

# Screening of immunosuppressive cells from colorectal adenocarcinoma and identification of prognostic markers

**Fazhan Li**

Zhengzhou University <https://orcid.org/0000-0002-8406-6124>

**Jun Zhou**

Second Xiangya Hospital

**Zedong Li**

Second Xiangya Hospital

**Leiyi Zhang** (✉ [zhangleiyi1234@163.com](mailto:zhangleiyi1234@163.com))

Second Xiangya Hospital

---

## Research

**Keywords:** COAD, Immunosuppressive cells, Th-17 cells, TCGA

**Posted Date:** April 27th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-23944/v1>

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

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Bioscience Reports on April 1st, 2021. See the published version at <https://doi.org/10.1042/BSR20203496>.

# Abstract

## Background

Colorectal cancer (CRC) is the most common type of gastrointestinal malignant tumor. Colorectal adenocarcinoma (COAD) is the most common type of colorectal cancer, and it is extremely harmful to human life and health. In recent years, the role of the immune system in the development of tumor-associated inflammation and cancer has received increasing attention.

## Results

In this study, we compiled the expression profiles of 262 patients with complete follow-up data from the Cancer Genome Atlas (TCGA) database as an experimental group and selected 65 samples from the GEO dataset (of which M0 totaled 46 samples). Validated as a verification group. First, we screened the immune cell Th-17 cells related to the prognosis of COAD disease, and then identified the Th-17 cells-related hub genes by constructing co-expression network analysis (WGCNA) and lasso regression analysis (LASSO), which determined that it may be related to Th-17 cells. Six genes associated with prognosis in patients with COAD: "KRT23 ULBP2 ASRGL1 SERPINA1 SCIN SLC28A2". Finally, we constructed a clinical prediction model and analyzed the predictive power of the model. These hub genes have been shown to be involved in the development of many diseases and are closely linked to digestive diseases. Our results suggest that the hub gene may influence the prognosis of COAD by regulating the immune infiltration of Th-17 cells.

## Conclusions

These newly discovered hub genes help to understand the mechanisms of COAD development and metastasis, thereby promoting the development of COAD and providing new therapeutic targets and biomarkers for COAD.

## Background

Colorectal cancer (CRC) is the most common type of gastrointestinal malignancy and is the third most commonly diagnosed cancer in the world and the fourth leading cause of cancer-related mortality(1). Among them, colorectal adenocarcinoma (COAD) is the most common type of colorectal cancer, with more than 1 million new cases and more than 600,000 deaths worldwide each year(2, 3). Risk factors for COAD include genetic mutations, environmental factors such as smoking and obesity, and various inflammatory diseases(4–6). So far, surgical treatment is still the most commonly used treatment. Although surgical treatment has prolonged the survival time of patients to a certain extent, the 5-year survival rate of patients who have metastasized is still less than 10%(7). Therefore, it is especially important to further determine the pathogenesis and explore new treatment options.

It has been found that lymphocyte infiltration in animal and human tumor cells has been more than a century(8). In the 1970s, Hamlin first reported the relationship between lymphocyte infiltration and tumor prognosis in breast cancer patients(9). The main prognosis and predictive immune biomarkers for lymphocyte infiltration to solid tumor tissue tumor progression. Studies have confirmed that the degree of tumor immune infiltration is positively correlated with tumor non-metastasis, and 80% of tumor-infiltrating cells are T cells(10). For example, CD8 + T cells are associated with the prognosis of colorectal tumors(11). In recent years, immunotherapy for advanced cancer has made great progress, such as malignant tumors such as melanoma and non-small cell lung cancer(12, 13). T helper 17 (Th-17 cells) is a subset of T cells differentiated from CD4 + T cells and secretes interleukin 17 (IL-17) and plays an important role in autoimmune diseases and defense responses(14). Th17 cells cause tissue damage and organ damage by secreting IL-17 pro-inflammatory cytokines(15). Th17 cells have been shown to play an important role in the pathogenesis of diseases such as gastric cancer(16, 17), ovarian cancer(18), non-small cell lung cancer(19), and breast cancer(20). Recent studies using the CRC murine model have shown that up-regulation of IL-17 signaling promotes tumor growth and progression, and Th17 cells and IL-17 play a direct role in tumor-associated inflammation and cancer development(21, 22). Interestingly, subsequent studies indicated that the presence of Th17 in the epithelium was significantly associated with patient survival, suggesting that Th 17 cells have a dual role in CRC(23). This result caught our attention and we verified and further explored it in this experiment.

In this study, we used data from the TCGA and GEO datasets to identify Th 17 cell-associated genes and to identify possible Th-17 cells by constructing co-expression network analysis (WGCNA) and lasso regression analysis (LASSO). WGCNA is a method to analyze gene expression patterns in multiple samples. Genes with similar expression patterns would be clustered, and correlation between modules and specific traits or phenotypes could be calculated. In this study, WGCNA-based methods were used to identify gene modules correlated with the development of colon cancer, and the most representative genes were identified as hub genes to help in screening colon cancer. A risk scoring model is constructed based on the finally determined hub gene to evaluate the prognosis of colorectal adenocarcinoma. The close relationship between these hub genes and Th-17 cells can provide new ideas for the study of COAD.

## Patients And Methods

### 1.Data collection

The expression profiles and clinical data of patients with COAD were obtained from the Cancer Genome Atlas (TCGA) cohort and the Gene Expression Integrated Library (GEO). This study included the expression profiles of 262 patients with complete follow-up data in the TCGA database (Histological type: colorectal adenocarcinoma, patients with M = 0 in TNM stage which means there is no distant metastasis (the tumor has not spread to other parts of the body), and the GEO dataset. 65 samples were taken (of which M0 totaled 46 samples). The TCGA biolinks package and the GEO query package are used to download TCGA and GEO data(24, 25). All data and sample information in this study are from the database and do not require review by the ethics committee.

## 2. Research method

### Flow chart Fig.1

### 3. Infiltration of Immune Cells

ssGSEA (single-sample gene set enrichment analysis) was performed on TCGA and GEO gene expression profile data to quantitatively analyze the infiltration of immune cells in tumor tissues, and 24 immune cells were infiltrated. ssGSEA, an enrichment score that calculates the extent to which genes in a particular gene set in a single sample are coordinated up- or down-regulated. ssGSEA ranks genes by their absolute expression in the sample and calculates enrichment scores by integrating the differences between empirically cumulative distribution functions at the gene level(26, 27).

### 4. Univariate Cox regression analysis and Kaplan - Meire's curve

We screened for immune cells that may affect COAD by single factor COX analysis in the TCGA and GEO data sets ( $p < 0.05$ ). The "survminer" package was then used for optimal separation statistics, which in turn divided gene expression into high and low groups and plotted Kaplan-Meier curves.

### 5. Construction of co-expression network and identification of related modules

The co-expression network is constructed from the "WGCNA" package in R(28). We first select genes with variances greater than all variance quartiles in the TCGA data set(Top quarter, because those genes with larger variances mean greater variation in different samples), identify the expression data for the selected genes, and then cluster the samples to detect outliers. Finally, the gene clustering module was determined based on clinical features and TOM-based differences(29), and the correlation between module characteristic genes (MEs) and clinical features was calculated to identify highly relevant gene clustering modules.

### 6. Hub gene determination and risk scoring model construction

In this study, we selected a module that was highly correlated with disease characteristics(screening based on residual tumor, Pathological stage, OS, OS events, and Th17 cells, where positive values indicate positive correlation, negative values indicate negative correlation, and absolute value of P value indicates the magnitude of correlation), and then we used a single factor COX analysis to screen for genes that were significantly associated with prognosis in the module ( $p$  value  $< 0.05$  was included in the study), and finally LASSO (Least Absolute Shrinkage and Selector Operation) The screening of COAD prognosis-related genes, based on lambda.min (lambda corresponding to the minimum mean error), was selected for the hub gene (LASSO was analyzed using the "glmnet" package in R). Central gene expression values weighted by LASSO regression coefficients yielded a risk score for each patient. The Survminer R software package finds the best cut-off value for this risk score. The Survminer R software package finds the risk score. The ROC and Kaplan-Meier curve are used to assess the prognostic ability of the risk score.

## Results

### 1. Quantify immune cell infiltration and analyze the relationship between immune cells and tumor recurrence

First, we used ssGSEA to quantify mRNA data from 24 immune cell infiltrations in TCGA and GEO samples, including: aDC, B cells, CD8 T cells, cytotoxic cells, DCs, eosinophils, iDCs, macrophages, Mast cells, neutrophils, NK CD56 bright cells, NK CD56dim cells, NK cells, pDC, T cells, T helper cells, Tcm, Tem, TFH, TFH, Tgd, Th1 cells, Th17 cells, Th2 cells, Treg. Immunocytes associated with COAD tumors were then screened for COX univariate analysis. We found that Th-17 cells had similar results in TCGA and GEO samples ( $p < 0.05$  in the TCGA and GEO data sets, and  $HR < 1$ .  $HR > 1$  is a risk factor for recurrence, and  $HR < 1$  is a recurrence protection factor.) (Fig. 2A), this indicates that Th-17 cells play a protective role in COAD.

The "survival" and 'survminer' packages were used to determine the relationship between Th-17 cells and COAD prognosis and a Kaplan-Meier curve was plotted (Fig. 2B). We found that the higher the degree of infiltration of Th-17 cells, the better the prognosis of COAD.

### 2. Construction of co-expression network and identification of related modules

In this study, we calculated the first quarter of the variance in the expression of each gene in all samples to construct a co-expression network with a total of 4,859 genes. Then, a hierarchical clustering tree was constructed for 4859 genes out of 262 samples, and 12 outlier samples were eliminated (Fig. 3). A hierarchical clustering tree is constructed for the remaining 250 samples. We choose the weighting coefficient  $\beta$  value to take 4 to construct the co-expression network, while moderately retaining the average connectivity of each gene node (Node) (Fig. 4). The similar modules were then combined by a hybrid dynamic shear tree method to identify 20 modules and color representation of the phenotype in each sample: white for low, red for high, and gray for missing (Fig. 5).

### 3. Correlation between modules and cancer

We selected a significant correlation module by comparing the clinical information of each module with COAD patients. Among them, the black module ( $cor = 0.36$ ,  $p = 4 \times 10^{-9}$ ) has the highest correlation with Th-17 cells (positive values indicate positive correlation and negative values indicate negative correlation), indicating that Th-17 cells may affect the prognosis of COAD patients by regulating the genes of the black module (Fig. 6).

### 4. hub genes identification

The black module has a total of 207 genes, of which 160 genes are identical to the GEO data set. To further identify the genes associated with COAD, we performed a COX univariate analysis on the same

160 genes in both data sets, setting a p-value of  $< 0.05$ . Eight prognostic related genes were screened. Then, LASSO regression analysis was used to screen 8 prognosis-related genes in the black module (the best lambda value was 0.009100874, and the more concise model lambda value in the SE was 0.06420486) (Fig. 7). It was finally determined that the six genes of “KRT23 ULBP2 ASRGL1 SERPINA1 SCIN SLC28A2” were related to the prognosis of patients. We used LASSO regression analysis and single factor COX analysis to screen for prognostic related genes. Based on P value  $< 0.05$ , the identified hub genes can be considered to be related to prognosis.

## 5. Risk score

The Survminer R package finds the best threshold for risk scores. Survminer R package to find the optimal cut-off for the risk score, optimal cut-off for the risk scores were found by Survminer R package, ROC and Kaplan–Meier curve were used to assess the prognostic capacity of the risk scores. We first divided the samples in TCGA and GEO into three groups based on the infiltration of Th-17 cells, and compared the high expression combined low expression group, and plotted the risk score distribution (Fig. 8), time-dependent ROC curve (Fig. 9) and survival analysis of the TCGA and GEO data sets (Fig. 10). The area under the ROC curve (AUC) of the OS prognostic model were: TCGA (12 months AUC: 0.747, 36 months AUC: 0.750); GEO (12 months AUC: 0.868, 36 months AUC: 0.706).

## Discussion

In this study, we first evaluated the amount of Th-17 cells in COAD patients. We found that Th-17 cells was significantly reduced in COAD patients, which means that Th-17 cells is closely related to COAD. In fact, Th-17 cells and its IL-17 have long been found to play an important role in the development of many types of malignancies (19, 20, 30–35). IL-17 promotes tumor growth through a variety of mechanisms, including by inhibiting specific immune cell infiltration in hepatocellular carcinoma and melanoma (36, 37), IL-17 can be remodeled by extracellular matrix To regulate the surrounding microenvironment and promote the invasion and metastasis of cancer cells (38–41). Recent mouse models of colorectal cancer (CRC) have shown that Th17 cells and IL-17 have a direct role in tumor-associated inflammation and cancer development (22).

However, the dual role of Th 17 cells in CRC should be taken seriously. The conclusions of this experiment further validate the protective role of Th 17 cells in the prognosis of CRC patients and identify the relevant hub genes. This result contributes to the study of immunologically relevant targeted therapies for COAD.

Although a large amount of data confirms that IL-17 plays an important role in the mechanism of tumorigenesis, there are relatively few studies to further determine its related genes. In this study, we finalized the relevant gene “KRT23 ULBP2 ASRGL1 SERPINA1 SCIN SLC28A2” that may affect Th-17 cells immune infiltration in COAD patients. We have now found that these genes are involved in the development of many diseases (42–49), In non-gastrointestinal tumors, ASRGL1 has shown a high correlation in the prognosis of patients with locally advanced lymph node-negative prostate cancer and

cervical cancer(42, 50); SLC28A2 helps increase the role of ribavirin (RBV) in the treatment of viral hepatitis Hyperuricemia (HUA) has been implicated(48, 51). KRT23 was identified as a specifically expressed gene in colorectal cancer and highly expressed in colorectal cancer(45). Comprehensive analysis of the lncRNA-miRNA-mRNA network also further confirmed that ULBP2 is associated with the prognosis of colorectal cancer(52). Serpina1 in the serine protease inhibitor protein family (SERPIN) as a biomarker in CRC can increase the predictive ability of CRC diagnosis(53), and data show that Snail and serpinA1 promote CRC progression through fibronectin. This result suggests that serpinA1 is a potential therapeutic target for novel prognostic biomarkers and CRC. Scinderin (SCIN) may be an important predictor of poor prognosis in patients with colorectal cancer liver metastasis (CRLM) and CRC(54) Our research provides new insights into these genes that are highly expressed in CRC, and this finding may help explain the dual role of Th17cells.

The results of this study further confirmed the role of these six genes in COAD. The results of this study further confirmed the protective effects of these six genes in COAD, and the immune infiltration regulation of these genes on Th 17 cells may be the cause of their dual effects. It also means “KRT23 ULBP2 ASRGL1 SERPINA1 SCIN SLC28A2”Probably as a potential target and biomarker for treatment. Although we used clinical samples and verified them with advanced analysis methods and multiple databases, some limitations should be mentioned in our research. On one hand, the expression levels of the genes we identified have not been determined through further experiments. In addition, the number of samples also has certain limitations. On the other hand, the relationship between the hub gene and Th17 is based on target prediction only. Further molecular biology experiments are needed to prove the interaction.

## Conclusion

In conclusion, we compiled the expression profiles of 262 patients with complete follow-up data from TCGA database as an experimental group and selected 65 samples from the GEO dataset. Then we identified the Th-17 cells-related hub genes by constructing co-expression network analysis (WGCNA) and lasso regression analysis (LASSO), which determined that it may be related to Th-17 cells. our results indicate that the gene associated with Th-17 cells “KRT23 ULBP2 ASRGL1 SERPINA1 SCIN SLC28A2” may be significant for human COAD. The difference in Th-17 cells between the M0 sample and the metastatic sample may be related to the above six genes. This result may provide a new understanding of the mechanism of tumor seeding and metastasis in COAD, and needs further study.

## Declarations

### Authors' contributions

LZD and ZJ contributed to the conception of the study. LZD and LFZ performed the data analyses. LZD contributed significantly to process data. LFZ wrote the manuscript. All of the authors read and approved the final manuscript.

## Ethical Statement

The data of this study are from TCGA, GEO, and do not involve animal experiments and human specimens, no ethics-related issues.

## Consent for publication

Not Applicable .

## Availability of data and material

The data of this study are from TCGA, GEO.

## Competing interests

The authors declare that they have no conflicts of interest.

## Funding

No funding support.

## Acknowledgments

We thank these researchers who gave their data for this analysis. It is cheerful to acknowledge their contributions.

## References

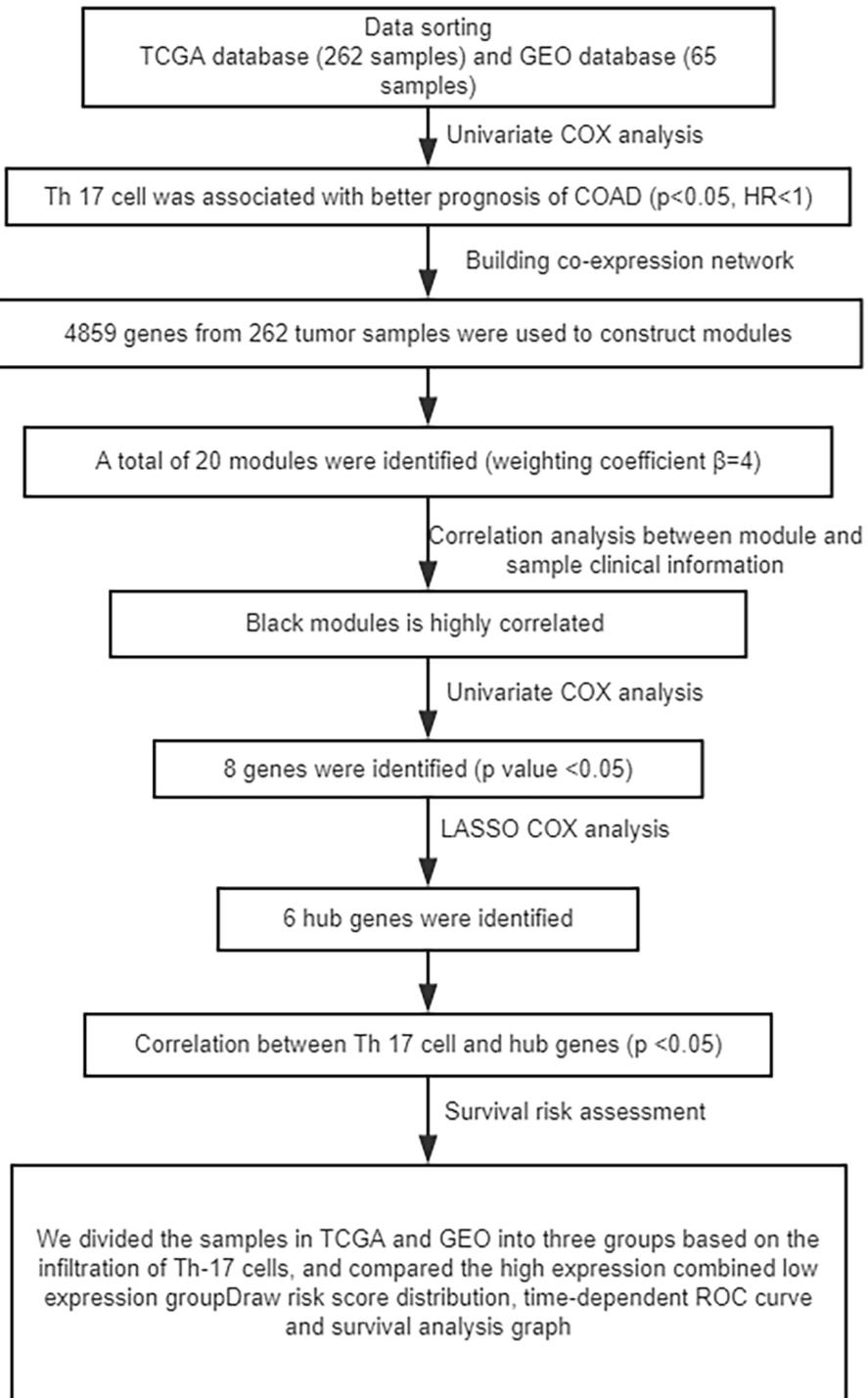
1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin.* 2018;68(1):7-30.
2. Zeng JH, Liang L, He RQ, Tang RX, Cai XY, Chen JQ, et al. Comprehensive investigation of a novel differentially expressed lncRNA expression profile signature to assess the survival of patients with colorectal adenocarcinoma. *Oncotarget.* 2017;8(10):16811-28.
3. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. *CA Cancer J Clin.* 2009;59(4):225-49.
4. Kang M, Edmundson P, Araujo-Perez F, McCoy AN, Galanko J, Keku TO. Association of plasma endotoxin, inflammatory cytokines and risk of colorectal adenomas. *BMC Cancer.* 2013;13:91.
5. Olsen HW, Lawrence WA, Snook CW, Mutch WM. Risk factors and screening techniques in 500 patients with benign and malignant colon polyps. An urban community experience. *Dis Colon Rectum.* 1988;31(3):216-21.
6. Terzic J, Grivennikov S, Karin E, Karin M. Inflammation and colon cancer. *Gastroenterology.* 2010;138(6):2101-14 e5.
7. Siegel R, DeSantis C, Virgo K, Stein K, Mariotto A, Smith T, et al. Cancer treatment and survivorship statistics, 2012. *CA Cancer J Clin.* 2012;62(4):220-41.

8. Maccarty WC. Longevity in Cancer: A Study of 293 Cases. *Ann Surg.* 1922;76(1):9-12.
9. Hamlin IM. Possible host resistance in carcinoma of the breast: a histological study. *Br J Cancer.* 1968;22(3):383-401.
10. Husby G, Hoagland PM, Strickland RG, Williams RC, Jr. Tissue T and B cell infiltration of primary and metastatic cancer. *J Clin Invest.* 1976;57(6):1471-82.
11. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science.* 2006;313(5795):1960-4.
12. Robert C, Thomas L, Bondarenko I, O'Day S, Weber J, Garbe C, et al. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N Engl J Med.* 2011;364(26):2517-26.
13. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med.* 2015;373(17):1627-39.
14. Melnik BC, John SM, Chen W, Plewig G. T helper 17 cell/regulatory T-cell imbalance in hidradenitis suppurativa/acne inversa: the link to hair follicle dissection, obesity, smoking and autoimmune comorbidities. *Br J Dermatol.* 2018;179(2):260-72.
15. Nalbandian A, Crispin JC, Tsokos GC. Interleukin-17 and systemic lupus erythematosus: current concepts. *Clin Exp Immunol.* 2009;157(2):209-15.
16. Liu X, Jin H, Zhang G, Lin X, Chen C, Sun J, et al. Intratumor IL-17-positive mast cells are the major source of the IL-17 that is predictive of survival in gastric cancer patients. *PLoS One.* 2014;9(9):e106834.
17. Iida T, Iwahashi M, Katsuda M, Ishida K, Nakamori M, Nakamura M, et al. Tumor-infiltrating CD4+ Th17 cells produce IL-17 in tumor microenvironment and promote tumor progression in human gastric cancer. *Oncol Rep.* 2011;25(5):1271-7.
18. Kato T, Furumoto H, Ogura T, Onishi Y, Irahara M, Yamano S, et al. Expression of IL-17 mRNA in ovarian cancer. *Biochem Biophys Res Commun.* 2001;282(3):735-8.
19. Chen X, Wan J, Liu J, Xie W, Diao X, Xu J, et al. Increased IL-17-producing cells correlate with poor survival and lymphangiogenesis in NSCLC patients. *Lung Cancer.* 2010;69(3):348-54.
20. Zhu X, Mulcahy LA, Mohammed RA, Lee AH, Franks HA, Kilpatrick L, et al. IL-17 expression by breast-cancer-associated macrophages: IL-17 promotes invasiveness of breast cancer cell lines. *Breast Cancer Res.* 2008;10(6):R95.
21. Grivennikov SI, Wang K, Mucida D, Stewart CA, Schnabl B, Jauch D, et al. Adenoma-linked barrier defects and microbial products drive IL-23/IL-17-mediated tumour growth. *Nature.* 2012;491(7423):254-8.
22. Wang K, Kim MK, Di Caro G, Wong J, Shalapour S, Wan J, et al. Interleukin-17 receptor signaling in transformed enterocytes promotes early colorectal tumorigenesis. *Immunity.* 2014;41(6):1052-63.
23. Amicarella F, Muraro MG, Hirt C, Cremonesi E, Padovan E, Mele V, et al. Dual role of tumour-infiltrating T helper 17 cells in human colorectal cancer. *Gut.* 2017;66(4):692-704.

24. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 2016;44(8):e71.
25. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics.* 2007;23(14):1846-7.
26. Finotello F, Trajanoski Z. Quantifying tumor-infiltrating immune cells from transcriptomics data. *Cancer Immunol Immunother.* 2018;67(7):1031-40.
27. 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(4):782-95.
28. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
29. Botia JA, Vandrovcova J, Forabosco P, Guelfi S, D'Sa K, United Kingdom Brain Expression C, et al. An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *BMC Syst Biol.* 2017;11(1):47.
30. Martin-Orozco N, Dong C. The IL-17/IL-23 axis of inflammation in cancer: friend or foe? *Curr Opin Investig Drugs.* 2009;10(6):543-9.
31. Su X, Ye J, Hsueh EC, Zhang Y, Hoft DF, Peng G. Tumor microenvironments direct the recruitment and expansion of human Th17 cells. *J Immunol.* 2010;184(3):1630-41.
32. Miyahara Y, Odunsi K, Chen W, Peng G, Matsuzaki J, Wang RF. Generation and regulation of human CD4+ IL-17-producing T cells in ovarian cancer. *Proc Natl Acad Sci U S A.* 2008;105(40):15505-10.
33. Hu J, Mao Y, Li M, Lu Y. The profile of Th17 subset in glioma. *Int Immunopharmacol.* 2011;11(9):1173-9.
34. Yamada Y, Saito H, Ikeguchi M. Prevalence and clinical relevance of Th17 cells in patients with gastric cancer. *J Surg Res.* 2012;178(2):685-91.
35. Zhuang Y, Peng LS, Zhao YL, Shi Y, Mao XH, Chen W, et al. CD8(+) T cells that produce interleukin-17 regulate myeloid-derived suppressor cells and are associated with survival time of patients with gastric cancer. *Gastroenterology.* 2012;143(4):951-62 e8.
36. Ma S, Cheng Q, Cai Y, Gong H, Wu Y, Yu X, et al. IL-17A produced by gammadelta T cells promotes tumor growth in hepatocellular carcinoma. *Cancer Res.* 2014;74(7):1969-82.
37. Wang L, Yi T, Kortylewski M, Pardoll DM, Zeng D, Yu H. IL-17 can promote tumor growth through an IL-6-Stat3 signaling pathway. *J Exp Med.* 2009;206(7):1457-64.
38. Bar-Eli M. Role of interleukin-8 in tumor growth and metastasis of human melanoma. *Pathobiology.* 1999;67(1):12-8.
39. Inozume T, Hanada K, Wang QJ, Yang JC. IL-17 secreted by tumor reactive T cells induces IL-8 release by human renal cancer cells. *J Immunother.* 2009;32(2):109-17.
40. Rosette C, Roth RB, Oeth P, Braun A, Kammerer S, Ekblom J, et al. Role of ICAM1 in invasion of human breast cancer cells. *Carcinogenesis.* 2005;26(5):943-50.

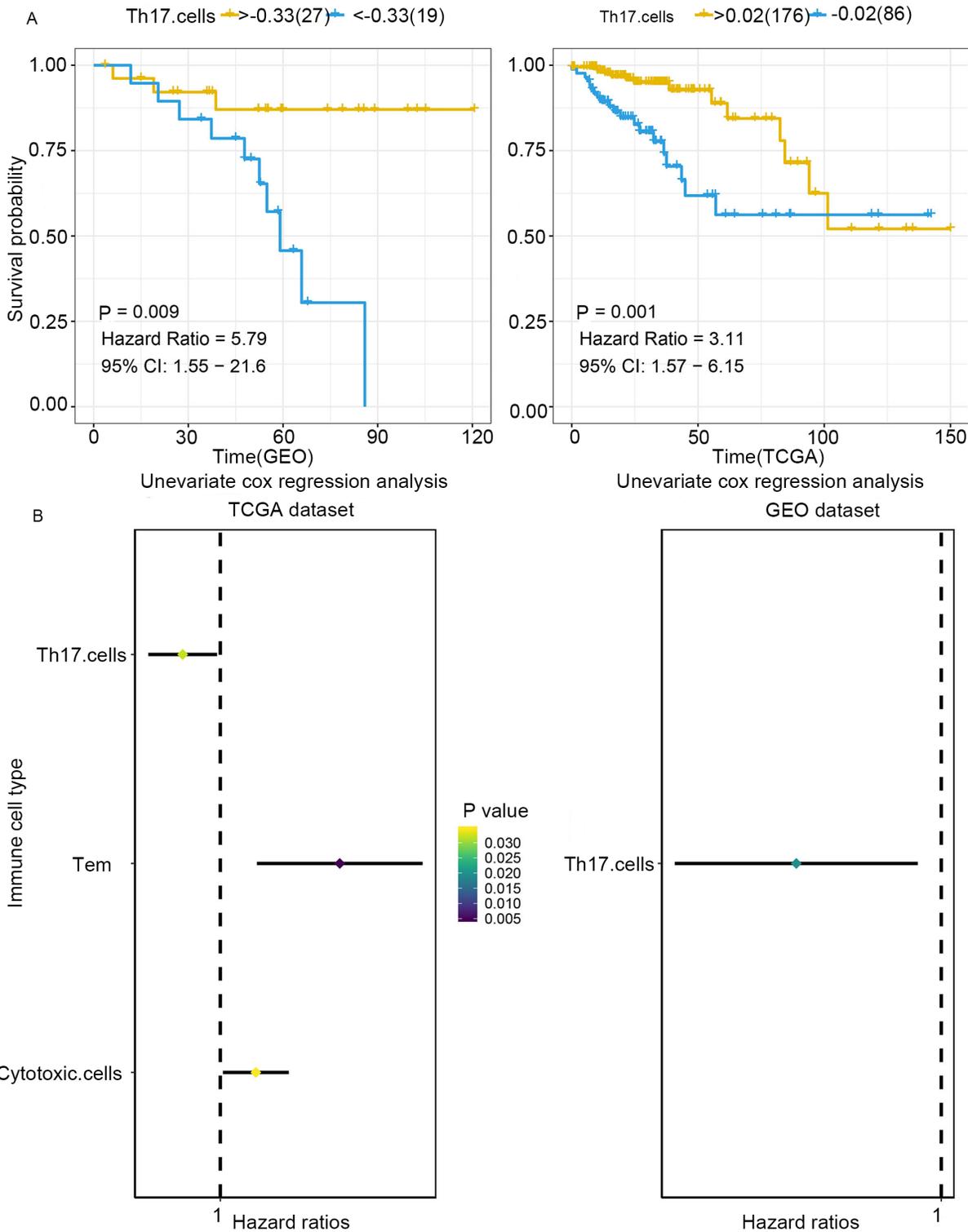
41. Roland CL, Harken AH, Sarr MG, Barnett CC, Jr. ICAM-1 expression determines malignant potential of cancer. *Surgery*. 2007;141(6):705-7.
42. Pudova EA, Lukyanova EN, Nyushko KM, Mikhaylenko DS, Zaretsky AR, Snezhkina AV, et al. Differentially Expressed Genes Associated With Prognosis in Locally Advanced Lymph Node-Negative Prostate Cancer. *Front Genet*. 2019;10:730.
43. Peng P, Wu W, Zhao J, Song S, Wang X, Jia D, et al. Decreased expression of Calpain-9 predicts unfavorable prognosis in patients with gastric cancer. *Sci Rep*. 2016;6:29604.
44. Anderson KJ, Cormier RT, Scott PM. Role of ion channels in gastrointestinal cancer. *World J Gastroenterol*. 2019;25(38):5732-72.
45. Flebbe H, Hamdan FH, Kari V, Kitz J, Gaedcke J, Ghadimi BM, et al. Epigenome Mapping Identifies Tumor-Specific Gene Expression in Primary Rectal Cancer. *Cancers (Basel)*. 2019;11(8).
46. Zhang ZH, Lian XY, Li XX, He PF, Lin J, Qian J. [Clinical Study of SCIN Expression and Dromoter Methylation in Patients with Chronic Myeloid Leukemia]. *Zhongguo Shi Yan Xue Ye Xue Za Zhi*. 2019;27(3):646-51.
47. Basyte-Bacevice V, Skieceviciene J, Valantiene I, Sumskiene J, Petrenkiene V, Kondrackiene J, et al. SERPINA1 and HSD17B13 Gene Variants in Patients with Liver Fibrosis and Cirrhosis. *J Gastrointestin Liver Dis*. 2019;28(3):297-302.
48. Zhou Z, Li Z, Wang C, Li X, Cheng X, Li C, et al. Common variants in the SLC28A2 gene are associated with serum uric acid level and hyperuricemia and gout in Han Chinese. *Hereditas*. 2019;156:4.
49. Weng J, Han X, Liu K, Yang J, Wei S, Zhang Y, et al. CD44 3'-Untranslated Region Functions as a Competing Endogenous RNA to Enhance NK Sensitivity of Liver Cancer Stem Cell by Regulating ULBP2 Expression. *Int J Biol Sci*. 2019;15(8):1664-75.
50. Xu J, Liu H, Yang Y, Wang X, Liu P, Li Y, et al. Genome-Wide Profiling of Cervical RNA-Binding Proteins Identifies Human Papillomavirus Regulation of RNASEH2A Expression by Viral E7 and E2F1. *mBio*. 2019;10(1).
51. Abdelkawy KS, El-Haggag SM, Ziada DH, Ebaid NF, El-Magd MA, Elbarbry FA. The effect of genetic variations on ribavirin pharmacokinetics and treatment response in HCV-4 Egyptian patients receiving sofosbuvir/daclatasvir and ribavirin. *Biomed Pharmacother*. 2020;121:109657.
52. Gao Z, Fu P, Yu Z, Zhen F, Gu Y. Comprehensive Analysis of lncRNA-miRNA- mRNA Network Ascertains Prognostic Factors in Patients with Colon Cancer. *Technol Cancer Res Treat*. 2019;18:1533033819853237.
53. Peltier J, Roperch JP, Audebert S, Borg JP, Camoin L. Quantitative proteomic analysis exploring progression of colorectal cancer: Modulation of the serpin family. *J Proteomics*. 2016;148:139-48.
54. Lin Q, Li J, Zhu D, Niu Z, Pan X, Xu P, et al. Aberrant Scinderin Expression Correlates With Liver Metastasis and Poor Prognosis in Colorectal Cancer. *Front Pharmacol*. 2019;10:1183.

## Figures



**Figure 1**

Flow chart.

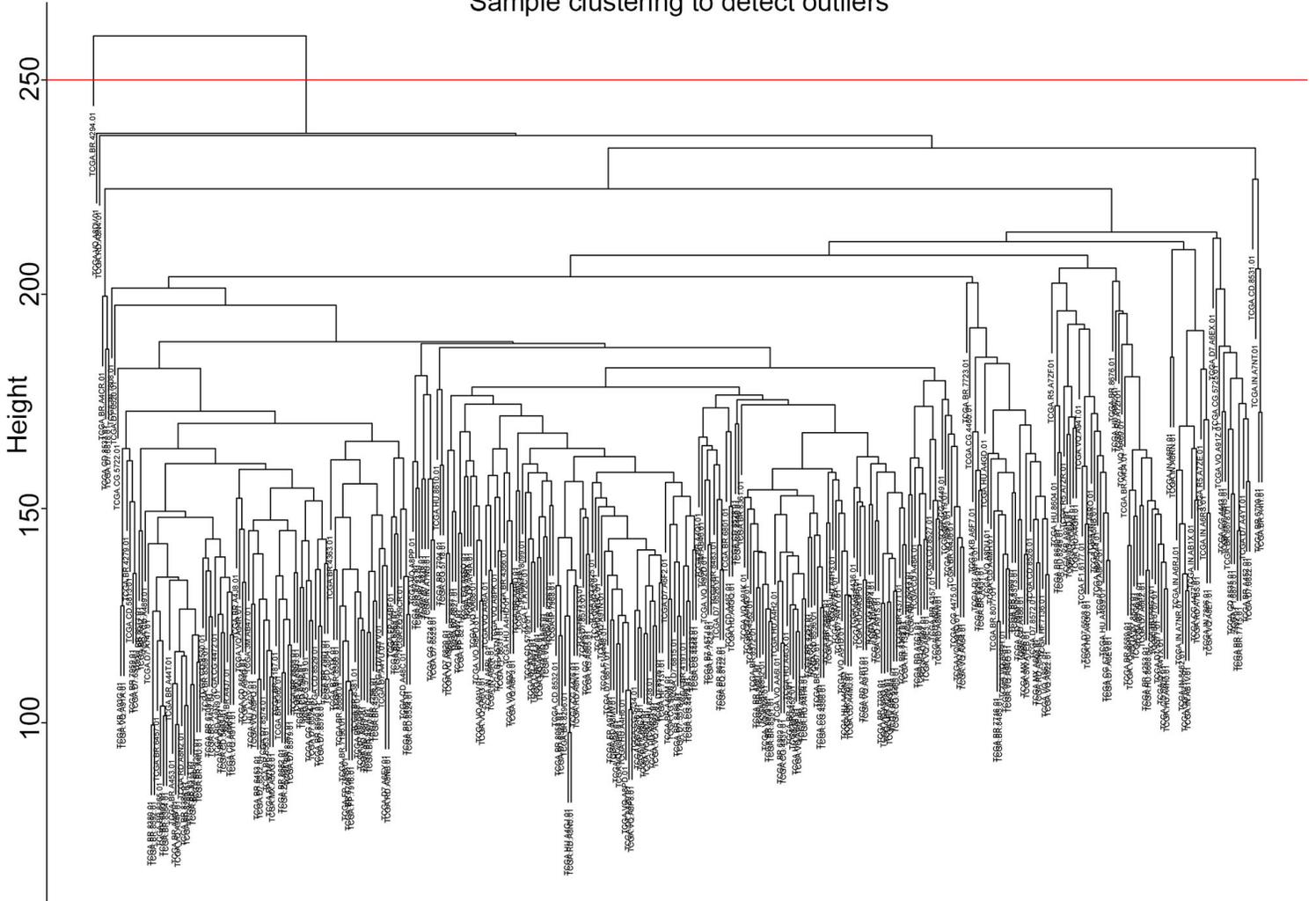


**Figure 2**

Figure 2A: Forrest plot of univariate Cox regression analysis in COAD (Th-17 cells are associated with prognosis and HR < 1, both as protective factors in both TCGA and GEO). Figure 2B: Kaplan-Meier curves of Th-17 cells in immune cells. Genetic analysis of the TCGA and GEO databases showed that the high-invasion group had a significant inhibitory effect on the prognosis of COAD, and Th-17 cells were

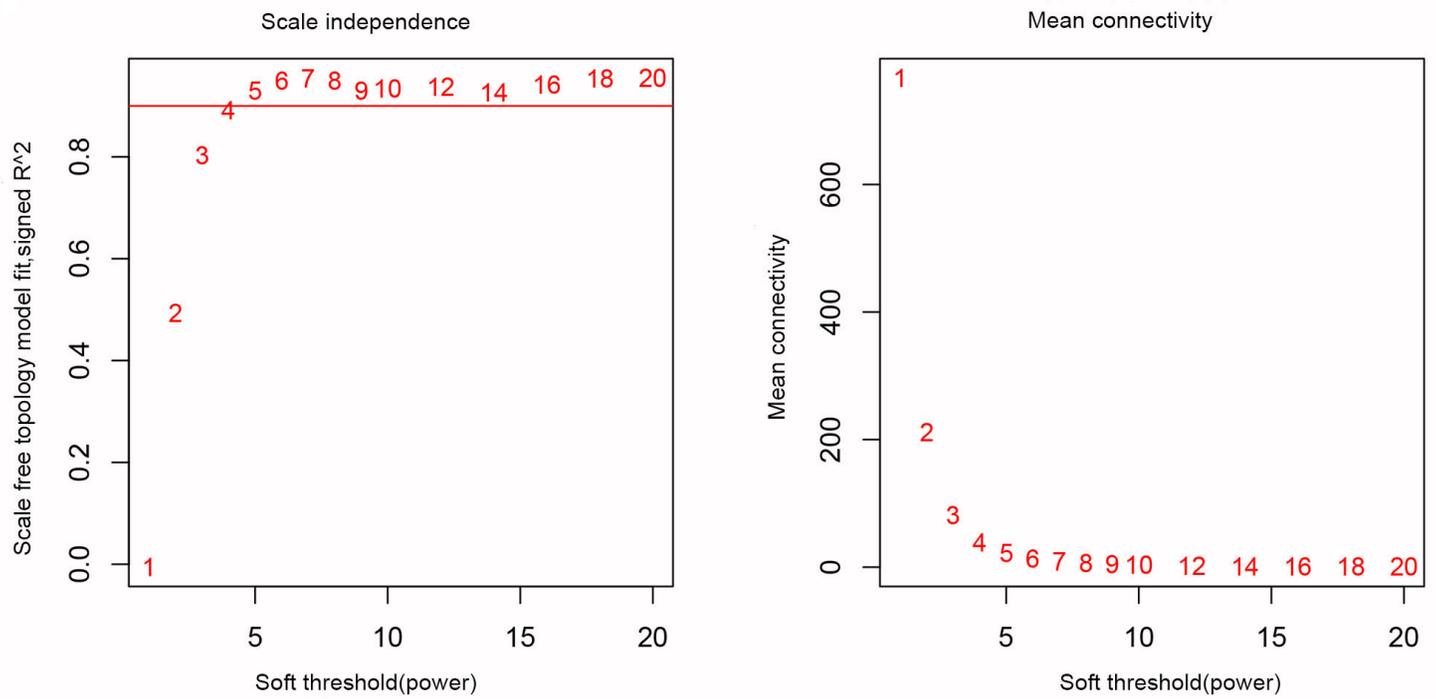
associated with a better prognosis (GEO sample:  $p = 0.009$ , Hazard Ratio = 5.79, 95% CI: 1.55 – 21.6) In the TCGA sample:  $p = 0.001$ , Hazard ratio = 3.11, 95% CI: 1.57 – 6.15).

### Sample clustering to detect outliers



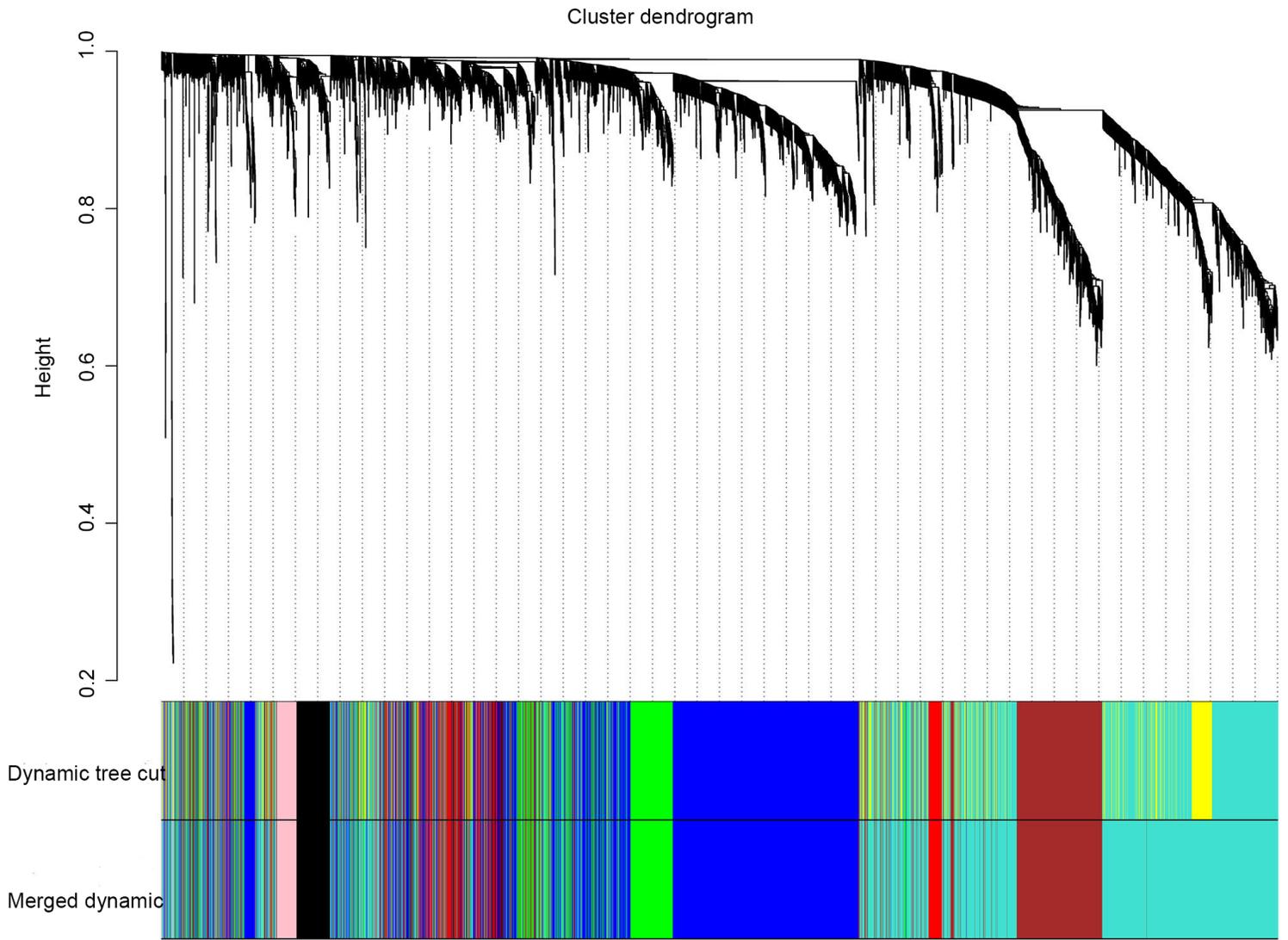
**Figure 3**

Systematic clustering of 262 COAD tumor samples and clinical information.



**Figure 4**

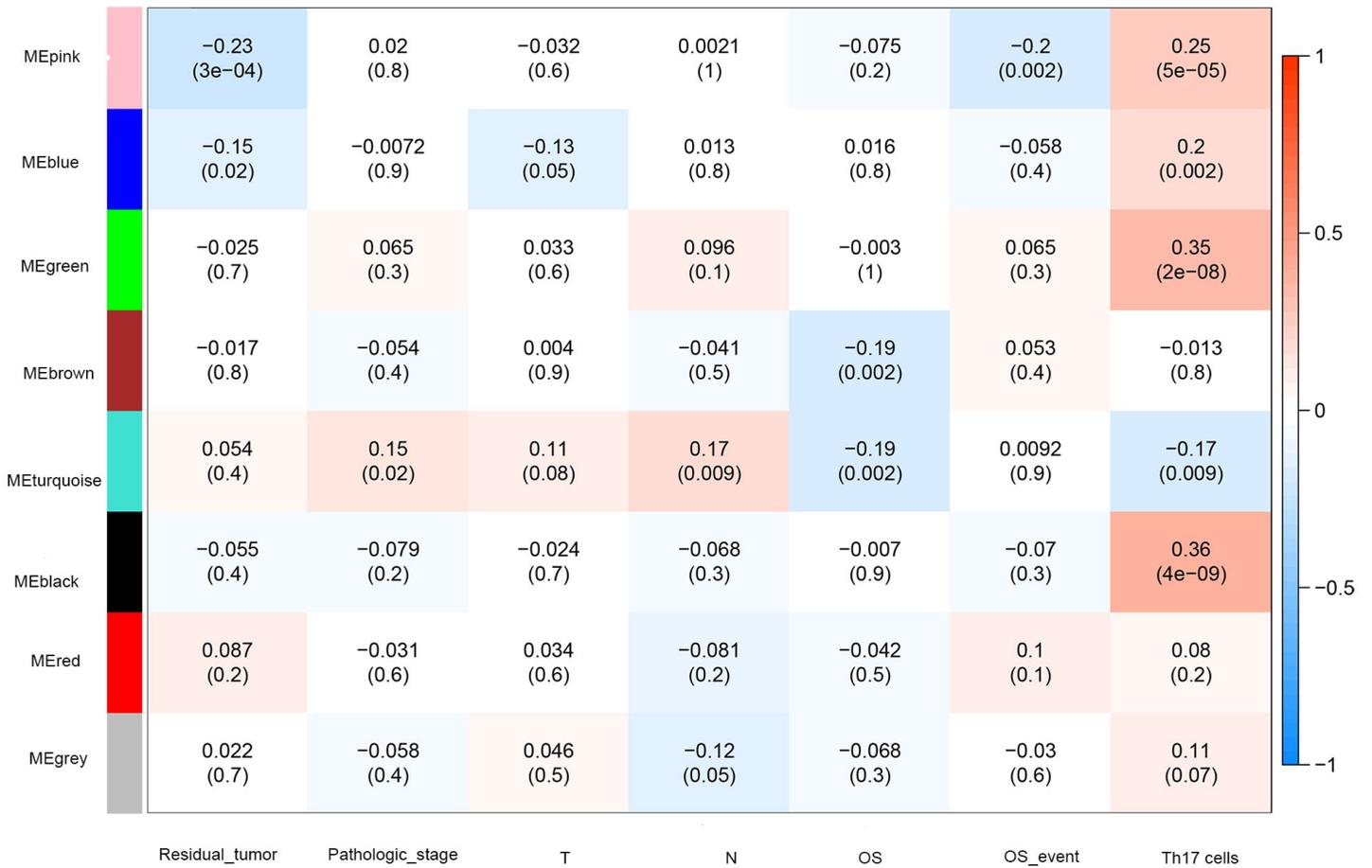
Scale-free conformance index and average connectivity calculated at different beta values (the numbers in the figure represent the corresponding soft threshold power. An approximate scale-free topology can be achieved at a soft threshold power of 4).



**Figure 5**

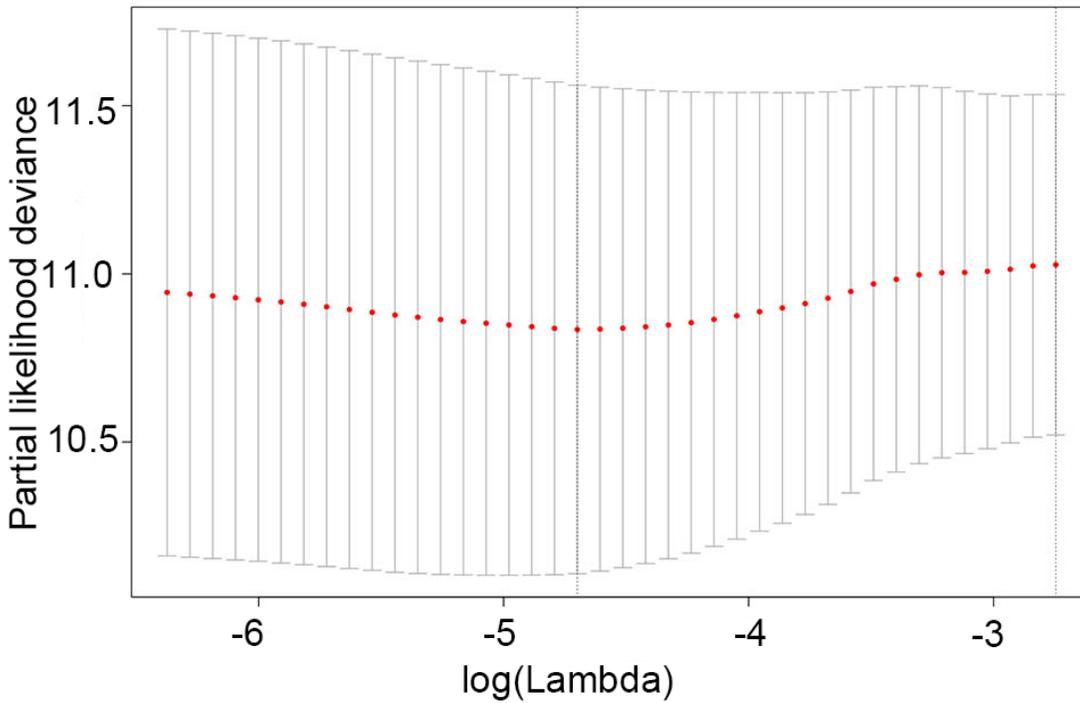
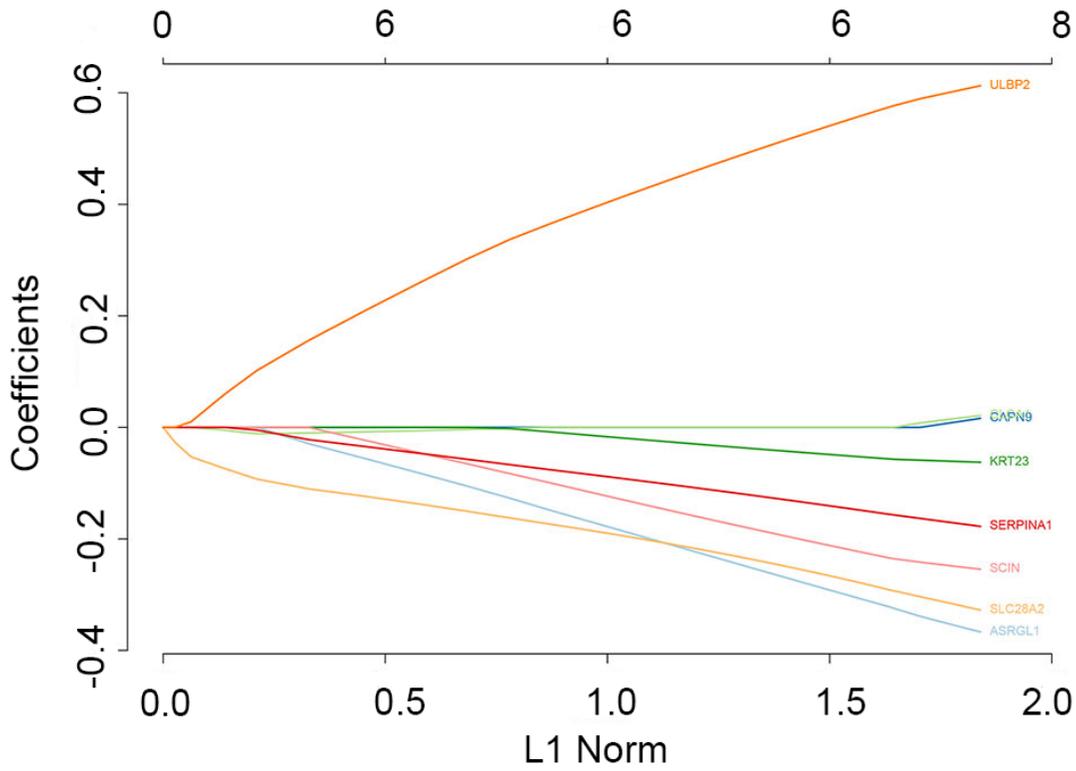
Gene cluster tree diagram. Based on consensus topological overlap, each color module represents a color-coding module containing a set of highly connected genes, each module containing at least 50 genes (larger modules are relatively more meaningful).

Module-trait relationships



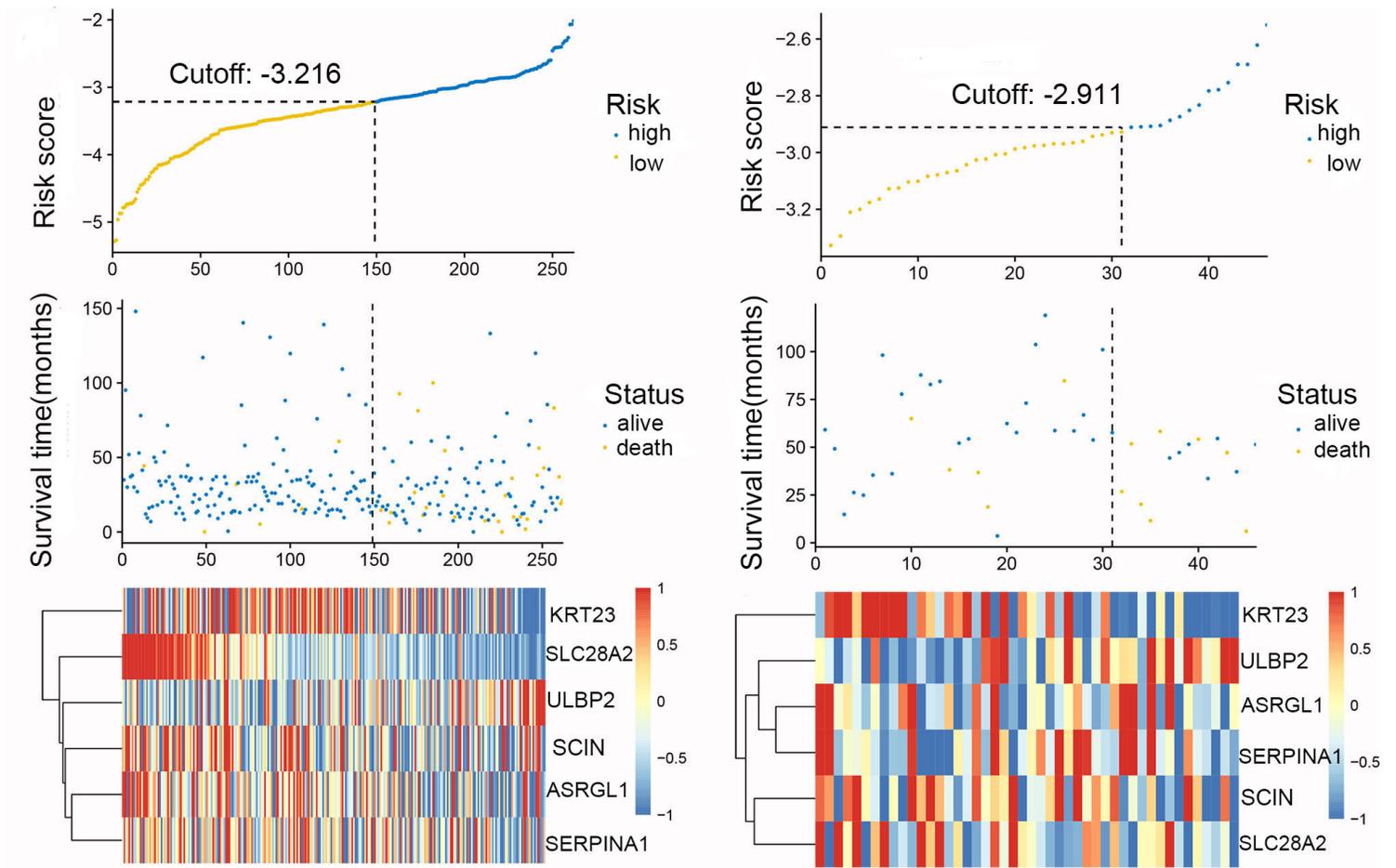
**Figure 6**

Heat map of different modules with different clinical features. Each row corresponds to a consistency module, and each column corresponds to one type of clinical information—Screening based on residual tumor, pathological stage, OS, OS events, and Th17 cells. The absolute value of the P value indicates the correlation size (the module name is displayed on the left side of each cell, and the associated intensity and direction are shown on the right side of the heat map).



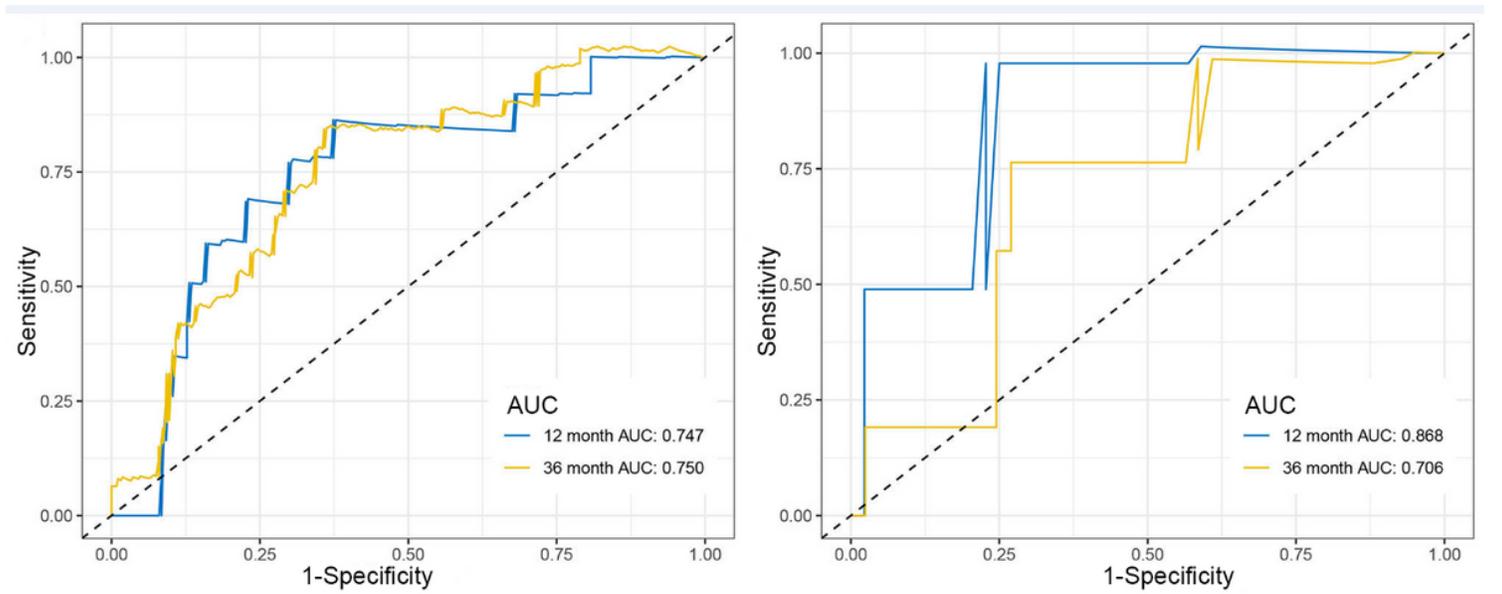
**Figure 7**

Figure A: Distribution of LASSO coefficients for eight related genes. Figure B: Partial likelihood bias of the LASSO coefficient distribution. The vertical dashed line indicates the minimum partial likelihood deviation.



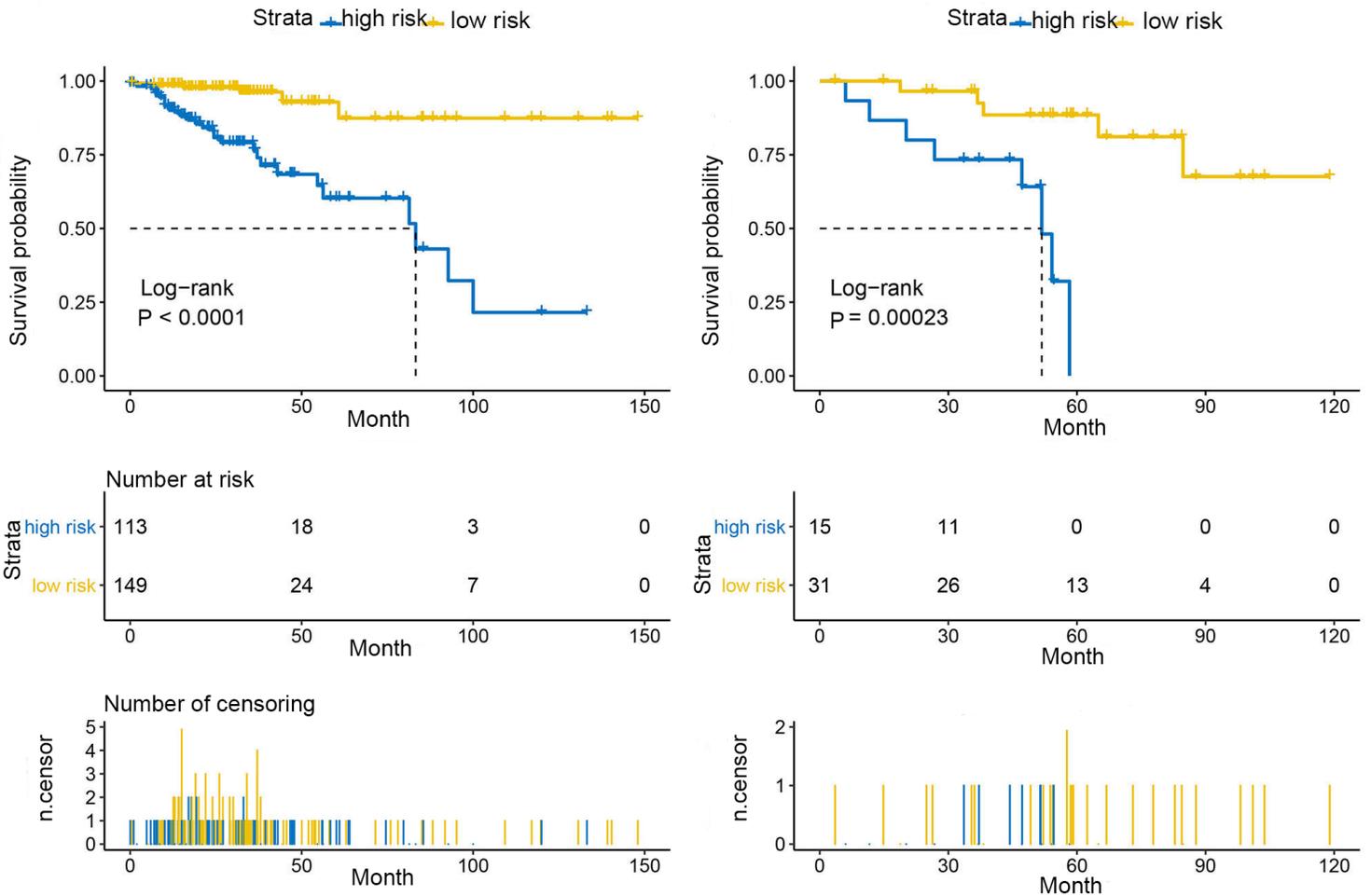
**Figure 8**

Relationship between risk scores and the expression levels of the six hub genes.



**Figure 9**

Time-dependent ROC curve analysis for clinical prediction models.



**Figure 10**

Relationship between high- and low-risk scores and OS. The OS of the low-risk score group was significantly higher than the high-risk score group.