

# Integrated Weighted Gene Co-expression Network Analysis Reveals ALB Associated With Prognosis of High-grade Serous Ovarian Cancer

**Bo Wang**

Shenzhen Baoan Women's and Children's Hospital, Jinan University

**Shan Chao**

Institutes for Shanghai Pudong Decoding Life

**Bo Guo** (✉ [bo\\_guo\\_baoan@163.com](mailto:bo_guo_baoan@163.com))

Maternal & Child Health Research Institute, Shenzhen Baoan Women's and Children's Hospital, Jinan University

---

## Research

**Keywords:** high-grade serous ovarian cancer, proteomics, WGCNA, hub gene, ALB

**Posted Date:** August 3rd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-761710/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** Ovarian cancer is the gynecologic tumor with the highest fatality rate, and high-grade serous ovarian cancer (HGSOC) is the most common and malignant type of ovarian cancer. One important reason for the poor prognosis of HGSOC is the lack of effective diagnostic and prognostic biomarkers. New biomarkers are necessary for improvement of treatment strategies and to ensure appropriate healthcare decisions.

**Methods:** To construct the co-expression network of HGSOC samples, we applied weighted gene co-expression network analysis (WGCNA) to assess the proteomic data downloaded from Clinical Proteomic Tumor Analysis Consortium (CPTAC), and module-trait relationship was then analyzed and plotted in a heat map to choose key module associated with HGSOC. Enrichment analysis was performed on the genes in the key modules to explore the functional information in which these genes participate. Hub genes with high connectivity in key module were identified by Cytoscape software. Furthermore, the true hub gene was selected through survival analysis, followed by expression verification with transcriptome dataset from TCGA and GTEx. Finally, the potential biological functions of hub gene were analyzed via single-gene Gene Set Enrichment Analysis (GSEA).

**Results:** After merging similar modules, a total of 9 modules were identified. Module-trait analysis revealed that the brown module ( $cor = 0.7$ ) was significantly associated with HGSOC. The results of enrichment analysis of the genes in the brown module show that these genes were related to the functions of the extracellular matrix, the complement system and the blood system. Ten hub genes with the highest connectivity were selected by protein-protein interaction analysis. After survival analysis and expression verification of hub genes, only ALB was confirmed to be the true hub gene and positively correlated with HGSOC prognosis. Single gene GSEA revealed that ALB was associated with cell material degradation.

**Conclusion:** We conducted the first gene co-expression analysis based on proteomic data from HGSOC samples, and found that ALB had prognostic value, which could be applied in the treatment of HGSOC in the future.

## Background

Ovarian cancer is the gynecologic cancer with highest mortality in the world, and ranks third in the incidence of gynecologic cancers, next to cervical cancer and endometrial cancer; 313,959 new cases and 207,252 deaths were estimated to have happened in 2020 (1). High-grade serous ovarian cancer (HGSOC), belonging to epithelial ovarian cancer (EOC), is the most common and fatal type of ovarian cancer, and the first-leading cause of ovarian cancer-related deaths (2–4). Since the ovaries are located deeply in the pelvic cavity and there are no specific symptoms and effective early screening methods of the early stages of HGSOC, most patients are often in an advanced clinical stage (stage III/IV) with metastasis at the time of diagnosis. The standard treatment of HGSOC is similar to other ovarian

cancers, which is to perform primary debulking surgeries followed by chemotherapy combining platinum and paclitaxel (2). But 80% of patients with advanced HGSOC will experience recurrence and chemotherapy resistance, leading to a 5-year survival rate of 30% (5), and more and more evidences indicate that the metastasis-prone characteristics of HGSOC play an important role in its relatively poor prognosis (6, 7). In advanced HGSOC, tumor cells have spread out from the ovaries and pelvic organs to the peritoneum and abdominal organs, which could impede the normal function of vital organs in the abdomen by uncontrolled proliferation, as well as function of circulatory and respiratory system by generating large amounts of ascites and increasing intra-abdominal pressure (5, 8), which eventually causing the HGSOC patients to die of various complications. However, patients with HGSOC present early stage tumor at diagnosis usually have a good prognosis after standard clinical interventions. Therefore, research on the biomarker of early diagnosis or prognosis of HGSOC is of great help to develop new effective therapies and improve the prognosis of patients with HGSOC.

Over the past two decades, with the breakthrough of high-throughput sequencing technology, a huge amount of sequencing data of various human tissues has been accumulated, which has greatly promoted the progress of biomedicine research, such as the screening of probable cancer prognostic biomarkers. Specifically in the field of ovarian cancer research, a variety of mRNAs or lncRNAs-based signatures have been identified for survival prediction in patients with ovarian cancer. For example, a study analyzed the differentially expressed genes in samples of ovarian cancer at different clinical stages and discovered specific gene co-expression modules related to the clinical stage and finally identified COL3A1, COL1A1, COL1A2, KRAS and NRAS as potential prognostic genes for ovarian cancer (9). Another study showed that 5 lncRNAs (LINC00664, LINC00667, LINC01139, LINC01419 and LOC286437) could be used as independent risk factors for ovarian cancer recurrence (10). Meanwhile, other researchers have proved that changes in expression levels of genes were related to platinum resistance in ovarian cancer (11). But all of these integrated analyses are performed at the transcriptome level, not at the proteome level. It is well known that protein is the main executor of life activities, and all life activities depend on the correct function of protein. And with the advancement of proteomics research methods and the accumulation of public proteomic data, performing integrated proteomics analysis is completely feasible at present.

In the post-genomic era, the consensus that the occurrence of complex diseases such as cancer is not determined by a single gene has gradually gained the approval of most researchers. Gene co-expression network research can provide us with information about gene expression correlations and potential functional relationships, thereby helping us understand biological systems and explore the relationship between the relevant functional genes (12). And weighted gene co-expression network analysis (WGCNA) is a method for concretizing this idea, which has been comprehensively applied to multiple cancer-associated studies to identify hub genes related to various traits.

Recent proteomics studies using ovarian tissue samples from HGSOC patients by Clinical Proteomic Tumor Analysis Consortium (CPTAC) have helped us better understand the tumor mechanism from a novel insight, and have also identified some candidate therapeutic targets (2, 13, 14). In this study, we

used WGCNA to re-analyze these published proteomic data in order to discover proteins and pathway related to occurrence and development of HGSOC, and identified a significant correlation between the brown module and HGSOC, clinical stage, histological grade and patient survival time. Ten hub genes were selected from this module and verified by survival analysis and other databases. Finally, it was identified that ALB might be a protein biomarker related to the prognosis of HGSOC. To our knowledge, this is the first study of prognostic biomarkers for HGSOC based on protein profile data, which provides us with some new insights into the occurrence and progress of HGSOC.

## Methods

### Proteomic data collection and pre-processing

The quantitative proteomic data and clinical information of HGSOC were obtained from CPTAC Data Portal ([https://cptac-data-portal.georgetown.edu/studies/filters/primary\\_site:Ovary](https://cptac-data-portal.georgetown.edu/studies/filters/primary_site:Ovary)). And the unshared peptides expression matrices analyzed through the Common Data Analysis Pipeline were used for subsequent data analysis. Samples lacking information about clinical stage, tumor histological grade and survival time and proteins with missing value of relative abundance among all samples were excluded from our study. Other sample inclusion criteria and data processing procedures were described in previous studies (2, 13, 14), in short, that were 1) five samples were removed causing without *TP53* mutation; 2) the median values of relative protein abundance over all proteins in every sample were calculated and re-centered to value of 0; 3) the normalized relative protein abundances of overlapping samples were averaged and used as their protein abundances. Finally, 2892 proteins were identified to perform WGCNA analysis.

### Co-expression network construction and module detection

We used the WGCNA package (Version 1.70, <https://CRAN.R-project.org/package=WGCNA>) to construct the co-expression network for the identified proteins (15). First, sample cluster analysis was carried out by the function `hclust` of WGCNA package to assess whether there were any significant outliers in the selected sample. Next, a suitable soft threshold power for scale-free network construction was calculated and chosen with the function `pickSoftThreshold` of WGCNA package. After that with the chosen soft threshold power value, an adjacency matrix was built to bring about weighted separation of co-expression (16). Co-expression similarity for paired proteins from adjacency matrix was calculated by the topological overlap dissimilarity measure, and we got a Topological Overlap Matrix (TOM) for next identification and similarity analysis of co-expression gene modules and combination of similar modules with the following major parameters: `deepSplit` of 2, `minModuleSize` of 15 and `mergeCutHeight` of 0.3. The resulting protein co-expression network was visualized as the heat map based on dissimilarity of TOM with hierarchical clustering dendrogram, and the number of proteins in each module was counted and plotted with the barplot (`ggplot2`, Version 3.3.5, <https://CRAN.R-project.org/package=ggplot2>).

### Identification of module-trait correlations and module preservation

The correlations between modules and clinical traits including sample type, clinical stage, pathological grade and survival time were assessed by Pearson Correlation Coefficients and a heat map was plotted to demonstrate the correlation value of interaction between modules and traits. The student *t*-test was used to get the *P* value of the correlation, and a *P* value of < 0.05 was considered as statistically significant. The brown module with highest value of Correlation Coefficients was mainly focused on and the correlation between Gene Significance (GS) for HGSOC and Module Membership (MM) in brown module was checked to identify module-trait associations.

## Functional annotation of modules

Gene ontology (GO) enrichment analysis and KEGG pathway enrichment analysis were performed for brown module via ClusterProfiler package (Version 3.16.1, <https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>) (17), and the top 5 results with adjusted *P* value of < 0.05 of enrichment analysis was visualized using chord diagram by GOplot package (Version 1.0.2, <https://CRAN.R-project.org/package=GOplot>) (18).

## Identification of hub genes

Genes closely connected to the intramodular nodes are regarded as hub genes which usually have more important biological function than other nodes (19). Protein-protein interaction (PPI) network analysis was performed via the online database Search Tool for Retrieval of Interacting Genes (STRING, Version 11, <https://string-db.org/>) (20), then the result of PPI analysis was imported to Cytoscape software (Version 3.8.0) to screen out hub genes with degrees ranking in the top 10 in the network of key modules using CytoHubba plug-in (Version 0.1) (21, 22).

## Kaplan-Meier survival analysis and validation of hub genes

Survival analysis was conducted and visualized for hub genes via the Survival package (Version 3.2–11, <https://CRAN.R-project.org/package=survival>) and Survminer package (Version 0.4.9, <https://cran.r-project.org/package=survminer>), respectively. The relative protein abundances and overall survival time from our data were used to plot the Kaplan-Meier curves. The cutoff values of the hub genes to separate the samples were determined by Survminer package. The hazard ratio (HR) was calculated with 95% confidence interval. Log-rank tests were performed to provide the statistical significance and *P* value of < 0.05 was considered as statistically significant. The protein level of hub genes between HGSOC and normal tissue were also validated using transcriptome data in The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) database and immunohistochemistry data in the Human Protein Atlas (HPA) database.

## Gene set enrichment analysis (GSEA)

GSEA of hub genes with our data was performed with GSEA software (Version 4.1.0) (23). The package “c5.all.v7.4.symbols.gmt” of the Molecular Signature Database (MSigDB, <https://www.gsea-msigdb.org/gsea/msigdb/>) was selected as reference gene set (24). The normalized enrichment scores and *P* value were generated, and *P* value of < 0.05 was considered as statistically significant.

## Results

### Identification of co-expression modules using WGCNA

It is believed that genes with comparable co-expression patterns are usually controlled by relative regulatory manner or have similar or parallel pathways of functional interaction (12). In this study, we downloaded the proteomic data from the CPTAC database according to Data Use Agreement. After the necessary quality control and manual check and screening, a matrix of relative abundance of 2892 proteins from 25 normal fallopian tube and 235 HGSOC samples with clinical information were selected to construct the co-expression networks. To ensure the reliability of the co-expression network, sample clustering analysis was performed to investigate the outliers among all samples, and no outliers were detected (Additional file 1). Finally, the relative abundances of these 2892 proteins and 260 samples were applied to identify the modules of co-expression genes (Additional file 2). To obtain scale free topology, a value of soft threshold power of 7 was selected based on scale independence analysis ( $R^2 = 0.927$ ), and the mean connectivity analysis was also relatively high under this soft threshold power (Fig. 1). Thirteen modules were generated and modules with high relevance of module eigengenes were combined for final 9 modules (Fig. 2a and b). The hierarchical clustering dendrogram of proteins also showed the same results of dynamic tree cut and merged dynamic (Fig. 2c). Numbers of proteins of each module were displayed in Fig. 2d and the detailed result was summarized in Additional file 3. Afterwards, the interactive relations among all modules and all proteins were visualized by a heat map plot based on TOM (Fig. 3).

### Identification of modules related to HGSOC

To determine whether co-expression modules were associated with sample types and clinical traits, the module-trait relationship analyses were performed (Fig. 4a). We identified that the brown module was positively related to HGSOC with the top relevance and survival time and negatively to stage and grade. In addition to brown module, tan module also showed a lower correlation with above traits compared to brown module. Then the relationship between GS for HGSOC and MM in brown module was performed and the Correlation Coefficients value was of 0.82 (Fig. 4b).

### Functional enrichment analysis of proteins in HGSOC-related module

To further explore the biological function of the proteins in the brown module related to HGSOC, GO and KEGG pathway enrichment analyses were performed for these proteins. The top 5 terms ranked by adjusted *P* value of enrichment analysis were plotted in chord diagram. The proteins in brown module were significantly enriched in term of extracellular matrix organization, extracellular structure organization, platelet degranulation, regulation of complement activation and complement activation of Biological Process category (Fig. 5a), and collagen-containing extracellular matrix, blood microparticle, vesicle lumen, secretory granule lumen and cytoplasmic vesicle lumen of Cellular Component category

(Fig. 5b), and extracellular matrix structural constituent, enzyme inhibitor activity, peptidase regulator activity, endopeptidase regulator activity and actin binding of Molecular Function category (Fig. 5c). And for the enrichment of KEGG pathway, these proteins were mainly enriched in complement and coagulation cascades, ECM-receptor interaction, focal adhesion, amoebiasis and carbon metabolism (Fig. 5d).

## Identification and validation of hub genes in HGSOC-related module

The 545 proteins in brown module were uploaded and analyzed by STRING to identify the interaction between them, and then the top 10 proteins (ALB, AKT1, APOB, C3, APOA1, FGA, FGG, SERPINA1, MAPK1 and AHSG) ranked by degree were selected as hub genes by cytoHubba plug-in of Cytoscape software. The network with neighbors generated by topological analysis method of Degree was shown in Fig. 6. Furthermore, the relationships between the 10 hub genes and the overall survival time of patients with HGSOC in our proteomic data were analyzed. Notably, low expression of SERPINA1, MAPK1, APOB and ALB significantly correlated with the poor survival time of HGSOC patients ( $P < 0.05$ ; Fig. 7a and Additional file 4). Meanwhile, we verified the expression levels of these 4 proteins in a transcriptome profile of HGSOC from the TCGA and GTEx database, which showed only the ALB was significantly expressed at lower level in HGSOC samples compared to normal samples (Fig. 7b and Additional file 5a-c), similar to the protein level of ALB from our proteomic data (Additional file 5d). Our results illustrated that ALB might be related to the risk of HGSOC. Finally, we also confirmed the protein level of ALB in HGSOC using the HPA database, and the result indicated that protein level of ALB in fallopian tube was significantly higher than that in HGSOC on the basis of immunohistochemistry (Fig. 8).

## Gene set enrichment analysis

To further understand the biological function of ALB in HGSOC, we performed GSEA based on our proteomic data. As shown in Fig. 9a and b, proteins in high-expression group of ALB were involved in “negative regulation of proteolysis”, “negative regulation of hydrolase activity”, “enzyme inhibitor activity”, “negative regulation of catalytic activity” and “regulation of proteolysis”, whereas “mRNA splice site selection”, “DNA biosynthetic process”, “regulation of signal transduction by P53 class mediator”, “signal transduction by P53 class mediator” and “post transcriptional regulation of gene expression” were enriched in low-expression group of ALB.

## Discussion

HGSOC remains the most common type of ovarian cancer with the highest incidence and the strongest fatality rate all over the world, and there is no definite research conclusion on its tumorigenesis mechanism. Meanwhile, due to the lack of effective early screening methods, most patients with high-grade serous ovarian cancer are at the advanced stage of tumor progression at the time of diagnosis, accompanied by extensive peritoneal metastasis, while most patients will experience tumor recurrence,

both of the above two factors together lead to a very poor prognosis for patients (2, 5). However, the 5-year survival rate of patients with early stages of HGSOC is as high as 92%, which is 62% higher than that with later stages of HGSOC (25), which suggests the possibility that patients with HGSOC can benefit from efficient early screening methods. Many researchers have conducted extensive research in this field, and have also discovered some novel diagnostic markers with clinical application value (9–11). But the objects of these studies are all at the transcriptome level, and protein, as a more direct manifestation of the life activities of cells, organs and even the body, may be a better research carrier for the screening of high-efficiency diagnostic or prognostic biomarkers. Furthermore, with the accumulation of proteomic data in public databases, it also provides more feasibility for this research idea.

WGCNA is a powerful method based on "guilt-by-association", and algorithms based on WGCNA transform the complexity of gene expression data sets into advantages, which can clarify gene relationships above the noise level (16). At present, WGCNA has been widely used in the field of tumor research, greatly promoting the biological interpretation of high-throughput sequencing data from clear cell renal cell carcinoma, pancreatic ductal adenocarcinoma (PDAC) as well as ovarian cancer, and assisting the discovery of tumor diagnosis and prognostic biomarkers (26–28). However, the number of studies applying WGCNA to analysis proteomic data is relatively small for the time being, but gratifying results have been achieved. Mantini et al. suggested that three proteins SPTBN1, KHSRP and PYGL might serve as potential prognostic biomarker for PDAC (29). While Johnson et al. identified proteins and biological processes in Alzheimer's disease brain that might serve as therapeutic targets and fluid biomarkers for the disease (30). These results also illustrate the effectiveness of WGCNA applied to proteomic data.

In this study, we obtained proteomic data of HGSOC samples from the CPTAC database. Then these data were assessed by WGCNA, and the brown module significantly related to HGSOC was identified. Interestingly, the brown module is not only significantly related to HGSOC, but also significantly related to the patient's survival time and significantly negatively related to the clinical stage and histological grade of HGSOC. The results of enrichment analysis of proteins in brown module show that most of these proteins are related to the organization and function of the extracellular matrix (ECM) components including collagen (COL1A1, COL1A2, COL4A1, COL4A2, COL6A1, COL6A2 etc.), proteoglycan (LUM, DCN), laminin (LAMA4, LAMB1, LAMB2, LAMC1) and other proteins as linkers to connect the above proteins (NID1, PRELP, TNXB), covering almost all types of ECM components, and most of these proteins were down-regulated in our data and were consistent with the results of previous studies (2). The metastasis-prone characteristics of HGSOC play an important role in its relatively poor prognosis (6, 7). Tumor invasion and metastasis is a complicated pathological process which involving interactions between tumor cells and various biologically active molecules from tumor microenvironment including ECM (31), and for malignant tumor cells derived from epithelial cells like HGSOC, epithelial-mesenchymal transition (EMT) is the key first biological step for these tumors cells to metastasize, which is accompanied by disorders of ECM composition and organization, and in turn enhancing tumor cell mobility and protecting tumor cells from immune attack via collagen remodeling, and finally promoting the invasion and metastasis of HGSOC (5, 32, 33). These facts not only prove the credibility of the

correlation between the brown module and the clinical stage, histological grade and survival time of patients, but also indirectly prove the validity of our analysis results.

Moreover, the PPI analysis was conducted to screen hub genes, which were further performed with survival analysis using our data and verified by a transcriptome dataset of HGSOC from the TCGA and GTEx database due to the lack of other available proteomic data. Finally, only ALB showed a good consistence between expression pattern and correlation with the patient's prognosis, that was, the expression level of ALB in tumor tissues was significantly down-regulated and had a significantly positive correlation with the patient's prognosis. Therefore, we speculate that ALB may be involved in the prognosis of HGSOC patients. ALB encodes the secreted and main protein of human blood, lymph, cerebrospinal and interstitial fluid, which plays important roles in a variety of physiological functions, such as maintaining colloid osmotic pressure and acting as the carrier protein for a wide range of endogenous molecules including hormones, fatty acids, as well as exogenous drugs (34). ALB has been reported to participate in the development and treatment of tumors with different mechanisms. First, previous studies have shown that the production of ALB is significantly inhibited during cancer-related systemic inflammation, which is regulated by a variety of cytokines and growth factors produced by tumor cells and immune cells (35). Secondly, the decrease in plasma ALB concentration reflects the poor nutritional condition of cancer patients, which may be related to the chemotherapy resistance of patients (36). Further, ALB might take part in antioxidant and anticancer effect (37, 38). In addition, the low serum ALB concentration is related to the poor survival time of patients with various cancers (35), as well as for ovarian cancer (39). However, these studies have not clarified the definite mechanism of how low plasma ALB levels lead to poor prognosis of patients. Meanwhile, there is no report about the relationship between ALB and the prognosis of HGSOC, and our research fills this gap. Moreover, the samples in our study are from tumor tissues of HGSOC patients, which may provide a certain research basis for the subsequent understanding of the role of ALB in the occurrence of HGSOC.

The limitation of this study is that the verification of our results was carried out in the transcriptome data due to the lack of other independent proteomic data. Secondly, due to the limitation of mass spectrometry technology, the relative abundance of most proteins is missing, and in order to ensure the reliability of the results, these proteins are not included into our analysis, which may cause bias to our conclusion.

## Conclusions

In summary, HGSOC-associated module was revealed through WGCNA. ALB was identified as a prognostic biomarkers for HGSOC, and its protein level was positively correlated with the survival time of HGSOC patients. These findings need to be confirmed by further clinical practices in the future.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### **Consent for publication**

Not applicable.

### **Availability of data and materials**

All data generated or analysed during this study are included in this published article and its supplementary information files.

### **Competing interests**

The authors declare that they have no competing interests

### **Funding**

No funding was obtained for this study.

### **Authors' contributions**

BW and BG conceived this study; BW and SC performed the analysis; BW and BG prepared and revised the manuscript. All authors have read and approved the submitted manuscript.

### **Acknowledgements**

Data used in this publication were generated by the Clinical Proteomic Tumor Analysis Consortium (NCI/NIH). We thank Haifeng Li from Institutes for Shanghai Pudong Decoding Life for assistance and valuable discussion with the data analysis.

## **References**

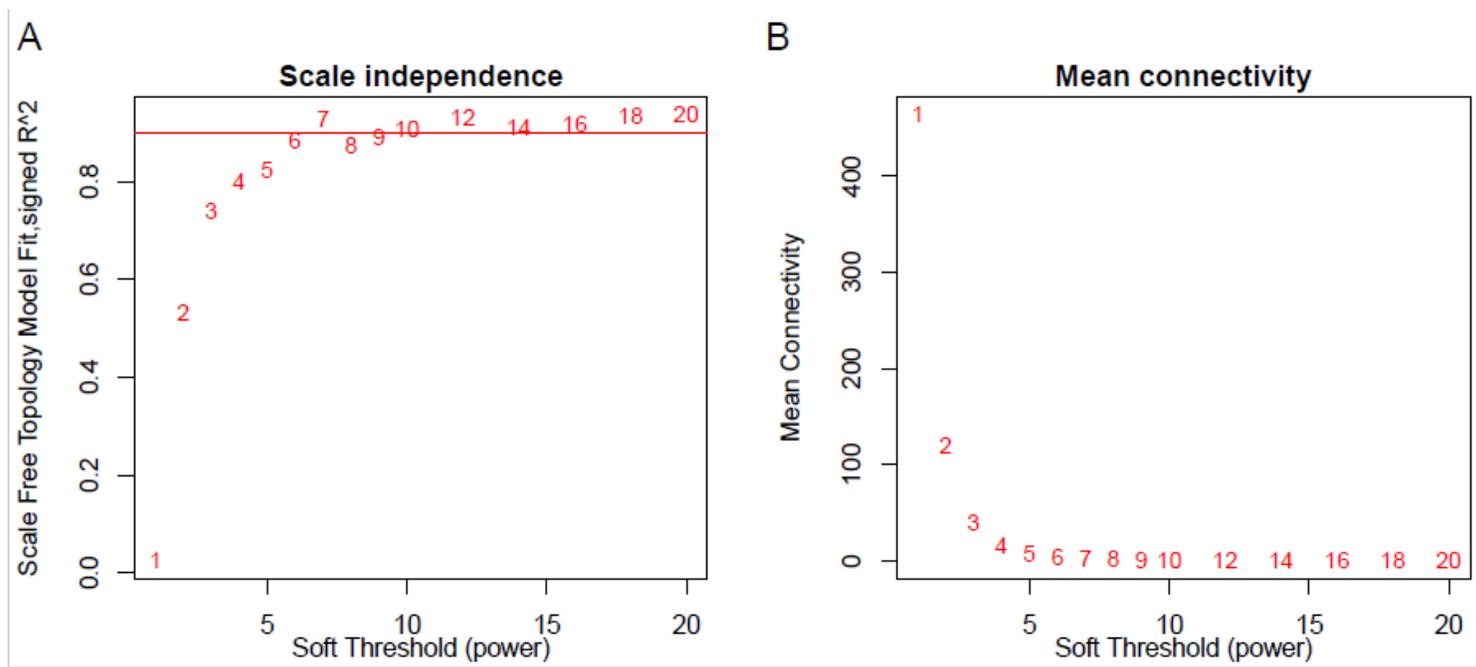
1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209–49.
2. Hu Y, Pan J, Shah P, Ao M, Thomas SN, Liu Y, et al. Integrated Proteomic and Glycoproteomic Characterization of Human High-Grade Serous Ovarian Carcinoma. *Cell reports.* 2020;33(3):108276.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics. 2018. *CA Cancer J Clin.* 2018;68(1):7–30.
4. Torre LA, Trabert B, DeSantis CE, Miller KD, Samimi G, Runowicz CD, et al. Ovarian cancer statistics, 2018. *CA Cancer J Clin.* 2018;68(4):284–96.
5. Natarajan S, Foreman KM, Soriano MI, Rossen NS, Shehade H, Fregoso DR, et al. Collagen Remodeling in the Hypoxic Tumor-Mesothelial Niche Promotes Ovarian Cancer Metastasis. *Cancer Res.* 2019;79(9):2271–84.

6. Zhang J, Liu XH, Li C, Wu XX, Chen YL, Li WW, et al. SNCG promotes the progression and metastasis of high-grade serous ovarian cancer via targeting the PI3K/AKT signaling pathway. *Journal of experimental clinical cancer research: CR*. 2020;39(1):79.
7. Yeung TL, Leung CS, Yip KP, Au Yeung CL, Wong ST, Mok SC. Cellular and molecular processes in ovarian cancer metastasis. A Review in the Theme: Cell and Molecular Processes in Cancer Metastasis. *American journal of physiology Cell physiology*. 2015;309(7):C444-56.
8. Tan DS, Agarwal R, Kaye SB. Mechanisms of transcoelomic metastasis in ovarian cancer. *Lancet Oncol*. 2006;7(11):925–34.
9. Yang L, Jing J, Sun L, Yue Y. Exploring prognostic genes in ovarian cancer stage-related coexpression network modules. *Medicine*. 2018;97(34):e11895.
10. Chen Y, Bi F, An Y, Yang Q. Identification of pathological grade and prognosis-associated lncRNA for ovarian cancer. *Journal of cellular biochemistry*. 2019;120(9):14444–54.
11. Zhang L, Zhang X, Fan S, Zhang Z. Identification of modules and hub genes associated with platinum-based chemotherapy resistance and treatment response in ovarian cancer by weighted gene co-expression network analysis. *Medicine*. 2019;98(44):e17803.
12. Topno R, Singh I, Kumar M, Agarwal P. Integrated bioinformatic analysis identifies UBE2Q1 as a potential prognostic marker for high grade serous ovarian cancer. *BMC Cancer*. 2021;21(1):220.
13. Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, et al. Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell*. 2016;166(3):755–65.
14. McDermott JE, Arshad OA, Petyuk VA, Fu Y, Gritsenko MA, Clauss TR, et al. Proteogenomic Characterization of Ovarian HGSC Implicates Mitotic Kinases, Replication Stress in Observed Chromosomal Instability. *Cell reports Medicine*. 2020;1(1).
15. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. 2008;9:559.
16. Zhang GH, Chen MM, Kai JY, Ma Q, Zhong AL, Xie SH, et al. Molecular profiling of mucinous epithelial ovarian cancer by weighted gene co-expression network analysis. *Gene*. 2019;709:56–64.
17. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology*. 2012;16(5):284–7.
18. Walter W, Sanchez-Cabo F, Ricote M. GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics*. 2015;31(17):2912–4.
19. Yuan Y, Chen J, Wang J, Xu M, Zhang Y, Sun P, et al. Identification Hub Genes in Colorectal Cancer by Integrating Weighted Gene Co-Expression Network Analysis and Clinical Validation in vivo and vitro. *Front Oncol*. 2020;10:638.
20. 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 research*. 2019;47(D1):D607-D13.

21. 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 research*. 2003;13(11):2498–504.
22. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC systems biology*. 2014;8(Suppl 4):11.
23. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50.
24. Liberzon A. A description of the Molecular Signatures Database (MSigDB) Web site. *Methods in molecular biology*. 2014;1150:153–60.
25. Ahn HS, Yeom J, Yu J, Kwon YI, Kim JH, Kim K. Convergence of Plasma Metabolomics and Proteomics Analysis to Discover Signatures of High-Grade Serous Ovarian Cancer. *Cancers*. 2020;12(11).
26. Wang Y, Chen L, Wang G, Cheng S, Qian K, Liu X, et al. Fifteen hub genes associated with progression and prognosis of clear cell renal cell carcinoma identified by coexpression analysis. *Journal of cellular physiology*. 2019;234(7):10225–37.
27. Su Q, Zhu EC, Qu YL, Wang DY, Qu WW, Zhang CG, et al. Serum level of co-expressed hub miRNAs as diagnostic and prognostic biomarkers for pancreatic ductal adenocarcinoma. *J Cancer*. 2018;9(21):3991–9.
28. Chen Y, Bi F, An Y, Yang Q. Coexpression network analysis identified Kruppel-like factor 6 (KLF6) association with chemosensitivity in ovarian cancer. *Journal of cellular biochemistry*. 2018.
29. Mantini G, Valles AM, Le Large TYS, Capula M, Funel N, Pham TV, et al. Co-expression analysis of pancreatic cancer proteome reveals biology and prognostic biomarkers. *Cellular oncology*. 2020;43(6):1147–59.
30. Johnson ECB, Dammer EB, Duong DM, Ping L, Zhou M, Yin L, et al. Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nat Med*. 2020;26(5):769–80.
31. Melzer C, Yang Y, Hass R. Interaction of MSC with tumor cells. *Cell communication signaling: CCS*. 2016;14(1):20.
32. Cheon DJ, Tong Y, Sim MS, Dering J, Berel D, Cui X, et al. A collagen-remodeling gene signature regulated by TGF-beta signaling is associated with metastasis and poor survival in serous ovarian cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2014;20(3):711–23.
33. Pearce OMT, Delaine-Smith RM, Maniati E, Nichols S, Wang J, Bohm S, et al. Deconstruction of a Metastatic Tumor Microenvironment Reveals a Common Matrix Response in Human Cancers. *Cancer discovery*. 2018;8(3):304–19.
34. Gburek J, Konopska B, Golab K. Renal Handling of Albumin-From Early Findings to Current Concepts. *International journal of molecular sciences*. 2021;22(11).

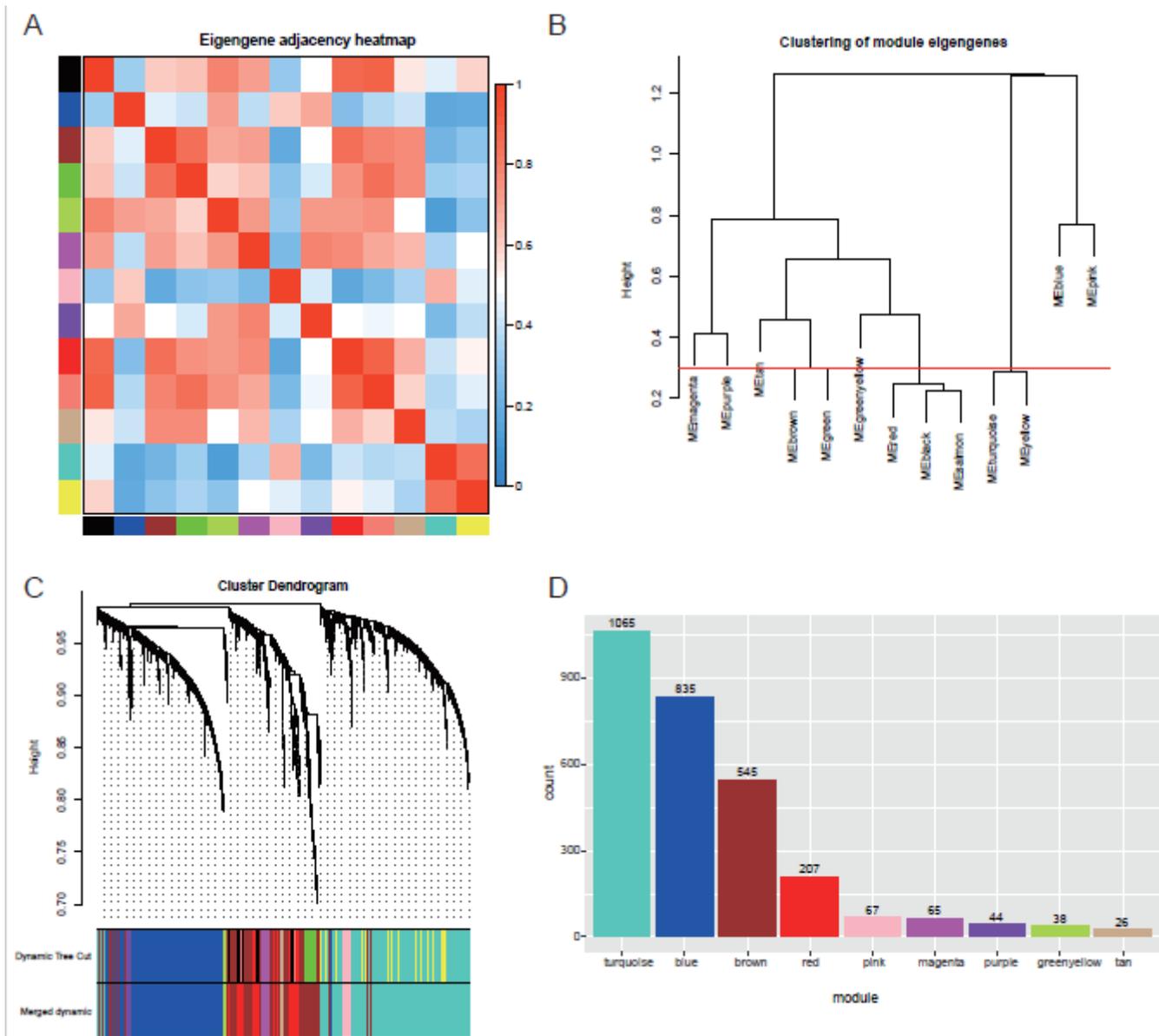
35. Lei J, Wang Y, Guo X, Yan S, Ma D, Wang P, et al. Low preoperative serum ALB level is independently associated with poor overall survival in endometrial cancer patients. *Future oncology*. 2020;16(8):307–16.
36. Ward EJ, Henry LM, Friend AJ, Wilkins S, Phillips RS. Nutritional support in children and young people with cancer undergoing chemotherapy. *The Cochrane database of systematic reviews*. 2015(8):CD003298.
37. Laursen I, Briand P, Lykkesfeldt AE. Serum albumin as a modulator on growth of the human breast cancer cell line, MCF-7. *Anticancer research*. 1990;10(2A):343–51.
38. Seaton K. Albumin concentration controls cancer. *J Natl Med Assoc*. 2001;93(12):490–3.
39. Schwartz GG, Tretli S, Klug MG, Robsahm TE. Women who develop ovarian cancer show an increase in serum calcium and a decrease in serum albumin. A longitudinal study in the Janus Serum Bank Cohort. *Gynecol Oncol*. 2020;159(1):264–9.

## Figures



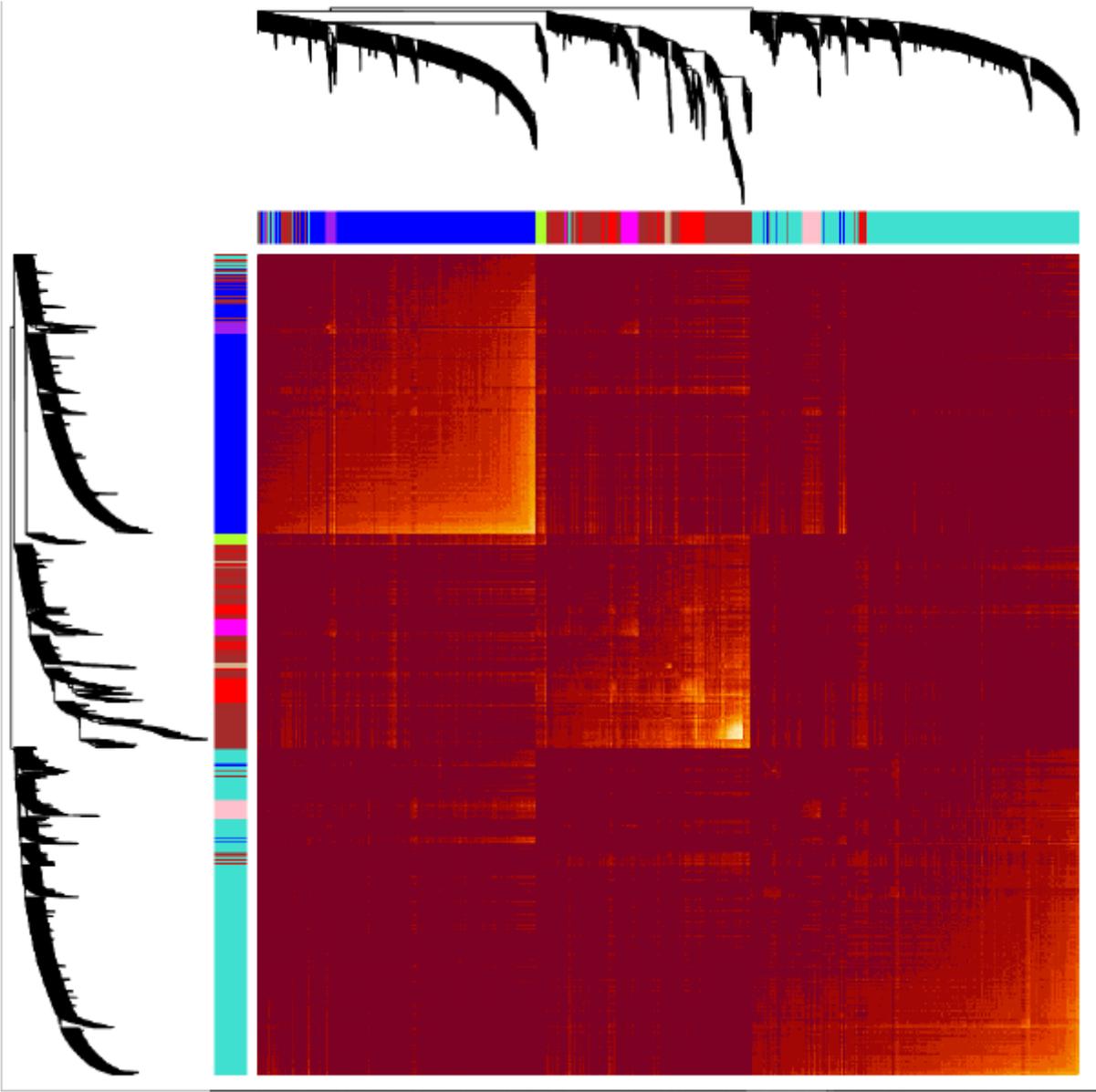
**Figure 1**

The scale independence and mean connectivity of co-expression modules based on different soft threshold.



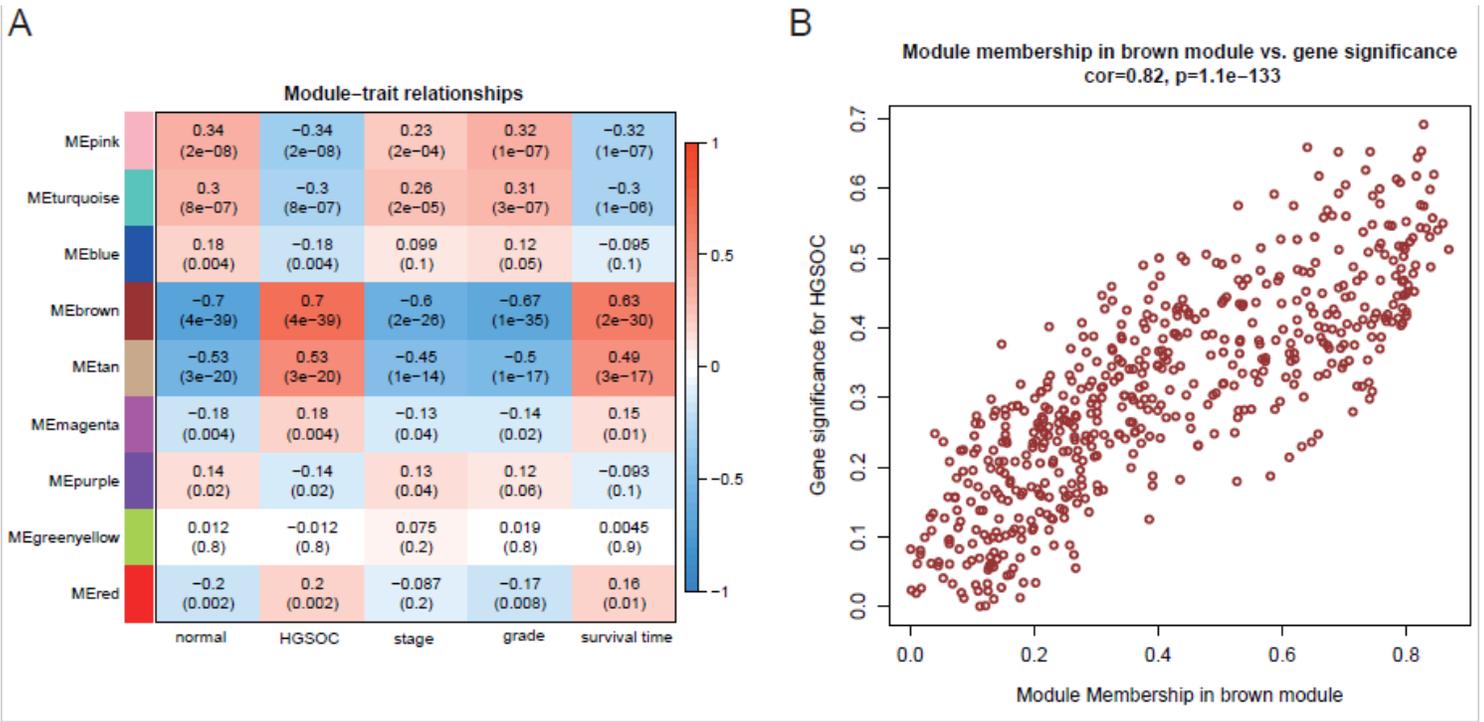
**Figure 2**

The identification of co-expression modules using WGCNA. (A) Eigengene adjacency analysis of different modules plotted with a heat map. Red represents a high correlation and blue represents a low correlation. (B) Hierarchical cluster analysis of different modules. The red line represents cut height of 0.3. (C) Hierarchical clustering of genes with dissimilarity based on topological overlap is shown with the modules detected and the merged modules. (D) Count of gene in different modules.



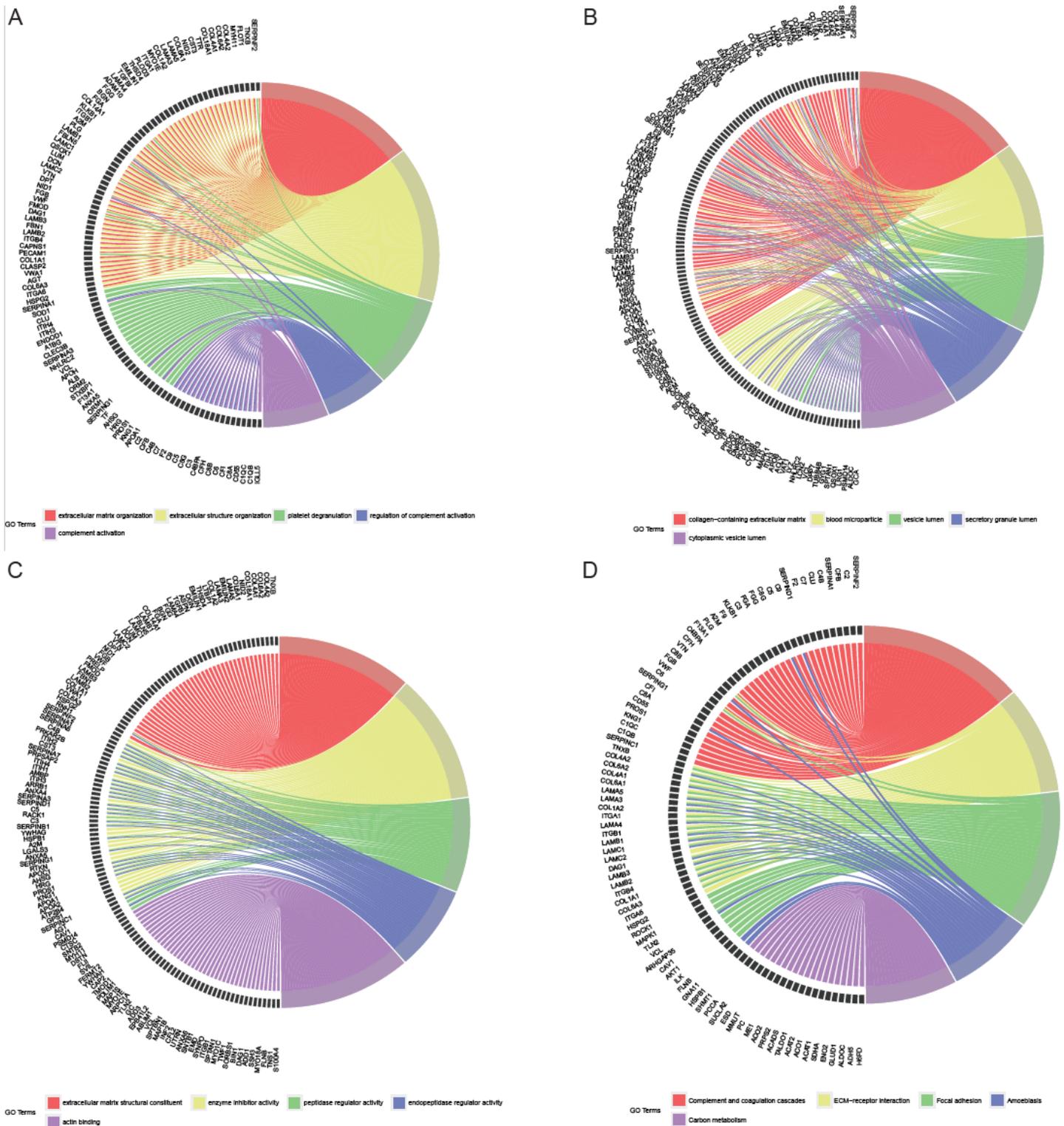
**Figure 3**

Interaction analysis of co-expression genes in different modules. The different colors of the horizontal and vertical axes represent different modules. The yellow brightness in the middle indicates the degree of correlation between different modules based on Topological Overlap Matrix (TOM).



**Figure 4**

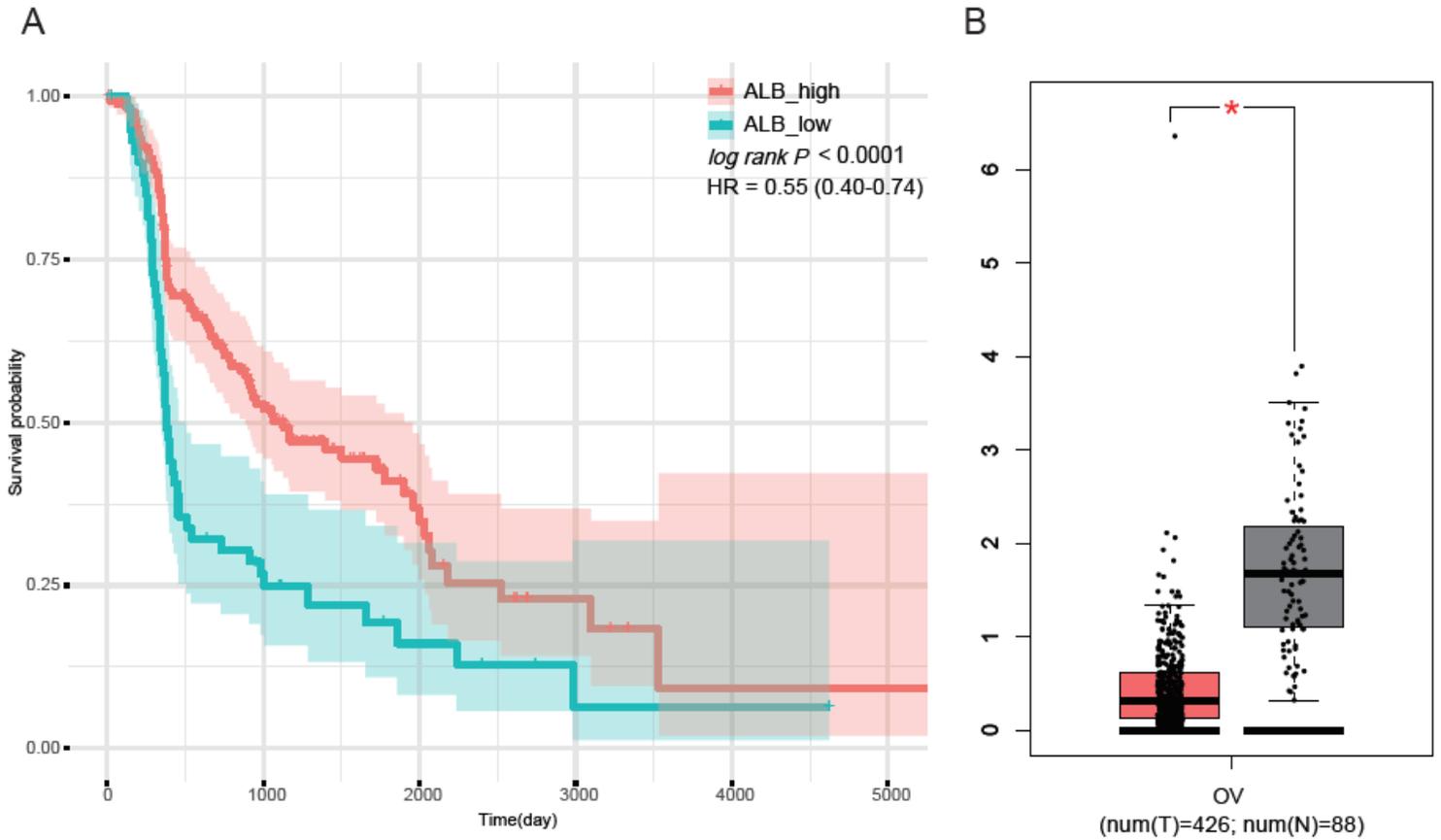
Module-trait relationship analysis. (A) The relationship between different modules and trait including sample type, stage, grade and survival time was plotted in a heat map. Red represents a positive correlation and blue represents a negative correlation. (B) Gene significance for HGSOC versus the module membership in the brown module is depicted as a scatter plot. The correlation is calculated through Pearson correlation analysis.



**Figure 5**

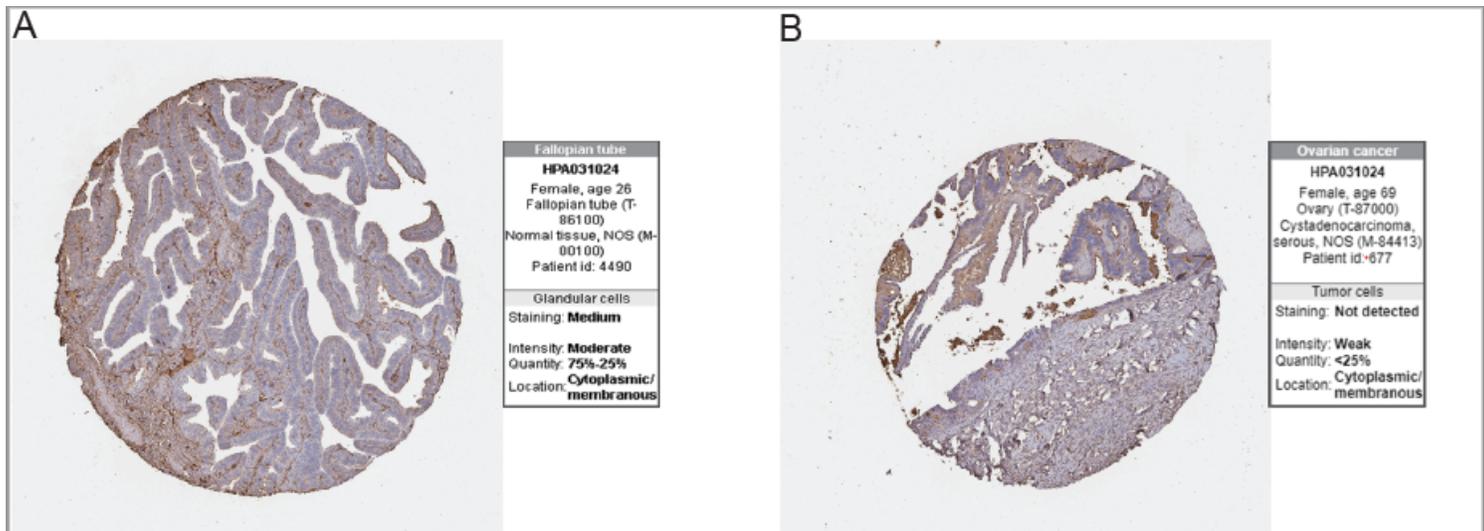
Enrichment analysis of proteins in brown module with the chord diagram. (A) Biological process category of GO enrichment analysis. (B) Cellular component category of GO enrichment analysis. (C) Molecular function category of GO enrichment analysis. (D) KEGG pathway enrichment analysis.





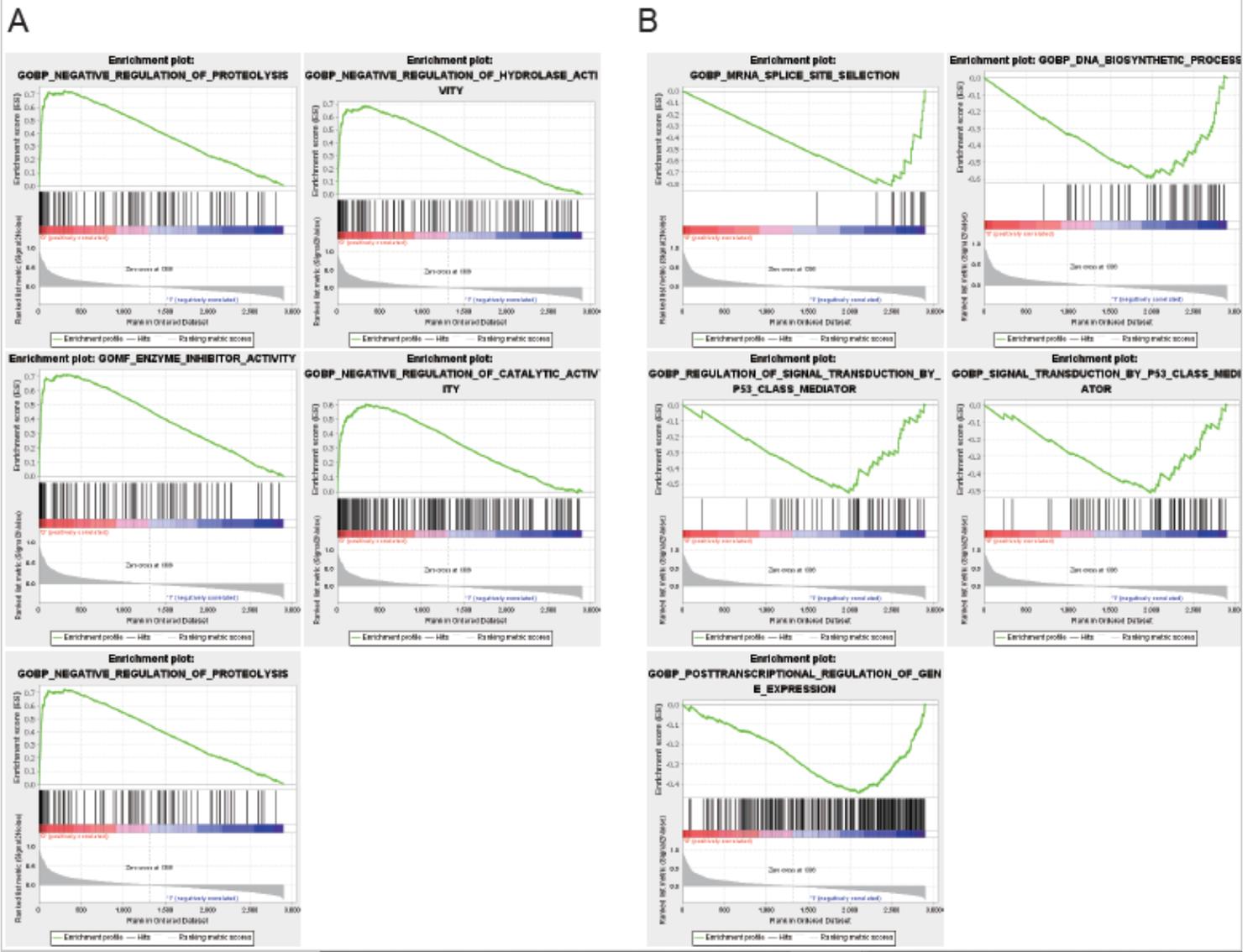
**Figure 7**

Survival analysis and verification of ALB. The survival analysis uses our proteomic data and the verification uses data from TCGA and GTEx.



**Figure 8**

Protein level of ALB in the Human Protein Atlas database. (A) Normal fallopian tube. (B) Ovarian cancer.



**Figure 9**

Gene set enrichment analysis. (A) The top 5 enriched entries of ALB high-expression group. (B) The top 5 enriched entries of ALB low-expression group.

## Supplementary Files

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

- [Additionalfile1.pdf](#)
- [Additionalfile2.xlsx](#)
- [Additionalfile3.xlsx](#)
- [Additionalfile4.pdf](#)
- [Additionalfile5.pdf](#)