BP	GO:0030198~extracellular matrix organization	9	1.69E-07
BP	GO:0030574~collagen catabolic process	8	1.12E-09
BP	GO:0030199~collagen fibril organization	7	2.22E-09
BP	GO:0001501~skeletal system development	7	4.43E-06
BP	GO:0007179~transforming growth factor beta receptor signaling pathway	5	2.01E-04
BP	GO:0071230~cellular response to amino acid stimulus	4	4.36E-04
CC	GO:0070062~extracellular exosome	23	4.61E-05
CC	GO:0005615~extracellular space	20	3.09E-08
CC	GO:0005576~extracellular region	20	5.12E-07
CC	GO:0031012~extracellular matrix	13	1.88E-10
CC	GO:0005578~proteinaceous extracellular matrix	10	2.16E-07
CC	GO:0005581~collagen trimer	8	2.06E-08
MF	GO:0005509~calcium ion binding	9	1.88E-03
MF	GO:0005201~extracellular matrix structural constituent	6	2.30E-06
MF	GO:0008201~heparin binding	6	1.56E-04
MF	GO:0048407~platelet-derived growth factor binding	4	5.01E-06
MF	GO:0016705~oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	4	8.01E-04
MF	GO:0004497~monooxygenase activity	4	8.43E-04

Table 3: KEGG analysis of differentially expressed genes in UC.

Category	Term	Count	Genes in the test set	*P*-Value
KEGG	hsa04974:Protein digestion and absorption	6	COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, COL6A3	3.38E-05
KEGG	hsa05146:Amoebiasis	4	COL1A1, COL1A2, COL3A1, CXCL8	8.29E-05
KEGG	hsa04611:Platelet activation	4	COL1A1,COL1A2,COL3A1,FCER1G	2.18E-04
KEGG	hsa04512:ECM-receptor interaction	5	COL1A1, COL1A2, COL6A3, THBS2, SPP1	6.52E-08
KEGG	hsa04510:Focal adhesion	5	COL1A1, COL1A2, COL6A3, THBS2, SPP1	2.21E-05
KEGG	hsa04151:PI3K-Akt signaling pathway	5	COL1A1, COL1A2, COL6A3, THBS2, SPP1	5.61E-04

Table 4: Functional roles of hub genes with a degree≥10

Gene Symbol	Full name	Degree	Function
MMP9	matrix metallopeptidase 9	16	Proteins of the MMP family are involved in the breakdown of extracellular matrix in normal physiological processes and the enzyme encoded by this gene degrades type IV and V collagens.
POSTN	periostin	13	This gene encodes a secreted extracellular matrix protein that functions in tissue development and regeneration, which plays a role in cancer stem cell maintenance and metastasis.
COL3A1	collagen type III alpha 1 chain	13	This gene encodes the pro-alpha1 chains of type III collagen. Mutations are associated with Ehlers-Danlos syndrome types IV, as well as aortic and arterial aneurysms.
COL1A1	collagen type I alpha 1 chain	13	This gene encodes the pro-alpha1 chains of type I collagen and mutations are associated with osteogenesis imperfecta types I-IV, Ehlers-Danlos syndrome type VIIA, Ehlers-Danlos syndrome Classical type, Caffey Disease and idiopathic osteoporosis.
COL1A2	collagen type I alpha 2 chain	13	This gene encodes the pro-alpha2 chain of type I collagen and their mutations are associated with osteogenesis imperfecta types I-IV, Ehlers-Danlos syndrome type VIIB, recessive Ehlers-Danlos syndrome Classical type, idiopathic osteoporosis, and atypical Marfan syndrome.
VCAN	versican	12	This gene is a member of the aggrecan/versican proteoglycan family. The protein encoded is a major component of the extracellular matrix and is involved in cell adhesion, proliferation, proliferation, migration and angiogenesis and plays a central role in tissue morphogenesis and maintenance. Mutations of this gene are the cause of Wagner syndrome type 1.
THBS2	thrombospondin 2	11	The protein encoded by this gene belongs to the thrombospondin family. It mediates cell-to-cell and cell-to-matrix interactions.
COL5A2	collagen type V alpha 2 chain	11	This gene encodes an alpha chain for one of the low abundance fibrillar collagens and product is closely related to type XI collagen and mutations are associated with Ehlers-Danlos syndrome, types I and II.
SPP1	secreted phosphoprotein 1	10	The protein encoded by this gene is involved in the attachment of osteoclasts to the mineralized bone matrix. The encoded protein is secreted and binds hydroxyapatite with high affinity and upregulates expression of interferon-gamma and interleukin-12.

Table 5: The cox proportional hazard model of the hub genes and six tumor-infifiltrating immune cells in UC (TIMER).

	Coef	HR	95%CI_l	95%CI_u	*P*-value
B_cell	-2.202	0.111	0.005	2.316	0.156
CD8_Tcell	1.485	4.416	0.334	58.445	0.26
CD4_Tcell	-1.708	0.181	0.004	8.031	0.377
Macrophage	3.56	35.17	3.085	400.986	0.004
Neutrophil	-3	0.05	0	5.703	0.215
Dendritic	-0.629	0.533	0.107	2.667	0.444
COL1A1	0.095	1.1	0.659	1.836	0.716
COL1A2	0.008	1.008	0.627	1.622	0.972
COL3A1	-0.147	0.863	0.573	1.302	0.483
COL5A2	0.161	1.175	0.872	1.582	0.29
MMP9	0.11	1.117	1.003	1.243	0.043
POSTN	-0.021	0.979	0.829	1.157	0.807
SPP1	0.011	1.011	0.926	1.105	0.801
THBS2	-0.031	0.97	0.833	1.129	0.694
VCAN	-0.027	0.973	0.801	1.183	0.786

HR: Hazard ratio; CI: Confident interval

Figures

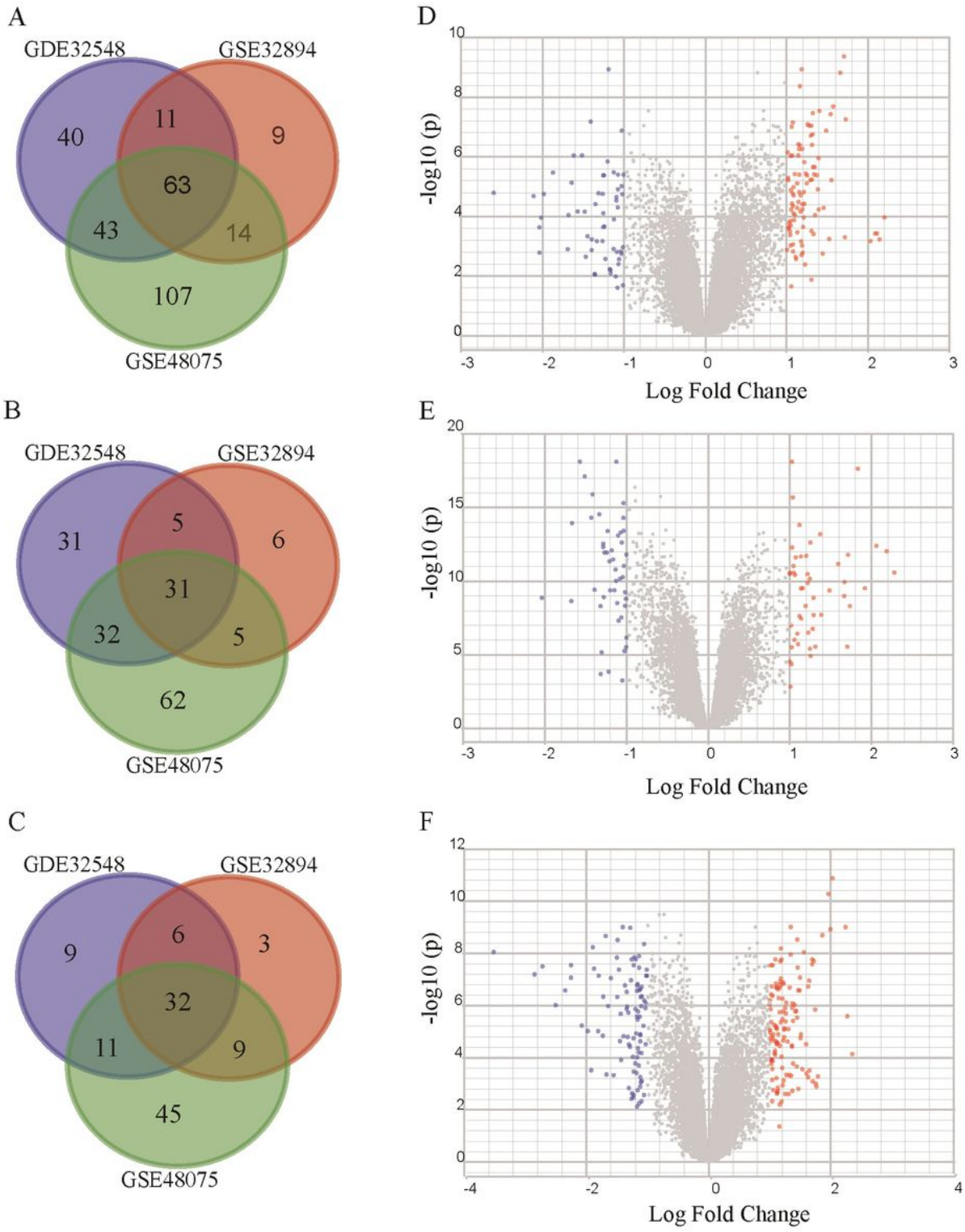

Figure 1

Venn diagram analyses showing the most significant module related to DEGs. Sixty-three DEGs were significantly differentially expressed among three groups (A), including 31 with significantly upregulation (B) and 32 with downregulation (C). DEGs were identified from GSE32548 (D), GSE32894 (E), and GSE48075 (F) gene expression profling datasets based on fold change ≥ 1 and adjusted P-value < 0.05. Red: upregulated genes, blue: downregulated genes, gray: non-significant genes.

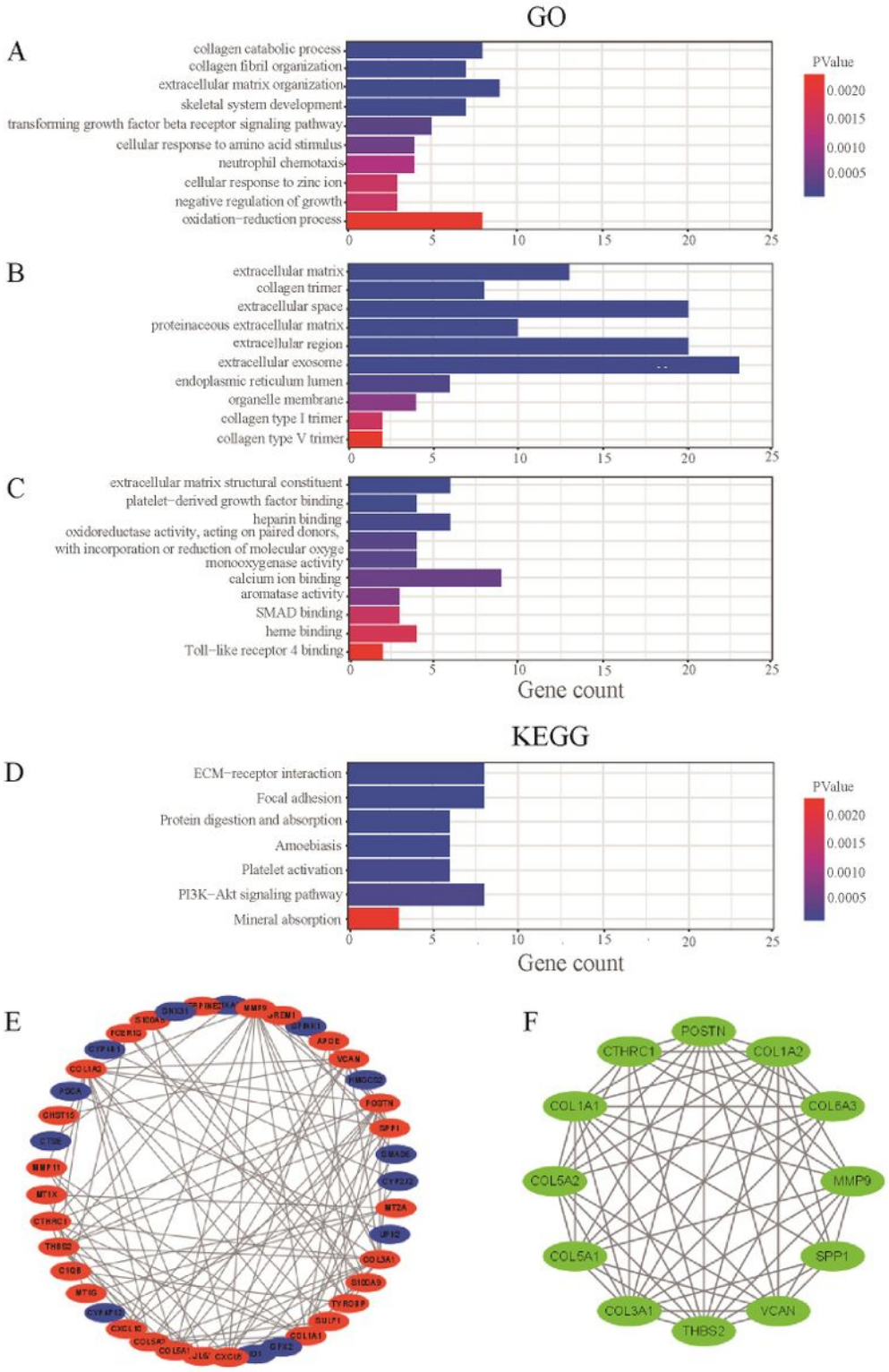

Figure 2

Enrichment analysis of genes related to DEGs. Represent the enrichment analysis results of genes involved in hub gene, namely BP (A), CC (B), MF (C) and KEGG (D). A PPI network of the DEGs was constructed using Cytoscape software (E). The most significant module of the PPI network includes 12 nodes and 57 edges (F). Red: upregulated genes, blue: downregulated genes, green: hub genes.

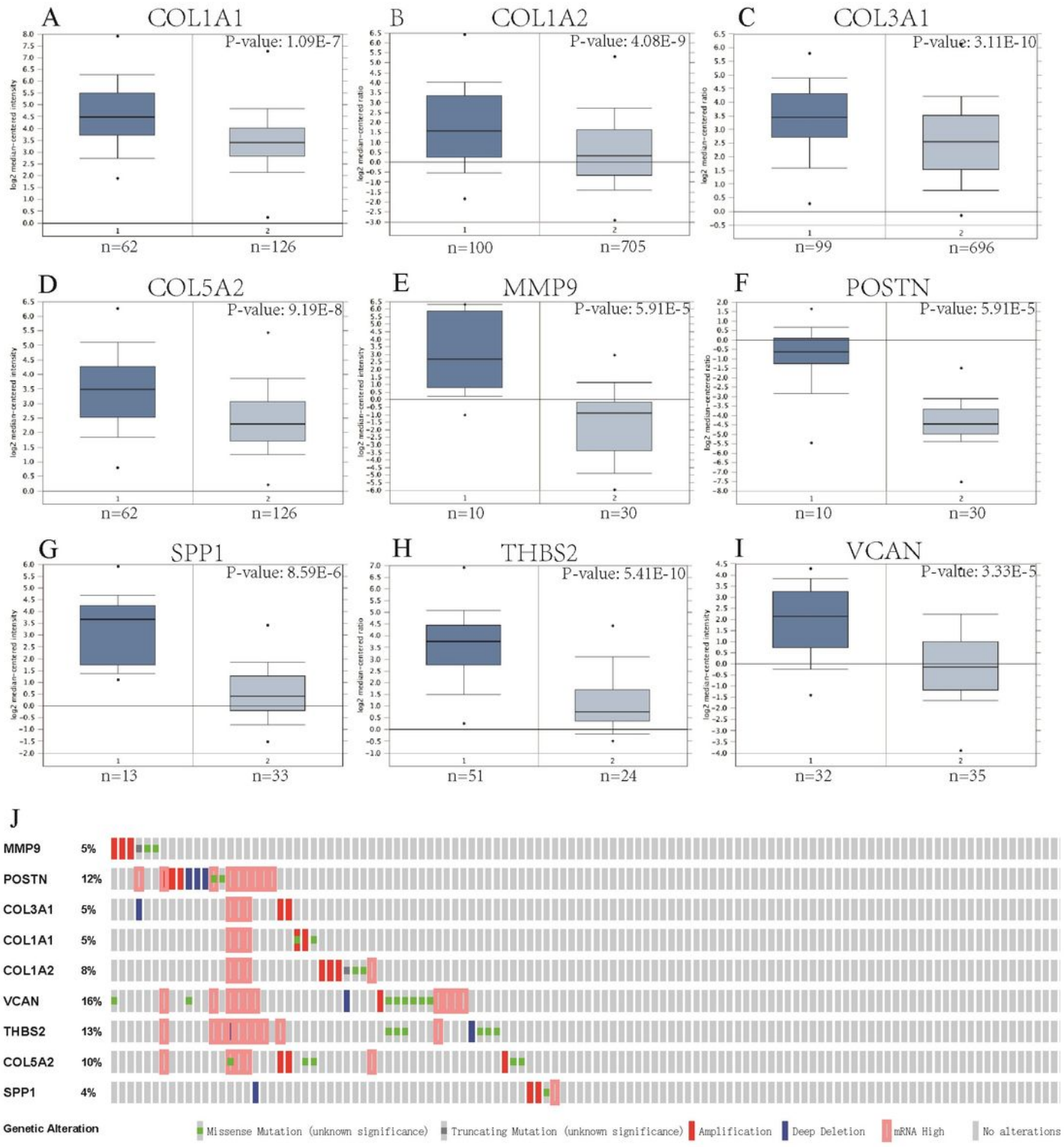

Figure 3

Expression of the hub genes was higher invasive UC, compared with superficial UC (A-I, Oncomine database). 1: invasive UC, 2: superficial UC. The genetic alteration profiles of the hub genes (J, cBioPort database).

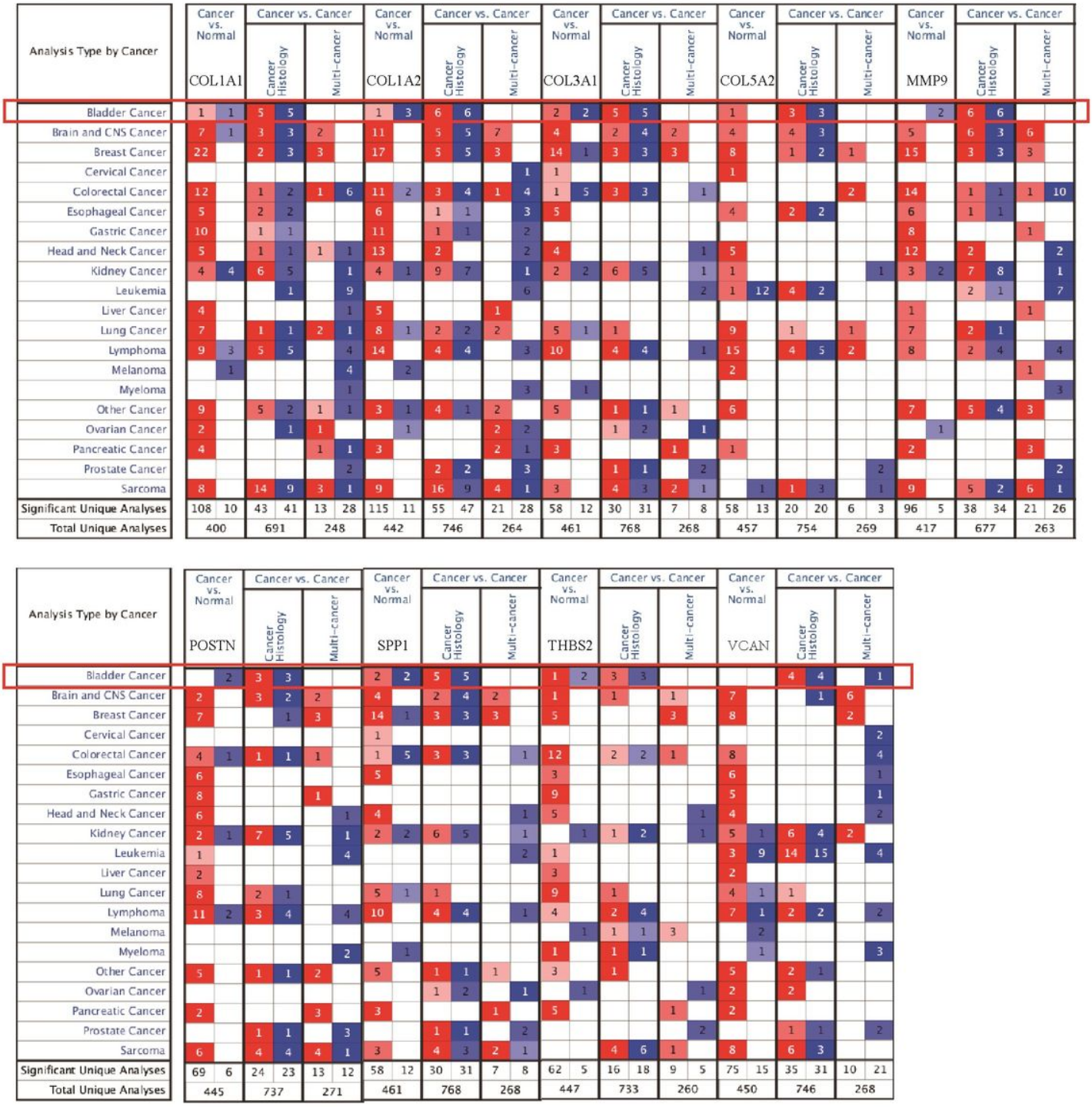

Figure 4

Transcriptional expression of the hub genes in 20 different types of cancer diseases (ONCOMINE database).

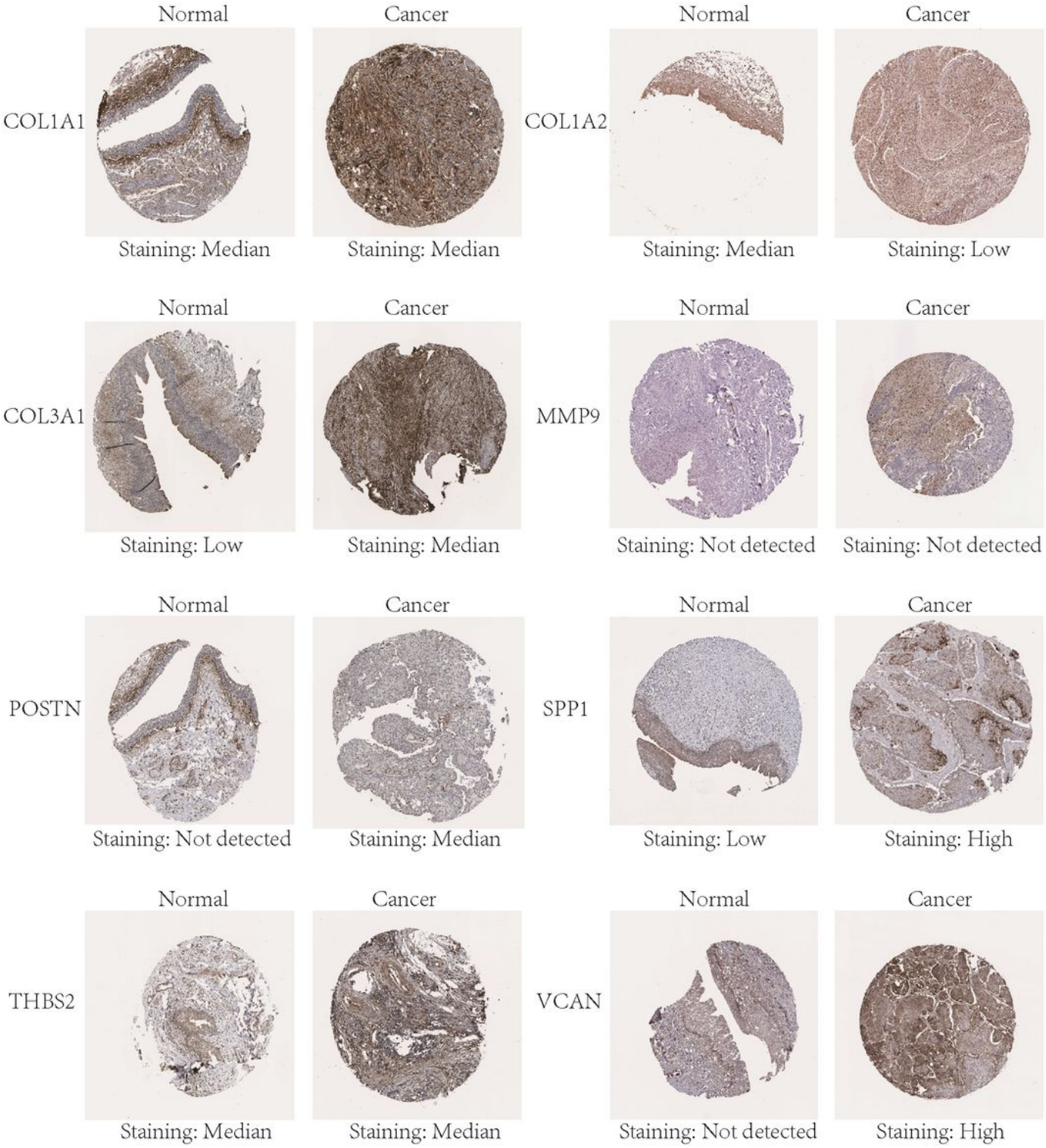

Figure 5

Representative immunohistochemistry images of distinct 8 hub genes in UC tissues and normal urothelial tissues (except for COL5A2, HPA database).

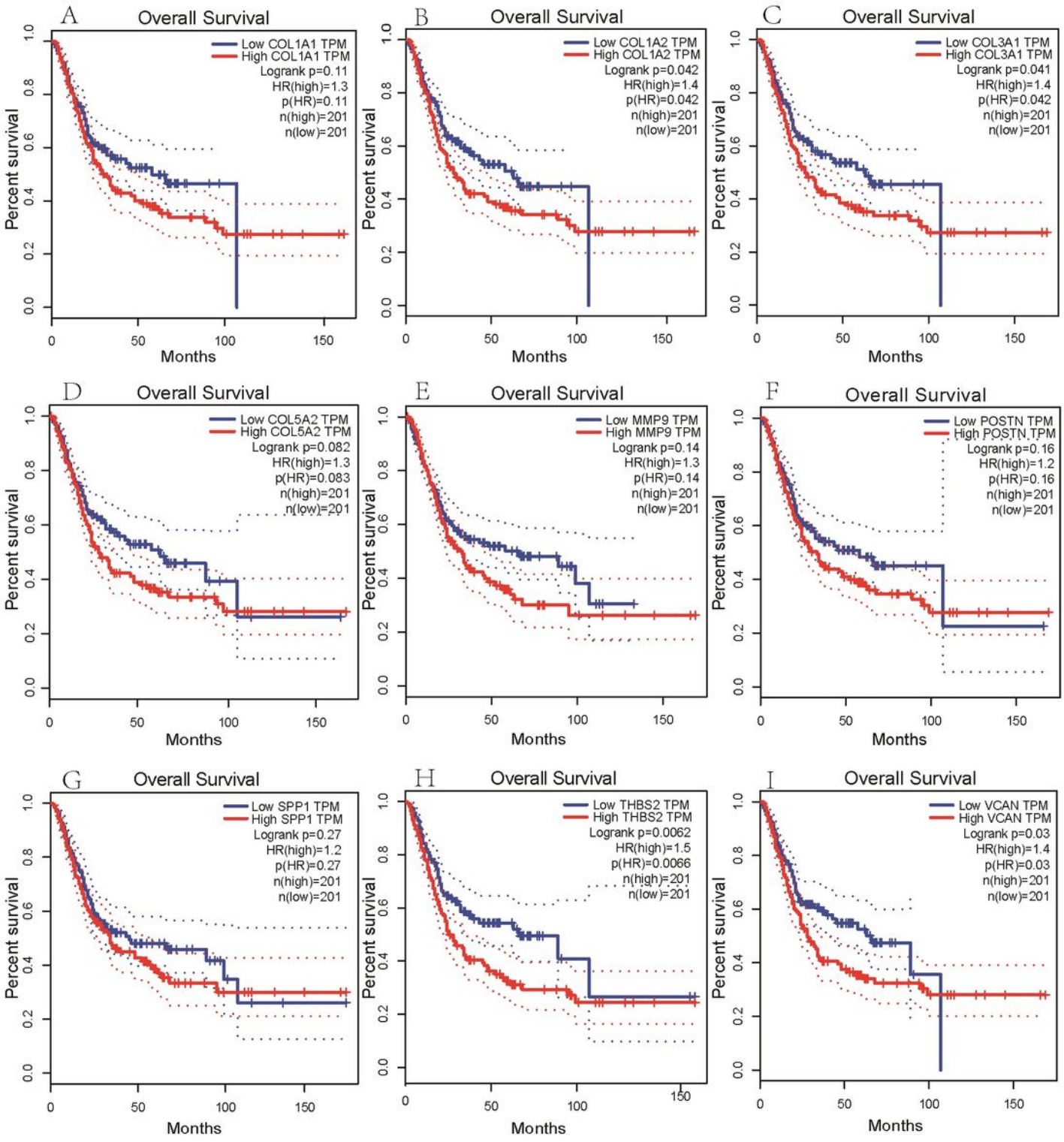

Figure 6

The prognostic value of the hub genes in UC patients in the OS curve (GEPIA database). The OS curve of COL1A1 (A), COL1A2 (B), COL3A1 (C), COL5A2 (D), MMP9 (E), POSTN (F), SPP1 (G), THBS2 (H) and VCAN (I).

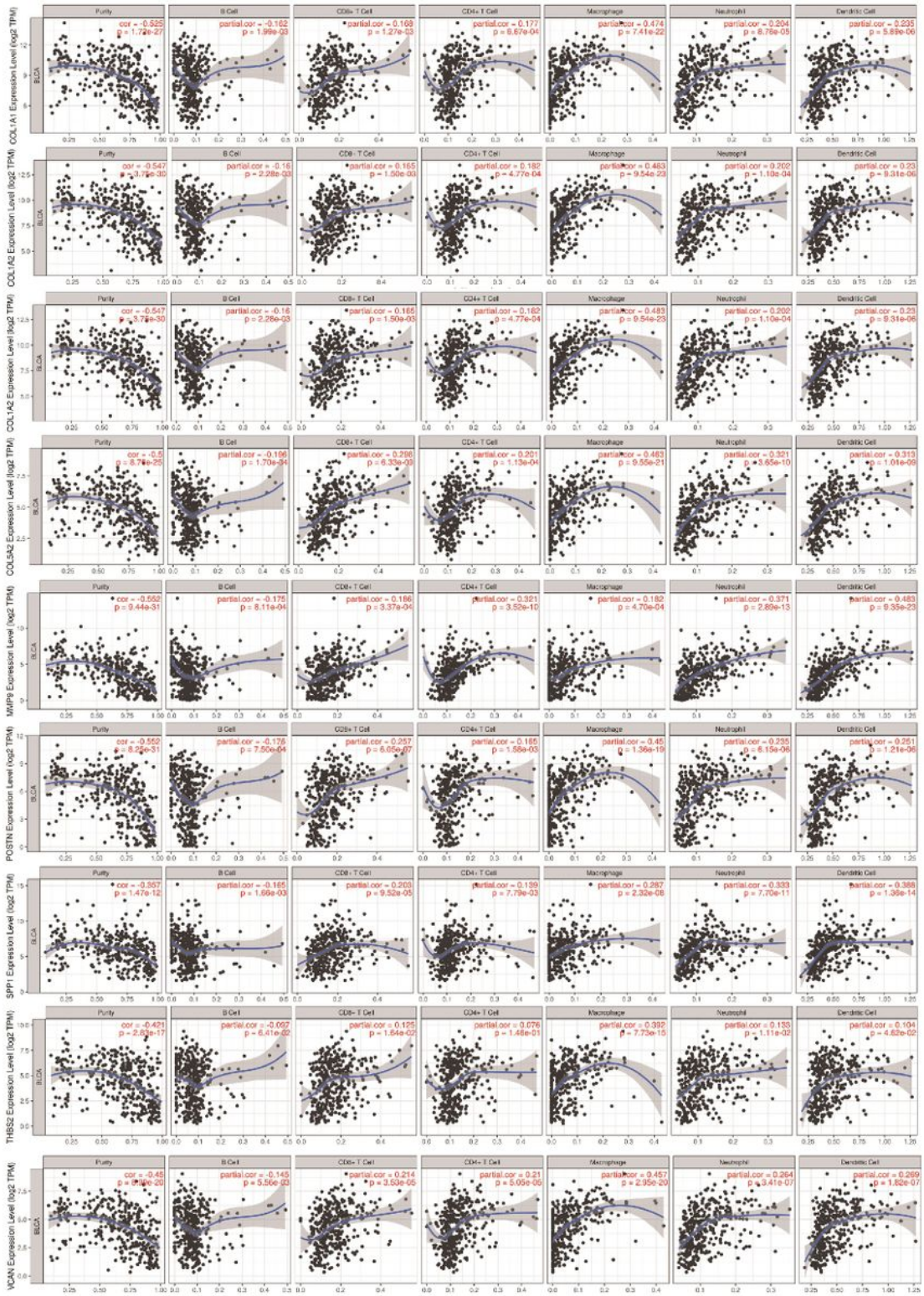

Figure 7

The correlation between differentially expressed hub genes and immune cell infiltration (TIMER database).

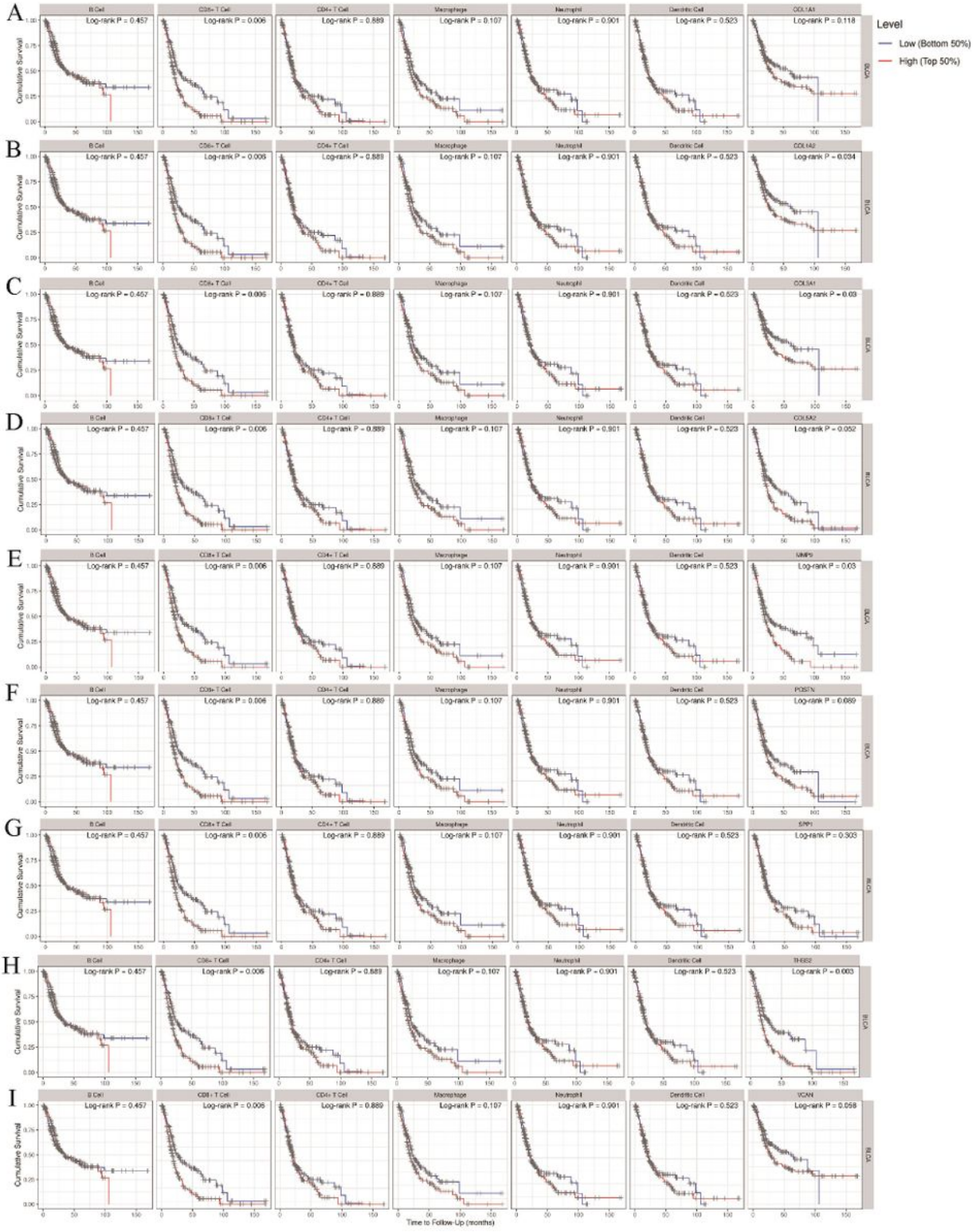

Figure 8

Cumulative survival was related to B cells, T cells, Macrophages, Neutrophils and DCs in UC. COL1A1 (A), COL1A2 (B), COL3A1 (C), COL5A2 (D), MMP9 (E), POSTN (F), SPP1 (G), THBS2 (H) and VCAN (I).

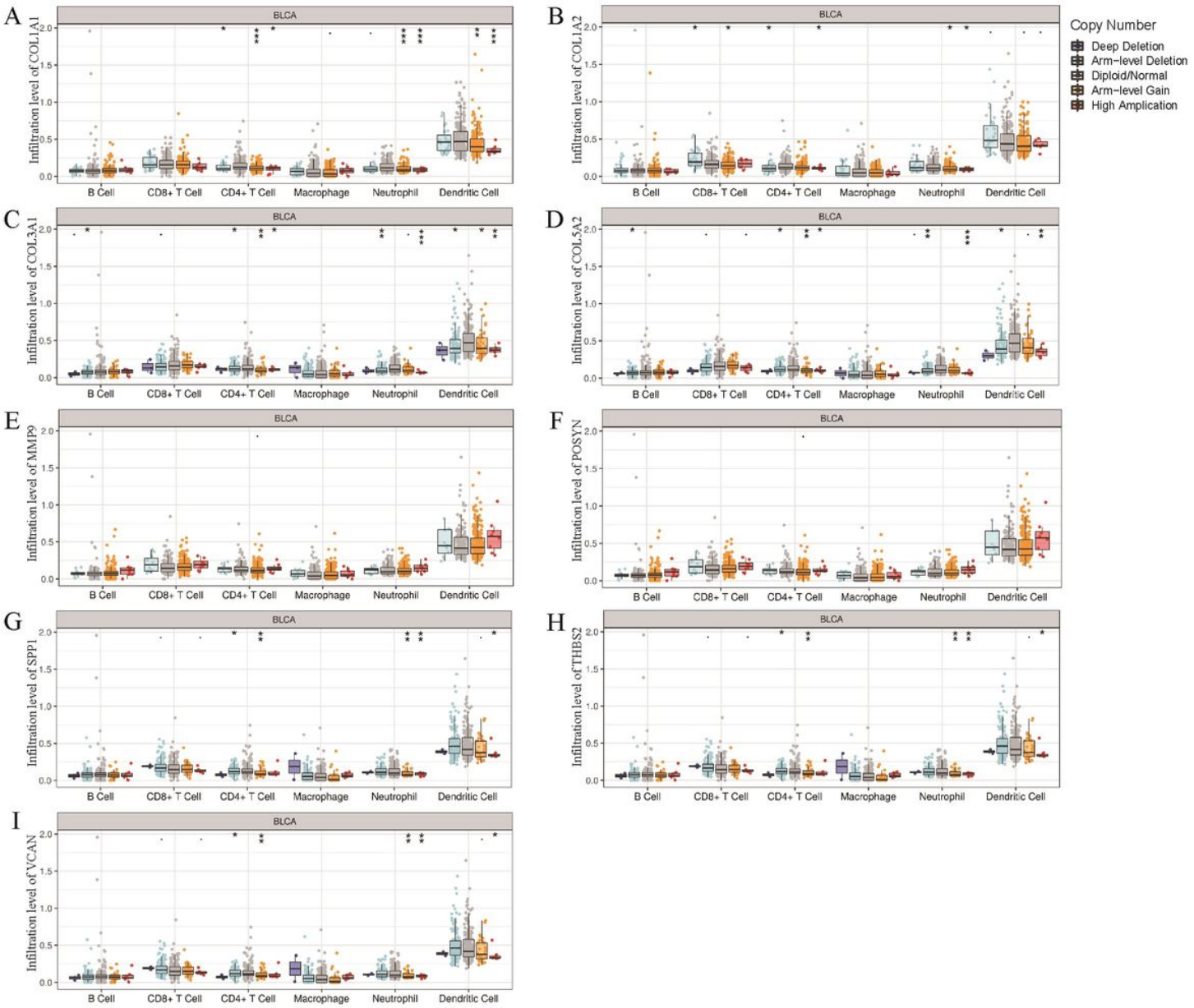

Figure 9

Association between hub gene copy number and immunocyte infiltration levels in UC cohorts. *P < 0.05, **p < 0.01, ***p < 0.001.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1.tif
- SupplementaryFigure2.tif
- SupplementaryFigure3.tif
- SupplementaryTable1.xls

- SupplementaryTable2.xls
- SupplementaryTable3.doc