

Identification of potential prognostic markers for osteosarcoma by integrated analysis

Yuxiang Ge

Department of Orthopedic Surgery, Minhang Hospital, Fudan University

Wang Ding

Department of Orthopedic Surgery, Minhang Hospital, Fudan University

Chong Bian

Department of Orthopedic Surgery, Minhang Hospital, Fudan University

Huijie Gu

Department of Orthopedic Surgery, Minhang Hospital, Fudan University

Jun Xu

Department of Orthopedic Surgery, Minhang Hospital, Fudan University

Xiaofan Yin (✉ xiaofanyinfdu@126.com)

Fudan University

Research article

Keywords: Osteosarcoma, Weighted gene co-expression network analysis (WGCNA), Prognosis, Hub gene identification

Posted Date: September 23rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-52760/v2>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Osteosarcoma (OS), one of the utmost common and malignant cancer, accounts for over 30% among skeletal sarcomas. Although great efforts have been made, the mechanism of OS still remains largely unknown. Here, we intend to identify gene modules and candidate biomarkers for clinical diagnosis of patients with OS, and reveal the mechanisms of OS progression.

Methods: Weighted gene co-expression network analysis (WGCNA) was conducted to build a co-expression network and investigate the relationship between modules and clinical traits. Functional enrichment analysis was performed on module genes. Protein-protein interaction (PPI) network was constructed to identify the hub gene and the expression level of hub genes was validated based on another dataset.

Results: A total of 9854 genes were included in WGCNA, and 17 gene modules were constructed. Gene module related with OS in sacrum was mainly enriched in skeletal system development, bone development and extracellular structure organization. Furthermore, we screened the top 10 hub genes and further validated 5 of the 10 (MMP13, DCN, GNG2, PCOLCE and RUNX2), the expression of which were upregulated as compared with normal tissues.

Conclusion: The hub gene we identified show great promise as prognostic markers for the management of OS and our findings also provide new insight for molecular mechanism of OS.

Introduction

Osteosarcoma (OS), as one of the utmost common and malignant cancer, mainly affects the adolescent and elderly individuals [1, 2]. OS can be divided into diverse sub-types associated with varying degrees of aggressiveness. Most OSs are high-grade malignancies, which is characterized by its aggressiveness and strong tendency for the development of pulmonary metastases[3]. The 5-year survival rate for patients with OS is less than 25%, indicating an extremely poor prognosis. Therefore, a better understanding of underlying mechanisms regarding the tumorigenesis and prognosis of OS plays a vital role to actively develop novel therapeutic targets[4, 5].

Previous researches have analyzed differentially expressed genes (DEGs) between metastatic and non-metastatic samples and functional enrichment analysis by microarray data of OS [6]. Several DEGs such as astrocyte elevated gene-1(AEG-1) and epidermal growth factor receptor (EGFR) were found to play an oncogenic role in the progression of OS[7, 8]. However, these analyses cannot reveal the connections and interactions among genes that are crucial in biological processes.

Weighted gene co-expression network analysis (WGCNA) is an effective tool to reveal the correlations between genes and phenotypes in different samples. It functions as a kind of method to investigate gene modules which includes associated genes and establishes a weighted gene co-expression network. WGCNA also serves a role to correlates gene modules of external sample trait [9]. As a result, WGCNA has

been globally applied to help discover potential biomarkers and therapeutic sites during various biological processes including cancer[10-12].

In our study, we proposed to build co-expression module by analyze the data of osteosarcoma and find the modules of co-expressed genes which affected specific sites of OS. We applied the GO and KEGG enrichment analysis on interested modules to investigate involved cell signal pathways and functional genes to understand the main gene function in specific module. Furthermore, module gene network was created and investigated hub genes in the module by Cytoscape software. Finally, we validated the top values of genes in the PPI network based on another data set and got 5 hub genes, such as MMP13, DCN, GNG2, PCOLCE and RUNX2, which may be associated with the OS development and clinical characteristics. Our study not only laid a theoretical foundation for further research and the discovery of new OS target biomarkers, but also helped in finding a novel and potential clinical therapy.

Materials And Methods

Data acquisition and processing

Series matrix files and data tables in microarray platform for WGCNA searched with keywords "Osteosarcoma" (Title) AND "Homo Sapiens"(Organism) were attained from NCBI Gene Expression Omnibus (GSE19276). The joint data set involved 47 osteosarcoma samples, including tibial group (n = 13), femoral group (n = 18), pelvic group (n = 7), Sacral group (n = 2), rib group (n = 2) and calcaneus group (n = 5). Before WGCNA analysis, we applied the array annotations to pair array probes with corresponding gene IDs. Then we removed the probes with multiple genes and counted mean expression of the gene values detected by multiple probes. An appropriate threshold was determined on the basis of different threshold of expression value. We applied WGCNA algorithm to assess the expression of different genes, and then we used flashClust to apply cluster analysis with proper threshold[9, 13, 14].

Construction of WGCNA and identification of modules

By constructing co-expression network gene modules related to osteosarcoma, we applied function soft Connectivity in WGCNA package to build a co-expression network in GSE19276 dataset. The top 5000 genes were screened by this algorithm for further analysis with the soft threshold set at 9. Weighted co-expression relationships among datasets was evaluated by paired Pearson correlations[9]. Hierarchical clustering analysis was performed by applying topological overlap matrix (TOM) and the clustering dendrogram was constructed by hierarchical clustering. Different branches in clustering dendrogram denoted diverse gene modules. All the genes were classified based on their expression patterns and weighted correlation coefficients of genes. Module was defined as the different genes which have similar patterns. Module-trait association was valued by matching each module and clinical data provided by GSE19276 dataset.

Functional and pathway enrichment of gene modules

Kyoto encyclopedia of genes genomes (KEGG) pathway analysis and gene ontology (GO) were performed to recognize the function and pathways regarding each module. The DAVID 6.8 (<https://david.ncifcrf.gov/>) database was used with the threshold $p < 0.05$ for the KEGG and GO [15, 16].

Hub Gene Analysis and Validation

Hub genes have been shown to be functionally significant and exceedingly connected with nodes among modules.

The protein-protein network (PPI) was constructed by uploading the genes among certain module to identify the importance of the top graded genes[17]. The Cytoscape software was further used to conduct the network analysis and we identified hub genes as genes of top 10 degree[18]. Independent gene expression profiles (GSE9460) containing osteosarcoma and healthy bone tissue were attained by the Gene Expression Omnibus. And the test sets were used to validate expression level of 10 genes with the highest degree. The dataset GSE9460 contained 15 primary osteosarcoma tissues and 3 healthy bone tissues.

Results

Data processing

A total of 47 tissue sample raw files were attained from NCBI Gene Expression Omnibus. Firstly, we performed Robust Multiarray Averaging (RMA) correction, transformed signals with \log_2 and normalized the data through quantile normalization. Then, median polishing probe sets were summarized by “affy” R package and 9854 probes were selected for further analysis.

Construction of Weighted Gene Co-Expression Network of osteosarcoma

The sample clustering dendrogram of osteosarcoma is shown in Fig. 1A. The soft-power threshold β was measured by scale independence and mean connectivity analysis of modules with different power values ranging from 1 to 14. As revealed in Figure 1B-C, $\beta = 9$ were selected for further analysis of the osteosarcoma. We detected gene modules based on the TOM matrix and 17 modules were generated in osteosarcoma for further analysis and labeled 1-17. The major module had 1440 genes with the minimum of 30 genes and module included 294 genes in average. Figure 2A-B showed the interaction relationship analysis among co-expression genes.

Module-clinical trait correlations

It is of importance to identify diverse genes associated with specific clinical feature so as to clarify hiding mechanism behind clinical trait. The clinical trait of osteosarcoma patients involving caecaneum, femur, pelvis, rib, sacrum and tibia were selected for the analysis. As is shown in Figure 3A, modules of similar expression forms related to specific clinical feature were recognized on the basis of eigengene and clinic traits correlation. In order to identify related eigengenes, the heatmap and eigengene dendrogram was

performed. As shown in Figure 3B-C, the dendrogram shows that green module significantly associated with osteosarcoma clinical traits (sacrum), and the scatterplot of Module Membership vs Gene Significance was plotted in green module (Figure 3D).

Functional enrichment analysis of genes in interested modules

In order to better understand the function of 266 candidate module genes, we used the online database DAVID, GO analysis, KEGG analysis using R package of DOSE and the clusterProfiler for pathway enrichment analysis[19, 20]. We conducted Go enrichment analysis in the selected green module[16]. As shown in Figure 4, the great majority of genes of green module were enriched in GO:0001501 (skeletal system development), GO:0060348 (bone development) and GO:0043062 (extracellular structure organization).

Module genes network analysis and hub genes identification

In order to understand the association between genes in the green module, we performed PPI network by Cytoscape software based on WGCNA. As shown in Figure 5, the co-expression networks of top-ranked genes for the selected green module was constructed. Genes interconnected with over 6 nodes were identified as hub nodes (MMP13, COL6A1, DCN, FAM20C, FBN1, GNG2, PCOLCE, PCOLCE2, RUNX2 and VCAN) and the expression level of above hub genes was depicted in Figure 6A.

Hub gene validation

The expression level of genes with the top 10 values were validated using the GSE9460 data set (Figure 6B). In the validation set, we found that 5 of the 10 genes (MMP13, DCN, GNG2, PCOLCE and RUNX2) were significantly up-regulated in osteosarcoma group ($p < 0.05$), consistent with the analysis of GSE19276.

Discussion

Osteosarcoma (OS) is one of the utmost common and malignant cancer, which mainly affects children and adolescents with the clinical feature of high metastatic potential. During the past several decades, despite great improvement in treatment strategies has been achieved to eliminate the local tumor, the survival rate for patients still gets unchanged in the long term [1, 2]. So as to improve survival rate of patients suffered from osteosarcoma, it is of great importance to put more efforts in underlying mechanism of osteosarcoma. Previous studied have demonstrated several target biomarkers of OS based on both gene expression profiling and micro-RNA profiling of the primary tumor[21]. Wang and Baumhoer et al. had analyzed the OS sample by RT-PCR or immunohistochemistry and observed that insulin-like growth factor-1 receptor (IGF-1R) was independent prognostic markers for OS and the high expression of IGF-1R was correlated with progression and metastasis of OS[22, 23]. A study performed by Duan et al. evaluated different miRNA in OS by Real-time PCR and reported that 3 miRNA were increased in osteoblast as compared with different kinds of OS cells, while 2 miRNAs were downregulated in

osteoblast [24]. Still, we still lack powerful prognostic biomarkers for patients with OS. WGCNA, an effective method to detect the complicated relationships between different genes and phenotypes, helps us appreciate signaling networks associated with clinical traits. The analysis of WGCNA helps us to identify modules related to biological process and hub genes, which might be biomarkers for identification or therapy[9]. By constructing WGCNA to detect gene expression of OS, Fan et al. identified matrix metalloproteinase 3 (MMP3) as a vital gene for OS development[25]. Likewise, Wang et al. reported the involvement of Matrix extracellular phosphoglycoprotein (MEPE) and Hemoglobin A2 (HBA2) in OS[26]. However, the precise mechanism of OS remains to be elucidated and more candidate biomarkers or therapeutic target needs to be discovered. Here, we utilized WGCNA to investigate co-expression modules associated with OS and forecasted candidate gene in specific bioprocess.

We first constructed 17 co-expression modules among 9854 genes in 47 OS samples via WGCNA, followed by module-trait correlations. As a result, we identified the green module that related to clinical traits (patient's OS location). Then the functional enrichment analysis was conducted in green module. Consistent with the previous study by Xiong and colleague, our result exhibited the green module was mainly enriched in pathway related to skeletal system development, bone development and extracellular structure organization[6]. Since the development of OS was accompanied with oncogenes and anti-oncogenes of abnormal levels, the enriched pathway that the green module involved in may regulate genes which were associated with skeletal system development. Heng et al. assessed the DEGs and microRNAs in OS samples and found that downregulated DEGs COL5A1 were enriched in cellular structure organization as well as extracellular collagen biosynthesis pathway[27]. Increased extracellular collagen degradation might promote the development of OS and metastasis to lungs, which indicated that genes involved in collagen synthesis might have a negative effect on extracellular matrix, which may be responsible for OS tumor invasion in an indirect manner.

In addition, PPI network was built and hub genes in OS-related regulatory network was recognized. In this study, we assessed the importance of genes via degree in network. The hub genes with highest degree were identified and COL6A1, MMP13, DCN, FAM20C, FBN1, GNG2, PCOLCE, PCOLCE2, RUNX2 and VCAN were viewed as hub genes. Furthermore, we validated 5 of 10 hub genes, MMP13, DCN, GNG2, PCOLCE, and RUNX2, had significantly high expression profile in the validation data set.

As is reported, numerous genes had correlations with OS. Matrix metalloproteinase-13(MMP13) had been recognized as one of the most important regulator involved in tumor osteolysis. Ma et al. identified MMP-13 as the target gene of proteolytic enzyme that regulated tumor-induced osteolysis[28]. Hirahata and colleagues showed that high MMP13 expression was of great importance for the survival of OS cells[29]. Decorin (DCN), Procollagen C-proteinase enhancer protein(PCOLCE) and Runx2 had an vital part in lung metastasis of osteosarcoma [30, 31]. Runx2 is a kind of protein that play major roles in bone formation. A study by Won assessed the expression of Runx2 in OS tissues and concluded that overexpression of Runx2 was associated with OS metastasis. Also, patients of osteosarcoma who had good reaction to chemotherapy usually had lower expression level of Runx2 compared with bad responders[32]. No reports have revealed that G Protein Subunit Gamma 2 (GNG2) was related to OS and further study was needed.

Conclusion

To sum up, the green module was viewed as a critical module correlated with progression of osteosarcoma patients. Furthermore, the hub genes in the module were confirmed through another dataset, which might help in identifying diagnostic and prognostic biomarkers of OS.

Declarations

Acknowledgments

We greatly appreciate the financial support from the National Natural Science Foundation of China (81772433).

Funding

This work was supported by grants from the National Natural Science Foundation of China (no. 81772433).

Availability of data and materials

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

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Authors' contributions

Conception and design: Yuxiang Ge, Xiaofan Yin.

Collection and assembly of data: Yuxiang Ge, Wang Ding, Chong Bian.

Data analysis and interpretation: Yuxiang Ge, Wang Ding, Chong Bian, Huijie Gu, Jun Xu.

Manuscript writing: Yuxiang Ge

Final approval of manuscript: All authors.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Abbreviations

WGCNA: weighted gene co-expression network analysis; OS: osteosarcoma; PPI: protein-protein interaction; DEGs: differentially expressed genes; EGFR: epidermal growth factor receptor; STAT3: signal transducer and activator of transcription 3; GEO: Gene Expression Omnibus; TOM: topological overlap matrix; GO: gene ontology; KEGG: Kyoto encyclopedia of genes genomes; STRING: Search Tool for the Retrieval of Interacting Gene; RMA: Robust Multiarray Averaging; RT-PCR: Reverse Transcription-Polymerase Chain Reaction; ME: module membership.

References

1. Klein MJ, Siegal GP. Osteosarcoma: anatomic and histologic variants. *Am J Clin Pathol.* 2006;125(4):555-81.
2. Wang LL. Biology of osteogenic sarcoma. *Cancer J.* 2005;11(4):294-305.
3. Kempf-Bielack B, Bielack SS, Jurgens H, Branscheid D, Berdel WE, Exner GU, et al. Osteosarcoma relapse after combined modality therapy: an analysis of unselected patients in the Cooperative Osteosarcoma Study Group (COSS). *J Clin Oncol.* 2005;23(3):559-68.
4. Conrad EU. ORTHOPAEDIC ONCOLOGY. DIAGNOSIS AND TREATMENT 2009.
5. Kansara M, Teng MW, Smyth MJ, Thomas DM. Translational biology of osteosarcoma. *Nat Rev Cancer.* 2014;14(11):722-35.
6. Xiong Y, Wu S, Du Q, Wang A, Wang Z. Integrated analysis of gene expression and genomic aberration data in osteosarcoma (OS). *Cancer Gene Ther.* 2015;22(11):524-9.
7. Pahl JH, Ruslan SE, Buddingh EP, Santos SJ, Szuhai K, Serra M, et al. Anti-EGFR antibody cetuximab enhances the cytolytic activity of natural killer cells toward osteosarcoma. *Clin Cancer Res.* 2012;18(2):432-41.
8. Fossey SL, Liao AT, McCleese JK, Bear MD, Lin J, Li PK, et al. Characterization of STAT3 activation and expression in canine and human osteosarcoma. *BMC Cancer.* 2009;9(1):81.
9. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9(1):559.
10. Wang H, Liu J, Li J, Zang D, Wang X, Chen Y, et al. Identification of gene modules and hub genes in colon adenocarcinoma associated with pathological stage based on WGCNA analysis. *Cancer Genet.* 2020;242:1-7.
11. Yin L, Cai Z, Zhu B, Xu C. Identification of Key Pathways and Genes in the Dynamic Progression of HCC Based on WGCNA. *Genes (Basel).* 2018;9(2):92.

12. Zhou XG, Huang XL, Liang SY, Tang SM, Wu SK, Huang TT, et al. Identifying miRNA and gene modules of colon cancer associated with pathological stage by weighted gene co-expression network analysis. *Onco Targets Ther.* 2018;11:2815-30.
13. Wan Q, Tang J, Han Y, Wang D. Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma. *Exp Eye Res.* 2018;166:13-20.
14. Ni Y, Zhang Z, Chen G, Long W, Tong L, Zeng J. Integrated analyses identify potential prognostic markers for uveal melanoma. *Exp Eye Res.* 2019;187:107780.
15. Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 2003;4(5):P3.
16. 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(1):25-9.
17. Zhu Y, Bian Y, Zhang Q, Hu J, Li L, Yang M, et al. Construction and analysis of dysregulated lncRNA-associated ceRNA network in colorectal cancer. *J Cell Biochem.* 2019;120(6):9250-63.
18. Wisniewski N, Cadeiras M, Bondar G, Cheng RK, Shahzad K, Onat D, et al. Weighted Gene Coexpression Network Analysis (WGCNA) Modeling of Multiorgan Dysfunction Syndrome after Mechanical Circulatory Support Therapy. *The Journal of Heart and Lung Transplantation.* 2013;32(4):S223.
19. Yu G, Wang LG, Yan GR, He QY. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics.* 2015;31(4):608-9.
20. 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.
21. Evola FR, Costarella L, Pavone V, Caff G, Cannavo L, Sessa A, et al. Biomarkers of Osteosarcoma, Chondrosarcoma, and Ewing Sarcoma. *Front Pharmacol.* 2017;8:150.
22. Wang YH, Han XD, Qiu Y, Xiong J, Yu Y, Wang B, et al. Increased expression of insulin-like growth factor-1 receptor is correlated with tumor metastasis and prognosis in patients with osteosarcoma. *J Surg Oncol.* 2012;105(3):235-43.
23. Baumhoer D, Elsner M, Smida J, Zillmer S, Rauser S, Schoene C, et al. CRIP1 expression is correlated with a favorable outcome and less metastases in osteosarcoma patients. 2011;2(12):970.
24. Duan Z, Choy E, Harmon D, Liu X, Susa M, Mankin H, et al. MicroRNA-199a-3p is downregulated in human osteosarcoma and regulates cell proliferation and migration. 2011;10(8):1337-45.
25. Fan H, Lu S, Wang S, Zhang S. Identification of critical genes associated with human osteosarcoma metastasis based on integrated gene expression profiling. *Mol Med Rep.* 2019;20(2):915-30.
26. Wang JS, Wang YG, Zhong YS, Li XD, Du SX, Xie P, et al. Identification of co-expression modules and pathways correlated with osteosarcoma and its metastasis. *World J Surg Oncol.* 2019;17(1):46.
27. Heng L, Jia Z, Bai J, Zhang K, Zhu Y, Ma J, et al. Molecular characterization of metastatic osteosarcoma: Differentially expressed genes, transcription factors and microRNAs. *Mol Med Rep.* 2017;15(5):2829-36.

28. Ma O, Cai WW, Zender L, Dayaram T, Shen J, Herron AJ, et al. MMP13, Birc2 (cIAP1), and Birc3 (cIAP2), amplified on chromosome 9, collaborate with p53 deficiency in mouse osteosarcoma progression. *Cancer Res.* 2009;69(6):2559-67.
29. Hirahata M, Osaki M, Kanda Y, Sugimoto Y, Yoshioka Y, Kosaka N, et al. PAI-1, a target gene of miR-143, regulates invasion and metastasis by upregulating MMP-13 expression of human osteosarcoma. *Cancer Med.* 2016;5(5):892-902.
30. Wang S, Zhong L, Li Y, Xiao D, Zhang R, Liao D, et al. Up-regulation of PCOLCE by TWIST1 promotes metastasis in Osteosarcoma. *Theranostics.* 2019;9(15):4342-53.
31. Shintani K, Matsumine A, Kusuzaki K, Morikawa J, Matsubara T, Wakabayashi T, et al. Decorin suppresses lung metastases of murine osteosarcoma. *Oncol Rep.* 2008;19(6):1533-9.
32. Massi D, De Nisi MC, Franchi A, Mourmouras V, Baroni G, Panelos J, et al. Inducible nitric oxide synthase expression in melanoma: implications in lymphangiogenesis. *Mod Pathol.* 2009;22(1):21-30.

Figures

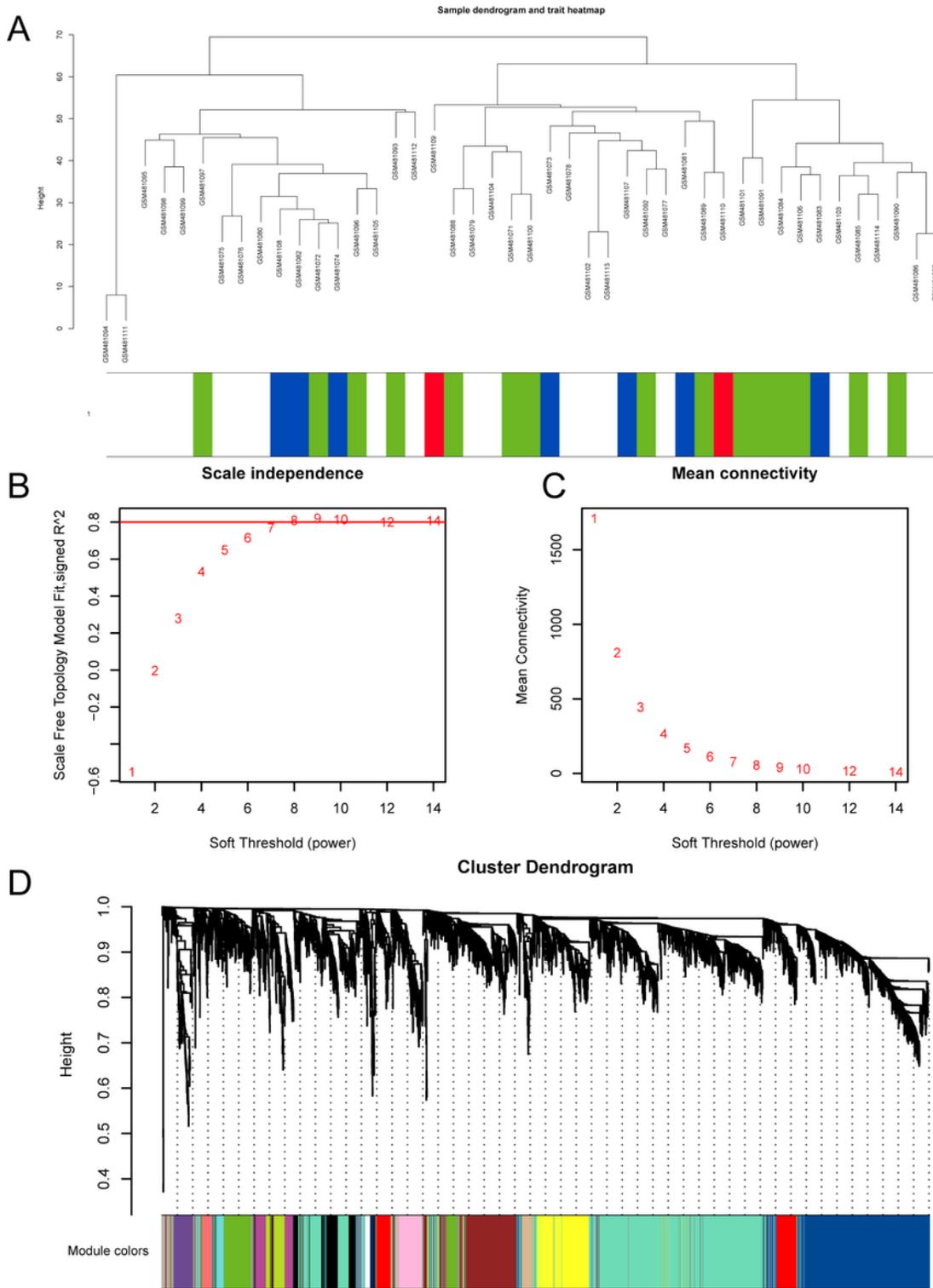


Figure 1

Weighted gene co-expression network analysis (WGCNA). (A) Sample dendrogram. (B-C) Scale independence and mean connectivity of network. The left part shows the correlation of soft-threshold with scale-free fit index. The right part shows the correlation of soft-threshold power on mean connectivity. (D) Clustering dendrograms of GSE19276.

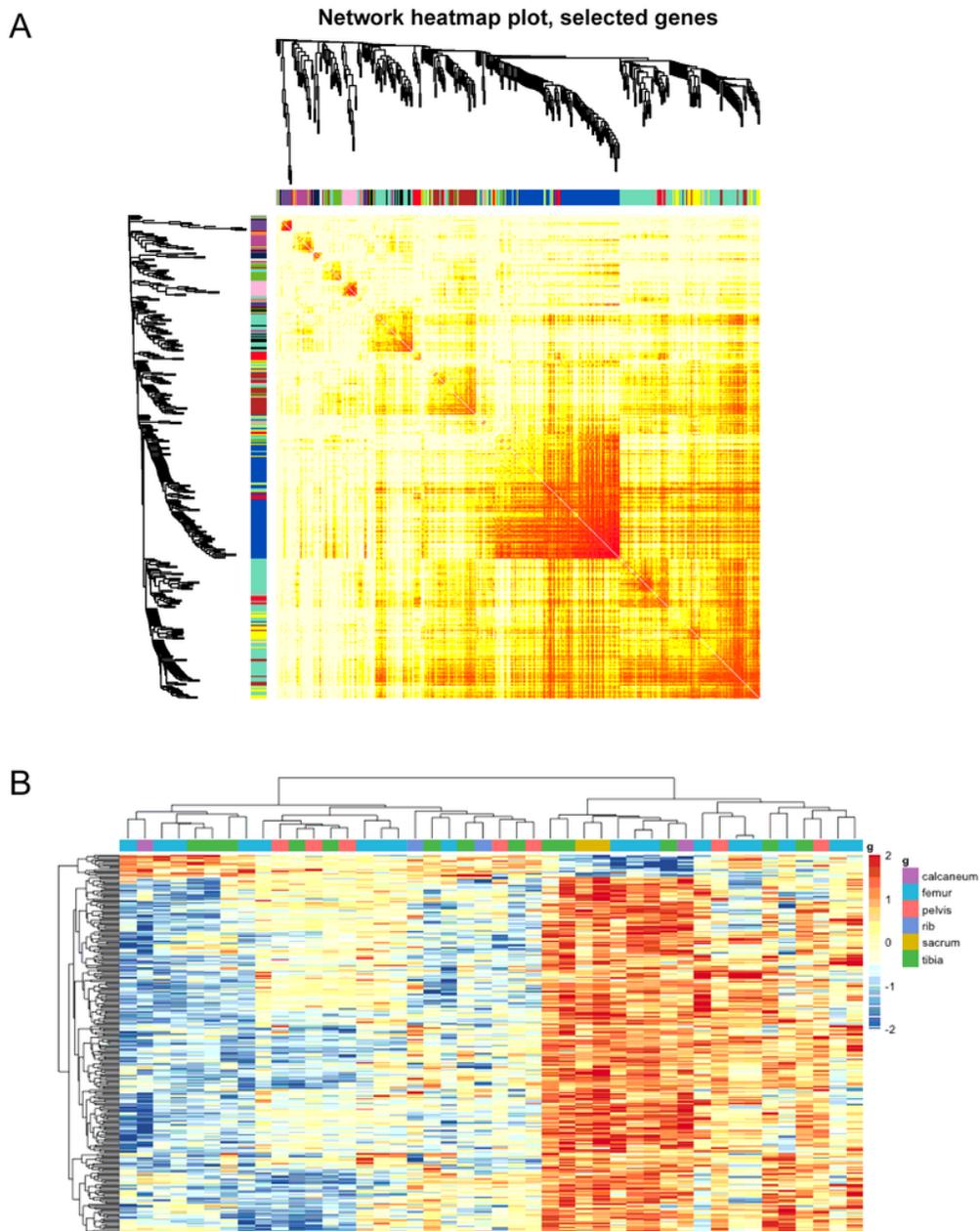


Figure 2

Visualizing the gene network using a heatmap plot. (A) Visualizing the gene network of topological overlap. Each row and column represents a gene. Light color indicates decreased topological overlap, and progressively darker indicates increased topological overlap. Darker squares along the diagonal represents different module. The gene dendrogram and module assignment are shown along the left and top. (B) Heatmap plot of gene expression in GSE19276.

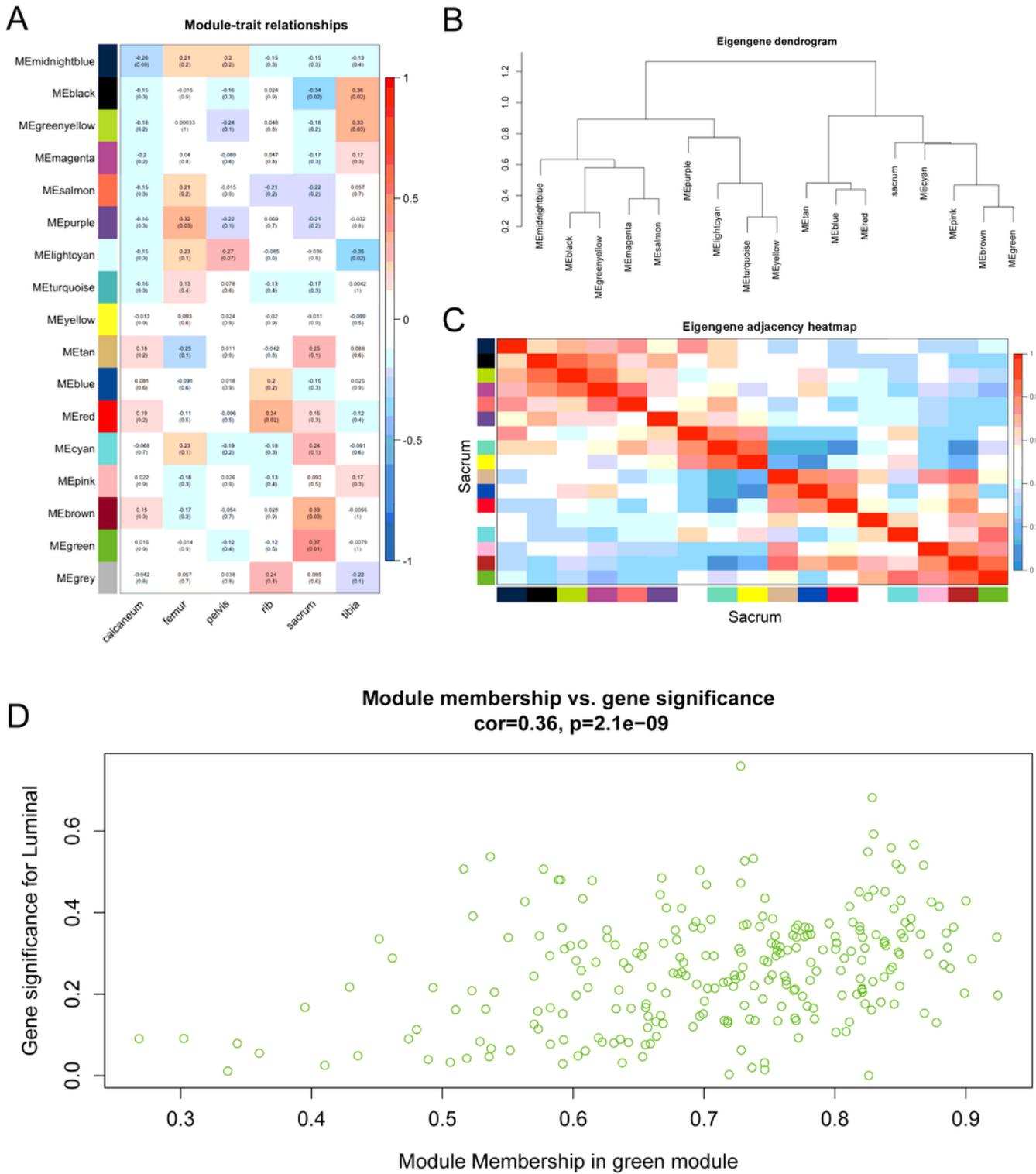


Figure 3

Correlation between module eigengenes and clinical trait of OS. (A) Module-trait relationships of WGCNA. Different gene modules in different colors represents diverse correlations. (B) Hierarchical clustering of module hub genes which gives summary of the modules in clustering analysis. (C) Eigengene adjacency heatmap between modules. (D) Scatterplots showing module membership versus gene significance for the green module.

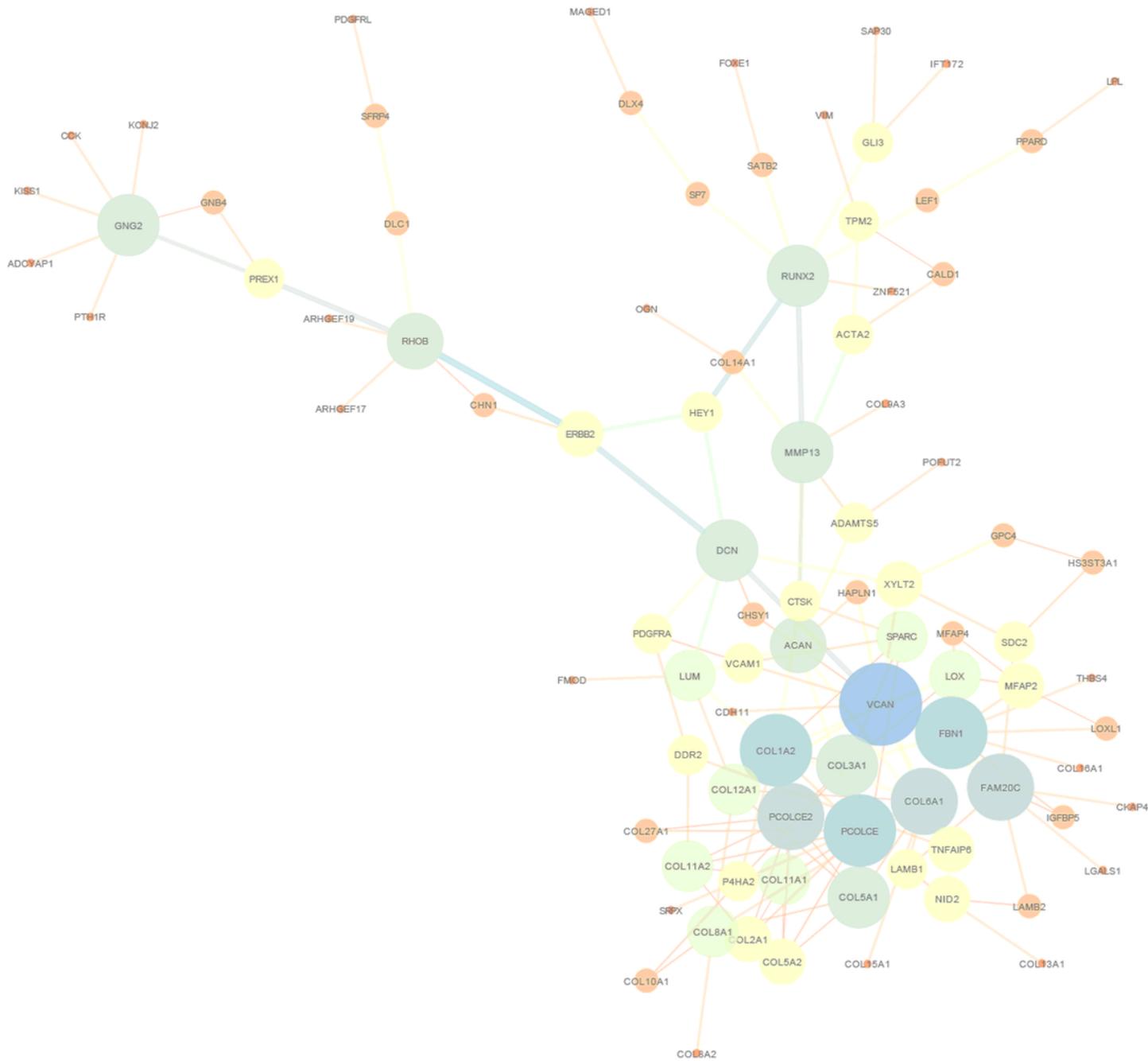


Figure 5

Protein-protein interaction (PPI) network analysis of genes in different expression patterns in the green module.

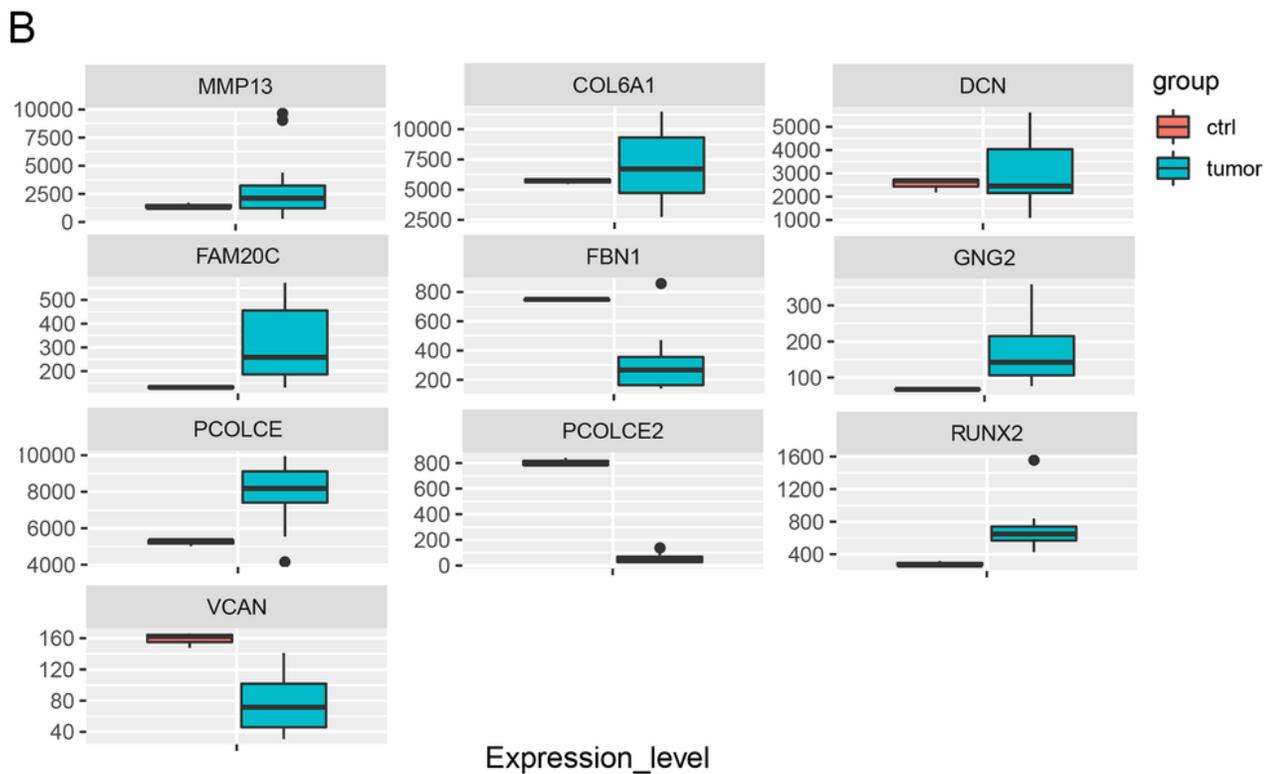
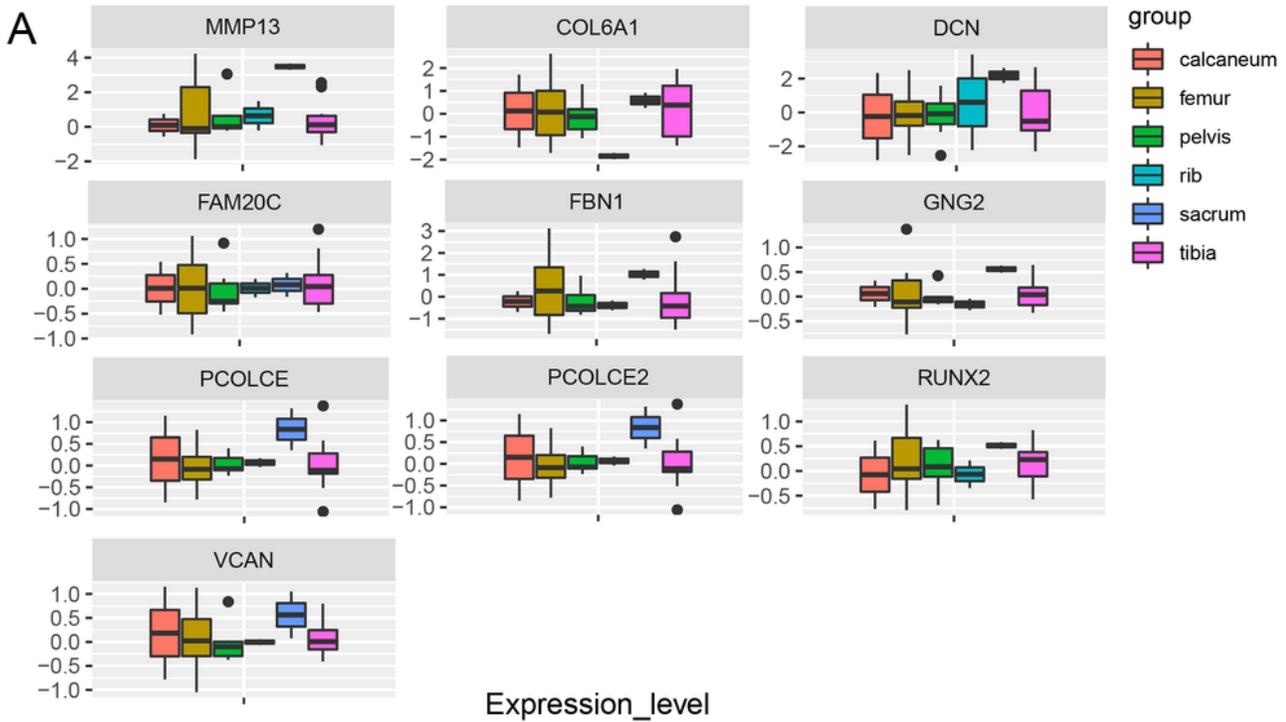


Figure 6

Validation of hub gene expression. (A) The expression of hub genes in OS patients for GSE19276 training set. (B) The expression of hub genes in 15 OS samples and 3 healthy controls for GSE9460 validation set.