

SPOCK1 and POSTN are Valuable Prognostic Biomarkers and Correlate With Tumor Immune Infiltrates in Colorectal Cancer

Mengting Li

Zhongnan Hospital of Wuhan University

Wenjie Li

Zhongnan Hospital of Wuhan University

Haizhou Wang

Zhongnan Hospital of Wuhan University

Yanan Peng

Zhongnan Hospital of Wuhan University

Qian Hu

Zhongnan Hospital of Wuhan University

Wanhui Wei

Zhongnan Hospital of Wuhan University

Caiqin Gan

Zhongnan Hospital of Wuhan University

Fan Wang

Zhongnan Hospital of Wuhan University

Lan Liu

Zhongnan Hospital of Wuhan University

Qiu Zhao (✉ qiuzhao@whu.edu.cn)

Wuhan University Zhongnan Hospital

Research

Keywords: SPOCK1, POSTN, tumor microenvironment, tumor-infiltrating immune cells, colorectal cancer

Posted Date: September 28th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-81480/v1>

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

[Read Full License](#)

Abstract

Background

Immune cells and stromal cells in the tumor microenvironment (TME) play a vital role in the initiation and progression of colorectal cancer (CRC). The study aimed to screen valuable prognostic biomarkers in CRC on the basis of stromal and immune scores.

Methods

We used the ESTIMATE algorithm to calculate the immune and stromal scores of CRC samples in TCGA. Then the CRC samples were divided into high and low score groups based on the median value of the immune and stromal scores. Differentially expressed genes (DEGs) associated with immune score and stromal score were screened. WGCNA and univariate COX regression analysis were performed to further identify key prognostic genes. The prognostic value of key genes was validated based on The Gene Expression Profiling Interactive Analysis (GEPIA) and GSE17536 dataset. TIMER and CIBERSORT algorithms were applied to analyze the correlations among key genes and tumor-infiltrating immune cells.

Results

1314 upregulated and 4 downregulated genes associated with immune score and stromal score were identified, which were significantly enriched in immune-related biological processes and pathways. Among these DEGs, SPOCK1 and POSTN were identified as key prognostic genes. High expression of SPOCK1 and POSTN was associated with advanced clinical stage, T stage, N stage, and poor prognosis of CRC. The results from CIBERSORT and TIMER revealed that SPOCK1 and POSTN were associated with tumor-infiltrating immune cells, especially macrophages and neutrophils. Besides, SPOCK1 and POSTN expression were positively correlated with the expression of immune checkpoints.

Conclusion

Collectively, our results indicate that SPOCK1 and POSTN may be novel prognostic biomarkers in CRC and correlate with immune infiltrates.

1. Introduction

Colorectal cancer (CRC) ranks the third most common malignancy around the world with high incidence and high mortality^[1]. The current treatments for CRC mainly include surgery, chemotherapy, radiation therapy, and other targeted therapies. However, the prognosis of CRC has never been satisfying, especially for patients with advanced stage^[2]. Genetic and epigenetic changes are closely related to the occurrence of tumors and the prognosis of CRC^[3,4]. Therefore, understanding the molecular mechanism of CRC is essential for developing preventive and treatment strategies of CRC.

Accumulating evidence confirmed the importance of tumor microenvironment (TME) in cancer development. The complex interaction between tumor cells and TME contributes to tumor progression and impacts the patient's clinical outcome^[5-7]. In this context, tumor-infiltrating immune cells and stromal cells, which are the main components of TME, have received extensive attention. Previous studies have demonstrated that stromal cells promote tumor angiogenesis and extracellular matrix remodeling during tumor progression^[8]. Meanwhile, numerous studies revealed that immune cells in the TME play important roles in tumorigenesis and are attractive therapeutic targets^[9-11]. Furthermore, it has been reported that infiltrating immune cells are involved in metastasis^[12]. In these biological processes, immune and stromal genes may influence the prognosis of cancer patients by regulating the abundance and function of immune cells and stromal cells. However, there is a lack of effective biomarkers related to immune and stromal score that can predict CRC patients' prognosis. Therefore, identifying these biomarkers is conducive to cancer diagnosis and explore new molecular targeted therapies.

With the emergence of high-throughput detection technology and the development of bioinformatics, transcriptome analysis has been widely used to assess the abundance of different types of cells in the TME. In this study, we applied the ESTIMATE scoring method to calculate the immune score and stromal score of each CRC sample from The Cancer Genome Atlas (TCGA) database. SPOCK1 and POSTN were identified as prognostic key genes associated with immune score and stromal score. The relationships between their expression with the clinical characteristics and prognosis of colorectal cancer were validated by kinds of analyses. CIBERSORT algorithm and Tumor Immune Estimation Resource (TIMER) were used to elucidate the association of SPOCK1 and POSTN expression with immune cell infiltration level.

In conclusion, our finding indicated that high expression of SPOCK1 and POSTN predicted poor prognosis in CRC. In addition, SPOCK1 and POSTN were enriched in a variety of immune-related pathways and were related to the regulation of tumor immune cells and the expression of multiple immune checkpoint genes, suggesting that they might be valuable prognostic biomarkers and targets for immunotherapy.

2. Materials And Methods

2.1 Data collection and processing

RNA-seq data in FPKM format and corresponding clinical information of CRC were downloaded from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>). There were 556 CRC samples with gene expression data in the TCGA. All samples were included in gene expression analysis. After excluding 45 patients with incomplete information, a total of 511 patients with complete clinical information such as survival time, survival status, age, gender, stage, T/M/N stage were included in survival analysis and clinical correlation analysis. The detailed patient characteristics of TCGA are displayed in Table S1. Moreover, as a validation dataset, the normalized gene expression profile of GSE17536 was downloaded from Gene Expression Omnibus (GEO) database

(<https://www.ncbi.nlm.nih.gov/geo/>). This dataset containing 177 CRC samples with complete clinical information.

2.2 Differentially expressed genes (DEGs) screening

The immune and stromal scores of each CRC sample were calculated based on the ESTIMATE algorithm, which was performed by R package “estimate”^[13]. The samples were divided into high-score and low-score groups respectively according to the median of the immune and stromal scores. Differentiation analysis of the gene expression was performed by using the R package “limma”, and DEGs were screened from the comparison between the high-score group and the low-score group. $|\log_2$ fold change (FC) $| \geq 1.0$ and false discovery rate (FDR) < 0.05 were set as the threshold for filtering DEGs. The intersection of the up-regulated and down-regulated DEGs in the immune and stromal score groups was identified by the online Venn diagrams tool (Venny 2.1, <https://bioinfogp.cnb.csic.es/tools/venny/>).

2.3 Functional enrichment analysis of the DEGs

Gene Ontology (GO) and Kyoto encyclopedia of Genes and Genomes (KEGG) analyses were performed for intersecting DEGs through “clusterProfiler”, “enrichplot”, “ggplot2” in R software. The terms with p-value < 0.05 were considered as statistically significant.

2.4 Construction of weighted co-expression network and identification of clinical significant modules

Weighted gene co-expression network analysis (WGCNA) was performed to identify hub gene modules associated with clinical. The expression profile data and phenotype data matrix of 1318 DEGs were selected for WGCNA^[14]. R package “WGCNA” was performed, and the scale-free topology fit index for several powers and mean connectivity were calculated to assess the soft-thresholding power for the network construction. The soft threshold value was determined when the scale independence value was 0.9. The adjacency was turned into topological overlap matrix (TOM), which could measure the network connectivity of genes. Based on the TOM dissimilarity with a minimum gene group size of 30 for the gene dendrogram, hierarchical clustering was carried out to classified similar genes into modules. The merge cut height was set as 0.25 to merge some modules founded on the dissimilarity of module eigengenes.

We calculated the correlation between co-expression modules and clinical features to identify clinically relevant modules. For intramodular analysis, Module membership (MM) and gene significance (GS) were calculated. Here we selected the significant module with a threshold of p-value < 0.05 .

2.5 Key genes screening and validation

Genes with higher MM and GS in the selected module would be defined as hub genes, and the ranked top 30 were selected. Second, prognostic genes were obtained from DEGs of the intersection by univariate Cox regression analysis with $p < 0.01$ as the cutoff value. Third, venn analysis was conducted to select the intersection of genes in the above two steps, and the genes were identified as key prognostic genes for subsequent analysis. Survival analyses (overall survival (OS) and disease-free survival (DFS)) of key

genes were carried out based on TCGA and the Genotype-Tissue Expression (GTEx) data in the Gene Expression Profiling Interactive Analysis (GEPIA) database (<http://gepia.cancer-pku.cn/>). For further validation, Kaplan-Meier curves based on GSE17536 were generated using the R package “survival”. $p < 0.05$ was considered significant.

2.6 Gene set enrichment analysis (GSEA)

According to the median expression value of key genes, CRC patients in TCGA were divided into high and low expression groups. Gene Set Enrichment Analyses (GSEA) was performed to identify enriched KEGG pathways in the two groups using the JAVA program (<https://www.broadinstitute.org/gsea>). The number of permutations was set at 1000 for each analysis. Nominal $p < 0.05$, and a false discovery rate (FDR) $q < 0.25$ were considered statistically significant. Multiple GSEA plots were visualized by “plyr”, “ggplot2”, and “grid” packages in R.

2.7 Correlation analysis between key genes and immune microenvironment

The CIBERSORT^[15] (<https://cibersort.stanford.edu/>) algorithm was performed to evaluate the relative abundances of 22 immune cells in all CRC samples based on RNA-seq from the TCGA database. Samples with $p < 0.05$ were considered qualified for further analysis. The 22 immune cells between high and low expression groups of key genes were visualized by using the “vioplot” package. The Tumor Immune Estimation Resource (TIMER) (<https://cistrome.shinyapps.io/timer/>) online database also known as a web server for analyzing tumor-infiltrating immune cells^[16]. We downloaded the immune infiltration levels of TCGA-CRC patients. The relationship between key genes expression and six immune cell subtypes (B cells, CD4+ T cells, CD8+ T cells, dendritic cells, macrophages, and neutrophils) were determined.

2.8 Statistical analysis

Statistical analyses were performed using R version 3.6.3 and SPSS V25.0. The Kaplan–Meier method was conducted to plot survival curves, and a log-rank was used as the statistical significance test. Differences among variables were compared using t-tests, nonparametric tests. The correlation of gene expression was evaluated by Spearman’s correlation. If not specified above, $p < 0.05$ was considered statistically significant.

3. Results

3.1 Analysis of differential gene expression profiles of immune and stromal scores in CRC

The immune scores and stromal scores of all CRC samples were calculated using the ESTIMATE algorithm and stratified into high and low groups according to their median levels. The gene expression profiles of the high- and low-score groups were compared. A total of 2033 DEGs were obtained based on immune scores (high vs. low), consisting of 1997 up-regulated and 36 down-regulated (Fig 1A, C, E, F).

Similarly, 2521 up-regulated and 23 down-regulated genes were identified from stromal scores (high vs. low) (Fig 1B, D, E, F). The Venn plots showed that 1314 up-regulated genes overlapped in the high immune/stromal score groups and 4 down-regulated genes overlapped in the low immune/stromal score groups.

3.2 Functional enrichment analyses of DEGs in the intersection

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed to evaluate the biological functions and pathways involved in these DEGs in the intersection. The results of GO enrichment analysis showed that these DEGs were almost enriched in immune-related terms. lymphocyte mediated immunity, immunoglobulin complex, and antigen binding were the most significant enrichment terms among biological processes, cellular components, and molecular functions (Fig 2A). KEGG analysis also displayed that these DEGs were mainly enriched in cytokine-cytokine receptor interaction, chemokine signaling pathway, and neuroactive ligand-receptor interaction (Fig 2B). The enrichment results suggested that these DEGs participate in immune-related activities and might regulate TME in CRC.

3.3 Construction of co-expression network and identification of hub genes

To further narrow candidate genes and screen hub genes related to clinicopathological factors, a total of 1318 DEGs in the intersection were included in the co-expression network analysis. 511 CRC patients with complete information including survival, stage, and T/M/N stage were selected.

The soft threshold was set to 9, with the lowest mean connectivity (Fig 3A). As shown in Fig 3B, the logarithm of the node with connectivity k $\log(k)$ was negatively correlated with the logarithm of the node probability $\log(p(k))$, and the correlation coefficient reached 0.94. The MEDissThres was set as 0.25 to merge similar modules, and five modules were eventually generated (Fig 3C). Among them, there were 240 genes in the blue module, 509 genes in the brown module, 85 genes in the green module, 321 genes in the grey module, and 163 genes in the yellow module. The grey module is a group of genes that could not fit into other modules. Correlation analysis indicated that the blue module was significantly associated with the clinical features of CRC (Fig 3D). Furthermore, the module membership in the blue module also has a significant correlation with clinical features such as survival time, clinical stage, T stage, and N stage (Fig 3E-H). We calculated the connectivity between genes in the blue module, and the top 30 genes with the highest connectivity were screened as hub genes for subsequent analysis.

3.4 Selection and validation of key genes associated with survival

Univariate Cox regression analysis was used to evaluate the prognostic values of DEGs in the intersection. Of the 1318 DEGs, 150 were identified to be significantly associated with the prognosis of CRC ($p < 0.01$). Venn analysis was used to obtain the intersection of 30 hub genes in the co-expression analysis and significant genes in univariate COX analysis (Fig 4A). As a result, SPOCK1 and POSTN were identified as key genes associated with survival. As shown in the forest plot, SPOCK1 (Hazard Ratio =

1.0861, 95%CI of ratio: 1.0279-1.1476, $p = 0.0033$), POSTN (Hazard Ratio = 1.0049, 95%CI of ratio: 1.0015-1.0084, $p = 0.0053$), age (Hazard Ratio = 1.0391, 95%CI of ratio: 1.0179-1.0608, $p = 0.0003$), stage (Hazard Ratio = 2.3940, 95%CI of ratio: 1.8654-3.0724, $p < 0.0001$), T stage (Hazard Ratio = 2.7387, 95%CI of ratio: 1.7842-4.2037, $p < 0.0001$), M stage (Hazard Ratio = 1.3957, 95%CI of ratio: 1.1575-1.6830, $p = 0.0005$), N stage (Hazard Ratio = 1.9759, 95%CI of ratio: 1.5428-2.5308, $p < 0.0001$) were relevant to OS among patients with CRC (Fig 4B).

The survival analysis of GEPIA data revealed that CRC patients with high expression of SPOCK1 (Hazard Ratio = 2, Logrank $p = 0.0015$, Fig 4C) and POSTN (Hazard Ratio = 1.7, Logrank $p = 0.017$, Fig 4E) had poor OS. Meanwhile, high expression of SPOCK1 (Hazard Ratio = 2, Logrank $p = 0.0016$, Fig 4D) and POSTN (Hazard Ratio = 1.6, Logrank $p = 0.022$, Fig 4F) were also significantly correlated with DFS. To validate the prognostic value of hub genes, we performed survival analyses based on an independent subset GSE17536. Consistent with the previous results, high expressions of SPOCK1 and POSTN were significantly associated with worse OS ($p < 0.01$, Fig 4G and 4I) and DFS ($p < 0.01$, Fig 4H and 4J). Summary above, SPOCK1 and POSTN have a greater prognostic value in CRC.

3.5 Association of SPOCK1 and POSTN expression with clinicopathological factors

To investigate the relationships between key prognostic genes and the clinical characteristics of CRC patients in TCGA, we compared the expression levels of SPOCK1 and POSTN in different clinical stage, T stage, M stage, and N stage. The results showed that differential expression of SPOCK1 was significantly associated with clinical stage (Fig 5A), T stage (Fig 5B), and N stage (Fig 5C). Moreover, differential expression of POSTN was also related to stage (Fig 5E), T stage (Fig 5F), and N stage (Fig 5G). However, the expression of SPOCK1 and POSTN had no significant influence on M stage (Fig 5D and 5H).

3.6 Gene set enrichment analysis of key genes

To further explore the underlying mechanism, we performed GSEA between high-expression groups and low-expression groups based on the median levels of SPOCK1 and POSTN expression, respectively. As shown in Fig 6, the genes in SPOCK1 high-expression group were mainly enriched in immune-related pathways like chemokine signaling, cytokine-cytokine receptor interaction, and ECM-receptor interaction (Fig 6A). However, the genes in the SPOCK1 low expression group were more significantly enriched in metabolic-related pathways, such as aminoacyl tRNA biosynthesis, butanoate metabolism, and citrate cycle TCA cycle (Fig 6B). In the POSTN high expression group, genes were also mainly enriched in B cell receptor signaling pathway, chemokine signaling pathway, cytokine-cytokine receptor interaction, and other immune-related pathways (Fig 6C). While, the genes in POSTN low-expression group were enriched significantly in oxidative phosphorylation, parkinsons disease, pentose phosphate pathway (Fig 6D). These results suggested that SPOCK1 and POSTN might have potential regulatory effects on the immune microenvironment of CRC.

3.7 Correlation of SPOCK1 and POSTN expression with immune infiltration levels

To confirm the relationships between SPOCK1 and POSTN expression with the immune microenvironment. We calculated the relative abundances of the 22 immune cells in all CRC samples using the CIBERSORT algorithm. The violin diagram provided visualization of the differences in 22 immune cells between the high and low expression groups of SPOCK1 and POSTN. The results showed that plasma cells ($p < 0.001$), resting CD4 memory T cells ($p = 0.003$), monocytes ($p = 0.005$), and activated dendritic cells ($p = 0.023$) were in higher proportions in the SPOCK1 low expression group, whereas macrophages M0 ($p < 0.001$), macrophages M1 ($p = 0.046$), and neutrophils ($p = 0.005$) were in higher proportions in the SPOCK1 high expression group (Fig 7A). The POSTN low expression group also had a higher proportion of plasma cells ($p < 0.001$), resting CD4 memory CD4 T cells ($p = 0.046$), and monocytes ($p = 0.002$), whereas the POSTN high expression group had a higher proportion of macrophages M0 ($p = 0.001$), macrophages M2 ($p = 0.003$), and neutrophils ($p < 0.001$) (Fig 7B). Additionally, we used the TIMER algorithm to determine the correlations between SPOCK1 and POSTN expression with 6 kinds of infiltration immune cells. The analysis showed that SPOCK1 and POSTN expression had significantly positive correlations with infiltrating levels of B cells, CD4+ T cells, CD8+ T cells, dendritic cells, macrophages and neutrophils (Fig 7C and 7D). From the above results, both SPOCK1 and POSTN were related to immune cell infiltration in CRC, especially macrophages and neutrophils.

3.8 Correlation of SPOCK1 and POSTN expression with immune checkpoint genes expression

Immune checkpoint molecules have been identified to suppress antitumor immune responses in solid tumors. Therefore, we analyzed the correlation between SPOCK1 and POSTN expression with immune checkpoint gene expression such as PD-L1, PD-1, PD-L2, CTLA4, TIM-3, and B7-H3. The results showed that SPOCK1 expression was significantly positively correlated with PD-1 ($R = 0.23$, $p = 5.9e-08$, Fig 8A), PD-L1 ($R = 0.33$, $p = 8.7e-16$, Fig 8B), PD-L2 ($R=0.59$, $p < 2.2e-16$, Fig 8C), CTLA4 ($R = 0.34$, $p = 5.8e-16$, Fig 8D), TIM-3 ($R = 0.62$, $p < 2.2e-16$, Fig 8E), and B7-H3 ($R = 0.33$, $p = 4.6e-15$, Fig 8F). Similarly, POSTN was also positively correlated with PD-1 ($R = 0.24$, $p = 1.1e-08$, Fig 8G), PD-L1 ($R = 0.45$, $p < 2.2e-16$, Fig 8H), PD-L2 ($R = 0.67$, $p < 2.2e-16$, Fig 8I), CTLA4 ($R = 0.37$, $p < 2.2e-16$, Fig 8J), TIM-3 ($R = 0.67$, $p < 2.2e-16$, Fig 8K), and B7-H3 ($R = 0.33$, $p = 3.6e-15$, Fig 8L). These suggested that high expression of SPOCK1 and POSTN might potentially regulate tumor immune evasion in CRC.

4. Discussion

In the presented study, we screened out differentially expressed genes (DEGs) based on the immune and stromal scores of CRC samples in TCGA. SPOCK1 and POSTN were identified as key prognostic genes for CRC. In the validation phases, the high expression of SPOCK1 and POSTN was shown to correlate with poor prognosis and the late clinical stage of CRC. Furthermore, we found that SPOCK1 and POSTN were involved in immune activities and were related to immune cell infiltration and immune evasion.

In TME, immune and stromal cells play a critical role in tumorigenesis. Growing studies have shown that the composition of the TME can predict patient prognosis and serve as an important target for cancer therapy^[17-20]. It has been reported that immune and stroma classification of CRC is relevant to precision

immunotherapy^[21]. In recent years, immunotherapy has made great progress in the treatment of CRC, especially in patients with microsatellite instability^[22,23]. Because the molecular mechanism of the TME in CRC is still unclear, and there is a lack of effective immune biomarkers, current immunotherapy has only achieved significant clinical effects in a subset of CRC patients^[24]. Therefore, it is necessary to take an in-depth understanding of the composition of TME and to investigate novel prognostic biomarkers for the immunotherapy of CRC.

Here, we use the ESTIMATE method to calculate the immune and stromal scores of CRC samples from TCGA. The CRC samples were divided into two groups based on the median value of the immune scores and stromal scores. DEGs between the two groups were selected. 1314 upregulated DEGs were in common among the high immune/stromal score groups and 4 downregulated DEGs were in common among the low immune/stromal score groups. Functional enrichment analysis showed that the biological processes and main pathways of these DEGs were related to immune-related terms, such as lymphocyte mediated immunity and cytokine-cytokine receptor interaction. These results indicate that these DEGs are highly correlated with immune response and tumor immune microenvironment. Based on the obtained 1318 DEGs, the WGCNA method was used to screen clinically relevant hub genes; while univariate Cox regression analysis was used to identify survival-related genes. The intersection involved SPOCK1 and POSTN as the key prognostic genes. We validated that high expression of SPOCK1 and POSTN was associated with the poor prognosis of CRC patients by applying OS and DFS analyses based on GEPIA and GSE17536. This is consistent with previous studies on POSTN^[25,26], but the prognostic value of SPOCK1 in CRC has not been reported before. Besides, high expression of SPOCK1 and POSTN was correlated to higher clinical stage, T stage, and N stage, but not with M stage, which is consistent with the WGCNA analysis. Accordingly, the results indicated the potential of SPOCK1 and POSTN as prognostic markers and therapeutic targets for TME in CRC.

SPOCK1 encodes a Ca²⁺-binding matricellular glycoprotein, which belongs to the SPARC family. SPOCK1 has been demonstrated to function in cell proliferation, migration, and apoptosis in certain types of cancer, such as pancreatic ductal adenocarcinoma^[27], lung cancer^[28], colorectal cancer^[29], and hepatocellular carcinoma^[30], indicating that SPOCK1 plays an important role in oncogenesis. To elucidate the biological roles of SPOCK1, we conducted GSEA to explore the relevant pathways. It was shown that the SPOCK1 high expression group is mainly enriched in immune-related pathways, such as chemokine signaling pathway, cytokine-cytokine receptor interaction, JAK-STAT signaling pathway. In addition, tumorigenesis-related pathways, such as ECM receptor interaction and MAPK signaling pathway, were also enriched in the SPOCK1 high expression group. Previous studies have shown that SPOCK1 is a key regulator of ECM and can mediate EMT in cancer cells^[31]. Altogether, these results suggested that SPOCK1 promotes the tumor progression by affecting these tumor and immune-related pathways, leading to a poor prognosis in CRC patients. POSTN, as a small extracellular matrix protein, plays a vital role in the regulation of cell-matrix interaction, which is considered to associate with TME and tumor progression^[32]. Accumulated evidence suggested that POSTN promotes tumor metastasis by regulating immune responses^[33,34]. In our study, the GSEA results showed that high expression levels of

POSTN were associated with many immune-related pathways, for instance, B cell receptor signaling pathway, chemokine signaling pathway, and cytokine-cytokine receptor interaction. It could be hypothesized that the regulation of immune-related signaling pathways may be involved in the regulatory role of POSTN in the clinical stage and prognosis in CRC.

Tumor-infiltrating immune cells (TIICs) are an important part of the complex TME that regulates tumorigenesis^[35]. Therefore, we adopted the CIBERSORT algorithm in the present study to further explore the relationships of SPOCK1 and POSTN with the 22 TIICs subsets of immune reaction. The results showed that there were seven types of immune cells, plasma cells, CD4 memory resting T cells, monocytes, macrophages M0, macrophages M1, activated dendritic cells, and neutrophils, that had significantly different proportions based on the expression levels of SPOCK1. Similarly, the expression levels of SPOCK1 were significantly correlated with infiltration levels of plasma cells, CD4 memory resting T cells, monocytes, macrophage M0, macrophage M2, and neutrophils. For validation, we used TIMER to analyze the correlation between the expression of SPOCK1 and POSTN with immune cell infiltration. Our results showed that SPOCK1 and POSTN were significantly associated with immune cell infiltration. Altogether, these results indicated that SPOCK1 and POSTN had strongest and consistent correlation with macrophage and neutrophil cells. It has been shown that increased intratumoral neutrophil in CRC is related to the acquisition of malignant phenotype and is an independent factor for the poor prognosis of CRC patients^[36,37]. The role of macrophages in CRC is controversial, but many studies have shown that tumor-infiltrating macrophages promote the metastasis of CRC, leading to poor prognosis^[38-41]. Zhang Y et al. found that SPOCK1 is related to the recruitment and differentiation of macrophages^[42]. Recent evidence revealed that POSTN could recruit M2 tumor-related macrophages and promote malignant growth^[43]. These are consistent with our conclusion. Taken together, it is possible that SPOCK1 and POSTN potentially regulate immune cell infiltration in CRC microenvironment.

Immune checkpoint molecules are often associated with tumor cell immune evasion and tumor progression. In return, immune checkpoint inhibitors have shown great success in the treatment of CRC^[22,44]. Our results showed that SPOCK1 and POSTN expression were positively correlated with the expression of immune checkpoints, including PD-1, PD-L1, PD-L2, CTLA4, TIM-3, and B7-H3. PD-L1/2 suppresses T-cell function through the PD-1 receptor, causing tumor cells to escape from immune surveillance. Previous studies have shown that PD-1, PD-L1, PD-L2, and CTLA4 expressions were associated with a poor prognosis in CRC patients^[45-48]. TIM-3 and B7-H3 are considered as novel promising targets for immunotherapy. Recent researches have revealed that TIM-3 and B7-H3 are also involved in evasion of cancer immune surveillance and CRC progression^[49,50]. These results suggest that SPOCK1 and POSTN may play a role in immune evasion, which partly explains their potential mechanisms for promoting tumor progression.

This study had some limitations. First, the roles of SPOCK1 and POSTN in CRC were analyzed based on TCGA or GEO data, so our results need to be verified with larger sample sizes. Second, the relationships

between SPOCK1 and POSTN with infiltration immune cells should be confirmed by in vitro and in vivo experiments.

5. Conclusion

In conclusion, SPOCK1 and POSTN were identified as key prognostic genes related to TME in CRC. High expression of SPOCK1 and POSTN was associated with late clinical stage and predicted a poor prognosis. Further analysis indicated that increased levels of SPOCK1 and POSTN were positively correlated with the infiltration level of macrophages and neutrophils and immune checkpoints expression. Therefore, SPOCK1 and POSTN could be used as prognostic biomarkers and therapeutic targets in CRC.

Declarations

Availability of data and materials

All the analysis data were accessed from TCGA database (<https://portal.gdc.cancer.gov/>) and GEO database (<https://www.ncbi.nlm.nih.gov/geo/>).

Ethical approval and consent to participate

The study didn't involve animal experiments and human specimens, no ethics related issues.

Funding

This study was supported by the National Science Foundation of China (No:81870390) and the Program of Excellent Doctoral (Postdoctoral) of Zhong nan Hospital of Wuhan University (Grant No. ZNYB2019003).

Acknowledgements

None.

Authors' contributions

Mengting Li and Qiu Zhao designed the study. Mengting Li, Wenjie Li collected the data, Wanhui Wei, Qian Hu, and Caiqin Gan analyzed the data. Haizhou Wang, Yanan Peng supervised the data and provided statistical advice. Mengting Li wrote the paper, Qiu Zhao, Lan Liu, and Fang Wang reviewed the paper. All authors approved the final manuscript.

Consent for publication

Not applicable.

Competing interests

All authors confirmed that there were no conflicts of interest with the contents of this article.

References

1. Siegel, R., Miller, K., Jemal, A. Cancer statistics, 2020. *CA: a cancer journal for clinicians* **70**, 7-30, doi:10.3322/caac.21590 (2020).
2. Miller, K., Nogueira, L., Mariotto, A., Rowland, J., Yabroff, K. et al. Cancer treatment and survivorship statistics, 2019. *CA: a cancer journal for clinicians* **69**, 363-385, doi:10.3322/caac.21565 (2019).
3. Lao, V., Grady, W. Epigenetics and colorectal cancer. *Nature reviews. Gastroenterology & hepatology* **8**, 686-700, doi:10.1038/nrgastro.2011.173 (2011).
4. Okugawa, Y., Grady, W., Goel, A. Epigenetic Alterations in Colorectal Cancer: Emerging *Gastroenterology* **149**, 1204-1225.e1212, doi:10.1053/j.gastro.2015.07.011 (2015).
5. Quail, D., Joyce, J. Microenvironmental regulation of tumor progression and metastasis. *Nature medicine* **19**, 1423-1437, doi:10.1038/nm.3394 (2013).
6. Belli, C., Trapani, D., Viale, G., D'Amico, P., Duso, B. et al. Targeting the microenvironment in solid tumors. *Cancer treatment reviews* **65**, 22-32, doi:10.1016/j.ctrv.2018.02.004 (2018).
7. Kamińska, K., Szczylik, C., Bielecka, Z., Bartnik, E., Porta, C. et al. The role of the cell-cell interactions in cancer progression. *Journal of cellular and molecular medicine* **19**, 283-296, doi:10.1111/jcmm.12408 (2015).
8. Bussard, K., Mutkus, L., Stumpf, K., Gomez-Manzano, C., Marini, F. Tumor-associated stromal cells as key contributors to the tumor microenvironment. *Breast cancer research : BCR* **18**, 84, doi:10.1186/s13058-016-0740-2 (2016).
9. Ngiow, S., Young, A. Re-education of the Tumor Microenvironment With Targeted Therapies and Immunotherapies. *Frontiers in immunology* **11**, 1633, doi:10.3389/fimmu.2020.01633 (2020).
10. Lei, X., Lei, Y., Li, J., Du, W., Li, R. et al. Immune cells within the tumor microenvironment: Biological functions and roles in cancer immunotherapy. *Cancer letters* **470**, 126-133, doi:10.1016/j.canlet.2019.11.009 (2020).
11. Zhang, Y., Zhang, Z. The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. *Cellular & molecular immunology* **17**, 807-821, doi:10.1038/s41423-020-0488-6 (2020).
12. Van den Eynde, M., Mlecnik, B., Bindea, G., Fredriksen, T., Church, S. et al. The Link between the Multiverse of Immune Microenvironments in Metastases and the Survival of Colorectal Cancer Patients. *Cancer cell* **34**, 1012-1026.e1013, doi:10.1016/j.ccell.2018.11.003 (2018).
13. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature communications* **4**, 2612, doi:10.1038/ncomms3612 (2013).
14. Langfelder, P., Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).

15. He, X., Hang, D., Wu, K., Naylor, J., Drew, D. et al. Long-term Risk of Colorectal Cancer After Removal of Conventional Adenomas and Serrated Polyps. *Gastroenterology* **158**, 852-861.e854, doi:10.1053/j.gastro.2019.06.039 (2020).
16. Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q. et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer research* **77**, e108-e110, doi:10.1158/0008-5472.Can-17-0307 (2017).
17. Efstathiou, J., Mouw, K., Gibb, E., Liu, Y., Wu, C. et al. Impact of Immune and Stromal Infiltration on Outcomes Following Bladder-Sparing Trimodality Therapy for Muscle-Invasive Bladder Cancer. *European urology* **76**, 59-68, doi:10.1016/j.eururo.2019.01.011 (2019).
18. Ozpiskin, O., Zhang, L., Li, J. Immune targets in the tumor microenvironment treated by radiotherapy. *Theranostics* **9**, 1215-1231, doi:10.7150/thno.32648 (2019).
19. Giraldo, N., Sanchez-Salas, R., Peske, J., Vano, Y., Becht, E. et al. The clinical role of the TME in solid cancer. *British journal of cancer* **120**, 45-53, doi:10.1038/s41416-018-0327-z (2019).
20. Mlecnik, B., Bindea, G., Angell, H., Maby, P., Angelova, M. et al. Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability. *Immunity* **44**, 698-711, doi:10.1016/j.immuni.2016.02.025 (2016).
21. Becht, E., de Reyniès, A., Giraldo, N., Pilati, C., Buttard, B. et al. Immune and Stromal Classification of Colorectal Cancer Is Associated with Molecular Subtypes and Relevant for Precision Immunotherapy. *Clinical cancer research : an official journal of the American Association for Cancer Research* **22**, 4057-4066, doi:10.1158/1078-0432.Ccr-15-2879 (2016).
22. Singh, P., Sharma, P., Krishnan, G., Lockhart, A. Immune checkpoints and immunotherapy for colorectal cancer. *Gastroenterology report* **3**, 289-297, doi:10.1093/gastro/gov053 (2015).
23. Kalyan, A., Kircher, S., Shah, H., Mulcahy, M., Benson, A. Updates on immunotherapy for colorectal cancer. *Journal of gastrointestinal oncology* **9**, 160-169, doi:10.21037/jgo.2018.01.17 (2018).
24. Ganesh, K., Stadler, Z., Cercek, A., Mendelsohn, R., Shia, J. et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nature reviews. Gastroenterology & hepatology* **16**, 361-375, doi:10.1038/s41575-019-0126-x (2019).
25. Ben, Q., Zhao, Z., Ge, S., Zhou, J., Yuan, F. et al. Circulating levels of periostin may help identify patients with more aggressive colorectal cancer. *International journal of oncology* **34**, 821-828, doi:10.3892/ijo_00000208 (2009).
26. Dong, D., Zhang, L., Jia, L., Ji, W., Wang, Z. et al. Identification of Serum Periostin as a Potential Diagnostic and Prognostic Marker for Colorectal Cancer. *Clinical laboratory* **64**, 973-981, doi:10.7754/Clin.Lab.2018.171225 (2018).
27. Li, J., Ke, J., Fang, J., Chen, J. A potential prognostic marker and therapeutic target: SPOCK1 promotes the proliferation, metastasis, and apoptosis of pancreatic ductal adenocarcinoma cells. *Journal of cellular biochemistry* **121**, 743-754, doi:10.1002/jcb.29320 (2020).
28. Wang, T., Liu, X., Tian, Q., Liang, T., Chang, P. Reduced SPOCK1 expression inhibits non-small cell lung cancer cell proliferation and migration through Wnt/ β -catenin signaling. *European review for medical*

- and pharmacological sciences* **22**, 637-644, doi:10.26355/eurrev_201802_14288 (2018).
29. Zhao, P., Guan, H., Dai, Z., Ma, Y., Liu, X. et al. Knockdown of SPOCK1 Inhibits the Proliferation and Invasion in Colorectal Cancer Cells by Suppressing the PI3K/Akt Pathway. *Oncology research* **24**, 437-445, doi:10.3727/096504016x14685034103554 (2016).
30. Li, P., Xiao, Z., Luo, J., Zhang, Y., Lin, L. MiR-139-5p, miR-940 and miR-193a-5p inhibit the growth of hepatocellular carcinoma by targeting SPOCK1. *Journal of cellular and molecular medicine* **23**, 2475-2488, doi:10.1111/jcmm.14121 (2019).
31. Sun, L., Li, S., Guo, Q., Zhou, W., Zhang, H. SPOCK1 Involvement in Epithelial-to-Mesenchymal Transition: A New Target in Cancer Therapy? *Cancer management and research* **12**, 3561-3569, doi:10.2147/cmar.S249754 (2020).
32. González-González, L., Alonso, J. Periostin: A Matricellular Protein With Multiple Functions in Cancer Development and Progression. *Frontiers in oncology* **8**, 225, doi:10.3389/fonc.2018.00225 (2018).
33. Schwanekamp, J., Lorts, A., Vagnozzi, R., Vanhoutte, D., Molkentin, J. Deletion of Periostin Protects Against Atherosclerosis in Mice by Altering Inflammation and Extracellular Matrix Remodeling. *Arteriosclerosis, thrombosis, and vascular biology* **36**, 60-68, doi:10.1161/atvbaha.115.306397 (2016).
34. Wang, Z., Xiong, S., Mao, Y., Chen, M., Ma, X. et al. Periostin promotes immunosuppressive premetastatic niche formation to facilitate breast tumour metastasis. *The Journal of pathology* **239**, 484-495, doi:10.1002/path.4747 (2016).
35. Domingues, P., González-Tablas, M., Otero, Á., Pascual, D., Miranda, D. et al. Tumor infiltrating immune cells in gliomas and meningiomas. *Brain, behavior, and immunity* **53**, 1-15, doi:10.1016/j.bbi.2015.07.019 (2016).
36. Rao, H., Chen, J., Li, M., Xiao, Y., Fu, J. et al. Increased intratumoral neutrophil in colorectal carcinomas correlates closely with malignant phenotype and predicts patients' adverse prognosis. *PloS one* **7**, e30806, doi:10.1371/journal.pone.0030806 (2012).
37. Mizuno, R., Kawada, K., Itatani, Y., Ogawa, R., Kiyasu, Y. et al. The Role of Tumor-Associated Neutrophils in Colorectal Cancer. *International journal of molecular sciences* **20**, doi:10.3390/ijms20030529 (2019).
38. Jedinak, A., Dudhgaonkar, S., Sliva, D. Activated macrophages induce metastatic behavior of colon cancer cells. *Immunobiology* **215**, 242-249, doi:10.1016/j.imbio.2009.03.004 (2010).
39. Barbera-Guillem, E., Nyhus, J., Wolford, C., Friece, C., Sampsel, J. Vascular endothelial growth factor secretion by tumor-infiltrating macrophages essentially supports tumor angiogenesis, and IgG immune complexes potentiate the process. *Cancer research* **62**, 7042-7049 (2002).
40. Afik, R., Zigmond, E., Vugman, M., Klepfish, M., Shimshoni, E. et al. Tumor macrophages are pivotal constructors of tumor collagenous matrix. *The Journal of experimental medicine* **213**, 2315-2331, doi:10.1084/jem.20151193 (2016).
41. Kang, J., Chen, J., Lee, C., Chang, J., Shieh, Y. Intratumoral macrophage counts correlate with tumor progression in colorectal cancer. *Journal of surgical oncology* **102**, 242-248, doi:10.1002/jso.21617

- (2010).
42. Zhang, Y., Chen, Q., Ross, A. Retinoic acid and tumor necrosis factor- α induced monocytic cell gene expression is regulated in part by induction of transcription factor MafB. *Experimental cell research* **318**, 2407-2416, doi:10.1016/j.yexcr.2012.07.011 (2012).
 43. Zhou, W., Ke, S., Huang, Z., Flavahan, W., Fang, X. et al. Periostin secreted by glioblastoma stem cells recruits M2 tumour-associated macrophages and promotes malignant growth. *Nature cell biology* **17**, 170-182, doi:10.1038/ncb3090 (2015).
 44. Passardi, A., Canale, M., Valgiusti, M., Ulivi, P. Immune Checkpoints as a Target for Colorectal Cancer Treatment. *International journal of molecular sciences* **18**, doi:10.3390/ijms18061324 (2017).
 45. Enkhbat, T., Nishi, M., Takasu, C., Yoshikawa, K., Jun, H. et al. Programmed Cell Death Ligand 1 Expression Is an Independent Prognostic Factor in Colorectal Cancer. *Anticancer research* **38**, 3367-3373, doi:10.21873/anticancer.12603 (2018).
 46. Li, Y., He, M., Zhou, Y., Yang, C., Wei, S. et al. The Prognostic and Clinicopathological Roles of PD-L1 Expression in Colorectal Cancer: A Systematic Review and Meta-Analysis. *Frontiers in pharmacology* **10**, 139, doi:10.3389/fphar.2019.00139 (2019).
 47. Wang, H., Yao, H., Li, C., Liang, L., Zhang, Y. et al. PD-L2 expression in colorectal cancer: Independent prognostic effect and targetability by deglycosylation. *Oncoimmunology* **6**, e1327494, doi:10.1080/2162402x.2017.1327494 (2017).
 48. Omura, Y., Toiyama, Y., Okugawa, Y., Yin, C., Shigemori, T. et al. Prognostic impacts of tumoral expression and serum levels of PD-L1 and CTLA-4 in colorectal cancer patients. *Cancer immunology, immunotherapy : CII*, doi:10.1007/s00262-020-02645-1 (2020).
 49. Yu, M., Lu, B., Liu, Y., Me, Y., Wang, L. et al. Tim-3 is upregulated in human colorectal carcinoma and associated with tumor progression. *Molecular medicine reports* **15**, 689-695, doi:10.3892/mmr.2016.6065 (2017).
 50. Sun, J., Chen, L., Zhang, G., Jiang, J., Zhu, M. et al. Clinical significance and regulation of the costimulatory molecule B7-H3 in human colorectal carcinoma. *Cancer immunology, immunotherapy : CII* **59**, 1163-1171, doi:10.1007/s00262-010-0841-1 (2010).

Figures

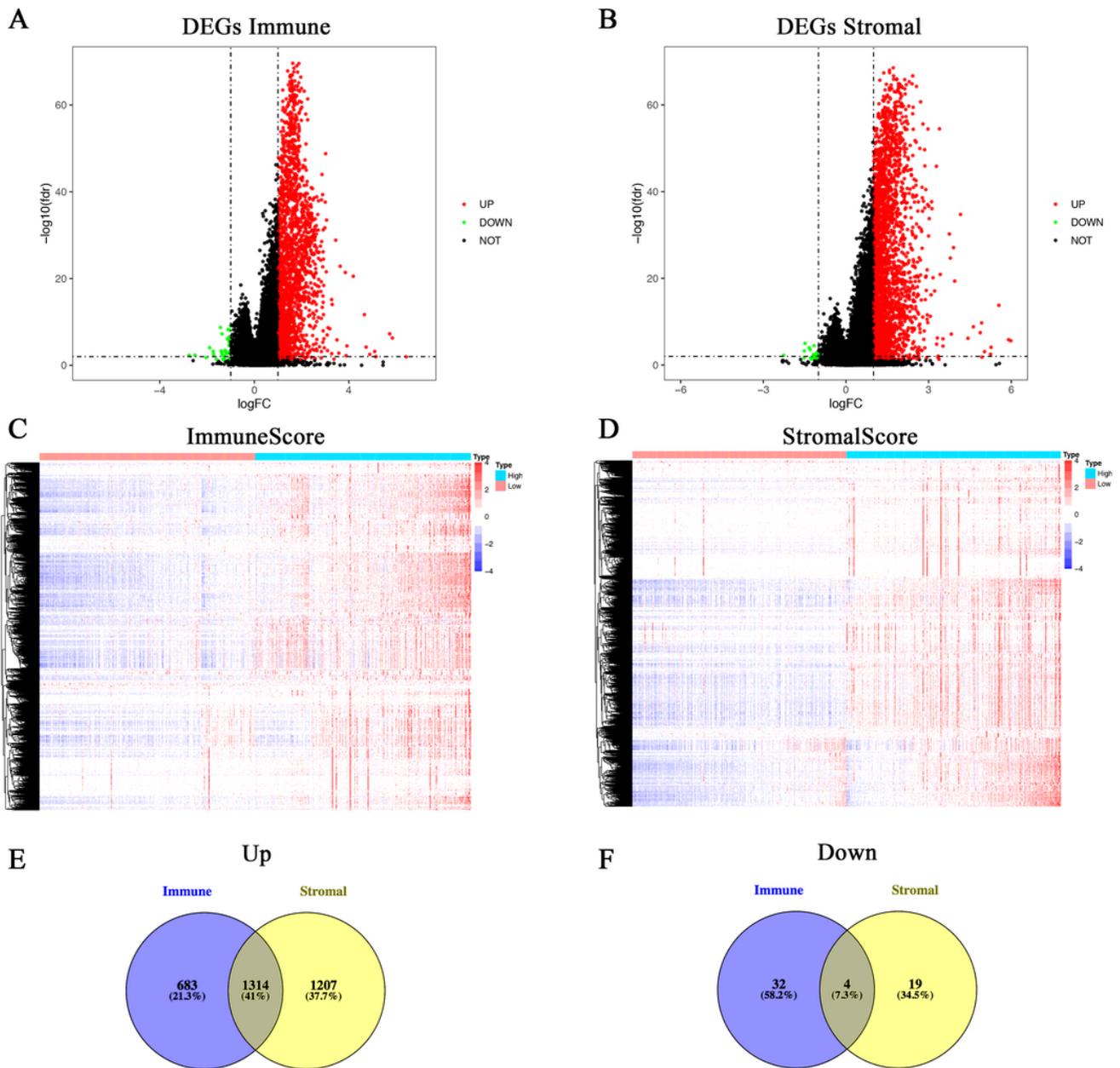


Figure 1

Identification of DEGs based on immune and stromal scores in CRC. (A, C) Volcano plot and heatmap for DEGs generated by comparison of the high immune score group vs. the low immune score group ($p < 0.05$, fold change >1). (B, D) Volcano Plot and Heatmap for DEGs generated by comparison of the high stromal score group vs. the low stromal score group ($p < 0.05$, fold change >1). (E) Venn diagram showing common up-regulated DEGs in the high immune/stromal score groups. (F) Venn diagram showing common down-regulated DEGs in the low immune/stromal score groups. DEGs, differentially expressed genes.

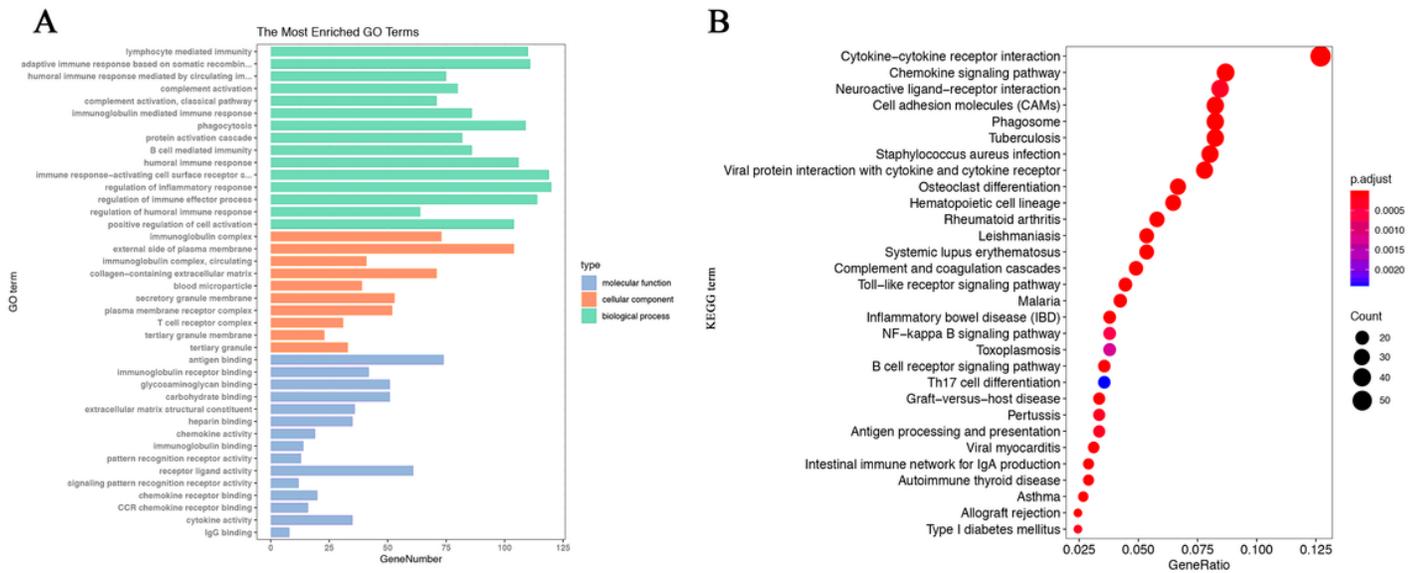


Figure 2

Functional enrichment analyses of DEGs in the intersection. (A) GO enrichment analysis for 1318 DEGs in the intersection. (B) KEGG enrichment analysis for 1318 DEGs in the intersection. GO, gene ontology; KEGG, Kyoto encyclopedia of Genes and Genomes.

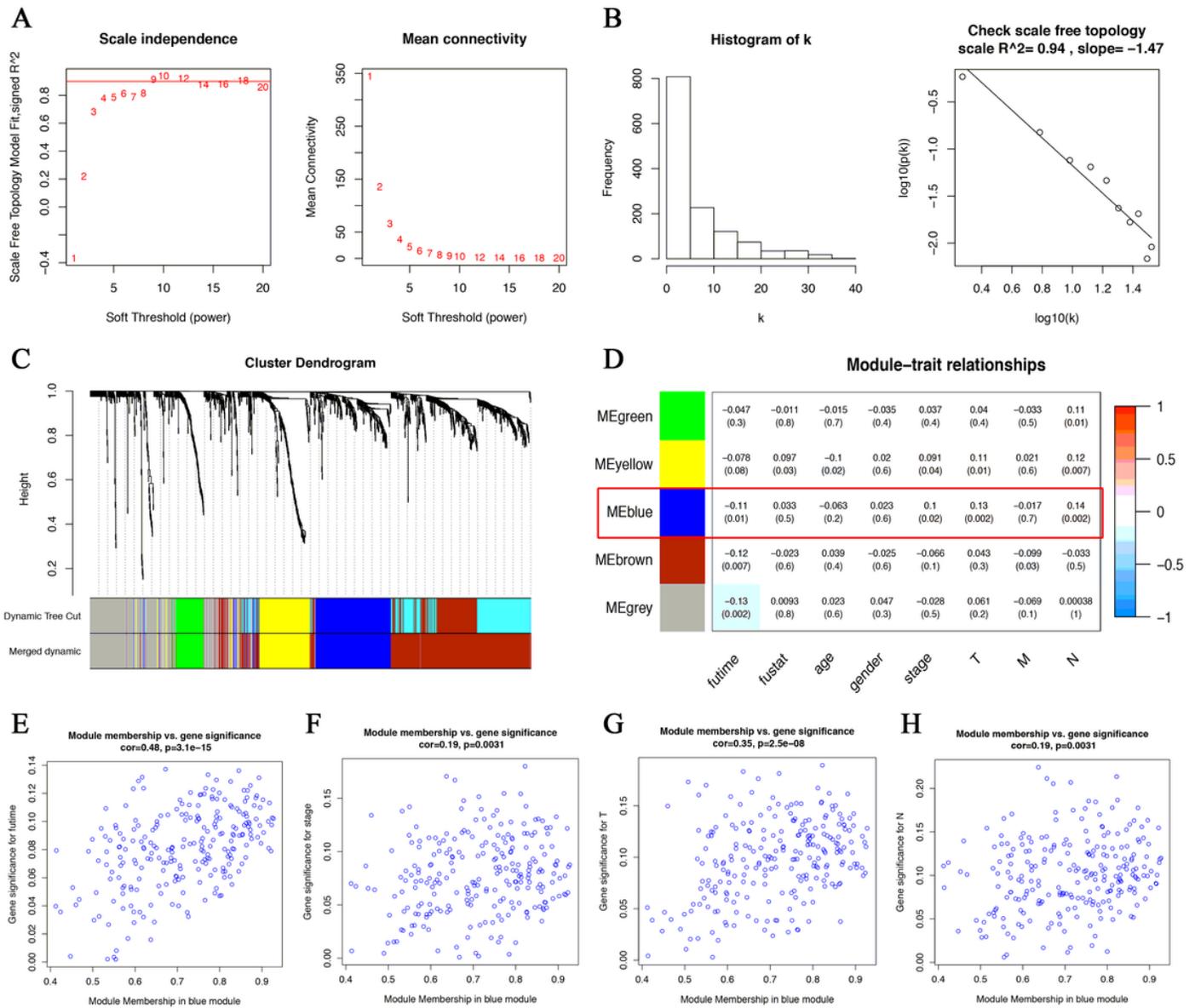


Figure 3

WGCNA analysis and the correlation between the key module and the clinical characters of CRC. (A) Scale independence and mean connectivity analysis. (B) visualized scale free topology. (C) The gene cluster dendrogram obtained by hierarchical clustering of adjacency-based dissimilarity. (D) The relationships between clustered modules and clinical traits. (E-H) Module-trait membership plot shows that the blue module is significantly correlated with survival time, clinical stage, T stage, and N stage. WGCNA, weighted gene co-expression network analysis.

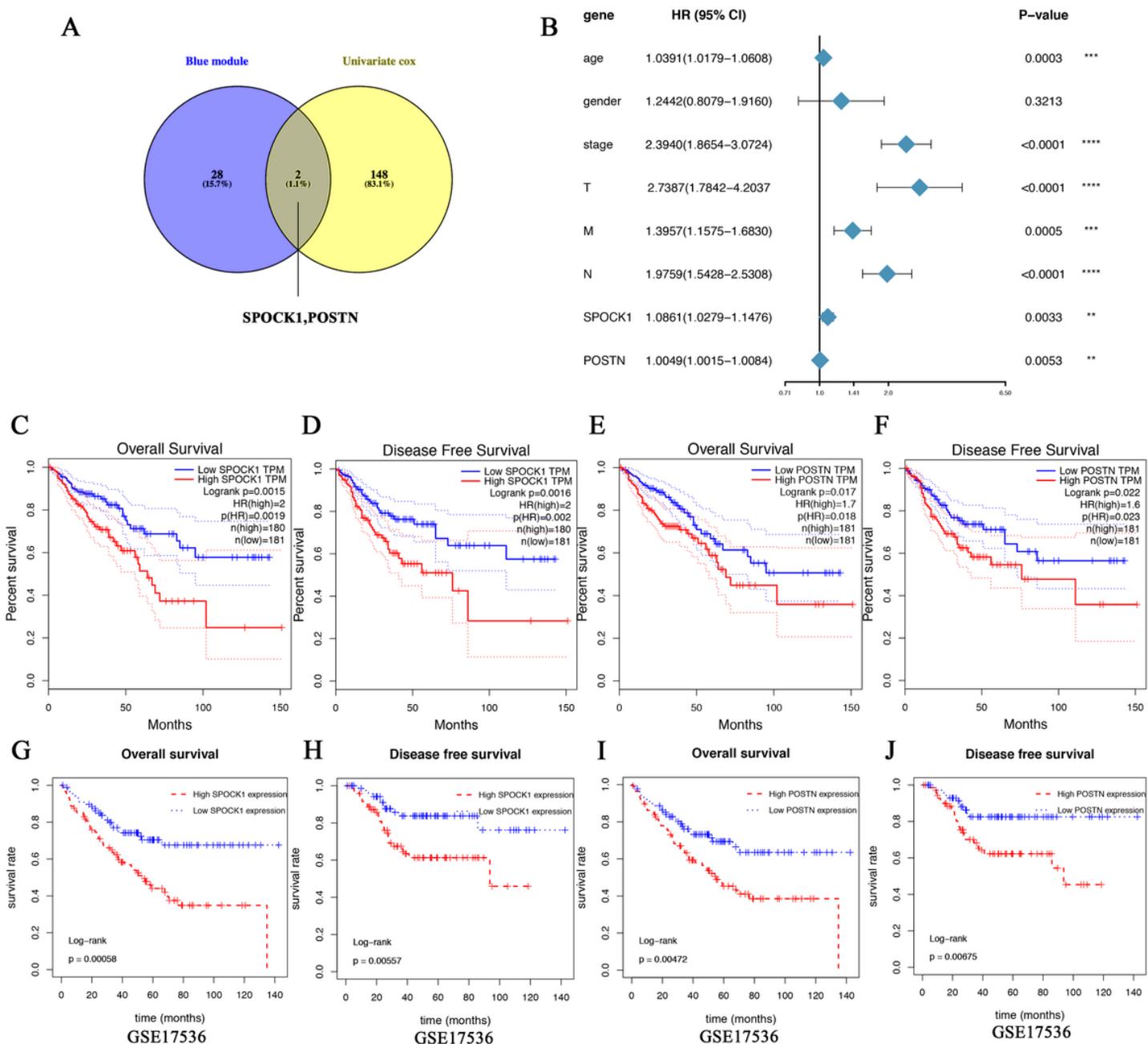


Figure 4

Identification and validation of key prognostic genes. (A) Venn plots showing the intersection between hub genes in the blue module and top significant genes in univariate COX. (B) Forest map showing the hazard ratio of SPOCK1, POSTN and clinical characters in CRC. (C, D) OS and DFS of SPOCK1 high and low expression CRC patients analyzed in GEPIA with TCGA data. (E, F) OS and DFS of POSTN high and low expression CRC patients analyzed in GEPIA with TCGA data. (G, H) OS and DFS of SPOCK1 high and low expression CRC patients analyzed in GSE17536 dataset. (I, J) OS and DFS of POSTN high and low expression CRC patients analyzed in GSE17536 dataset. GEPIA, the gene expression profiling interactive analysis; OS, overall survival; DFS, disease-free survival.

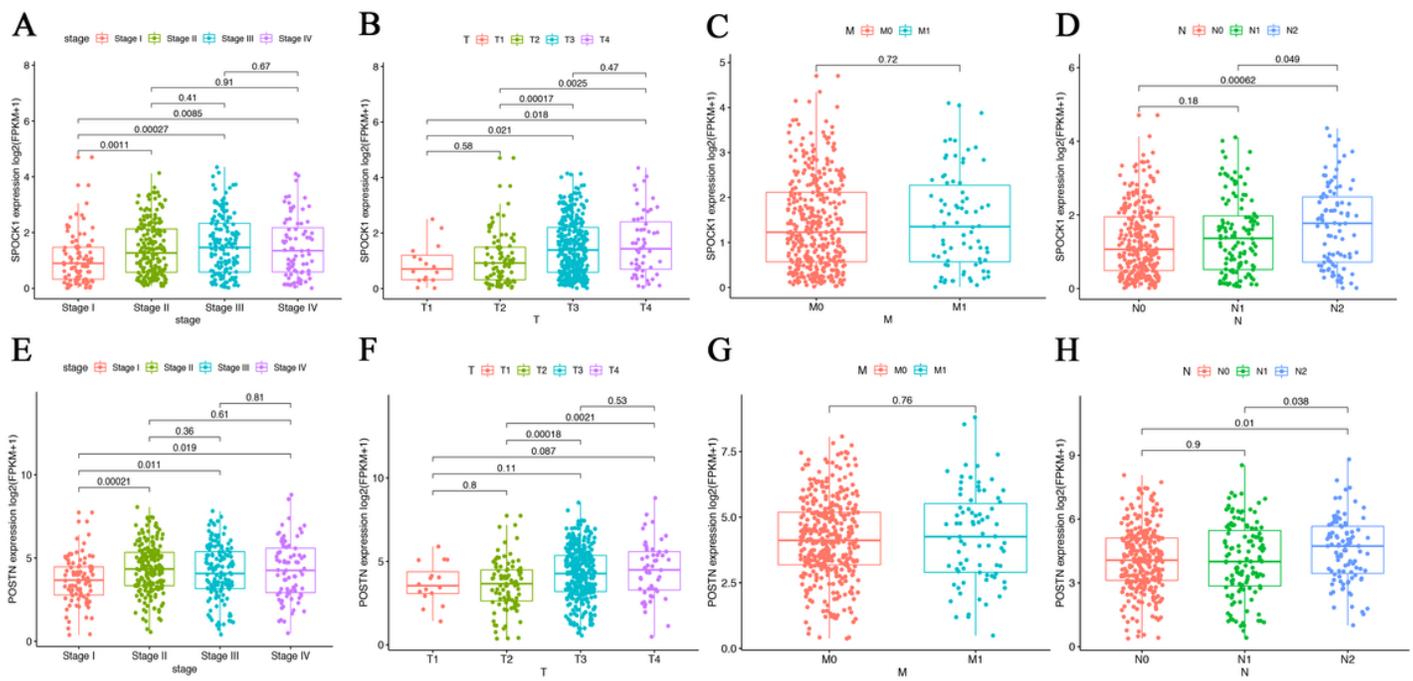


Figure 5

The association between SPOCK1 and POSTN expression with clinical features of CRC. (A-D) The correlation between with clinical stage, T stage, N stage, and M stage in low expression and high expression of SPOCK1 groups. (E-I) The correlation between clinical stage, T stage, N stage, and M stage in low expression and high expression of POSTN groups.

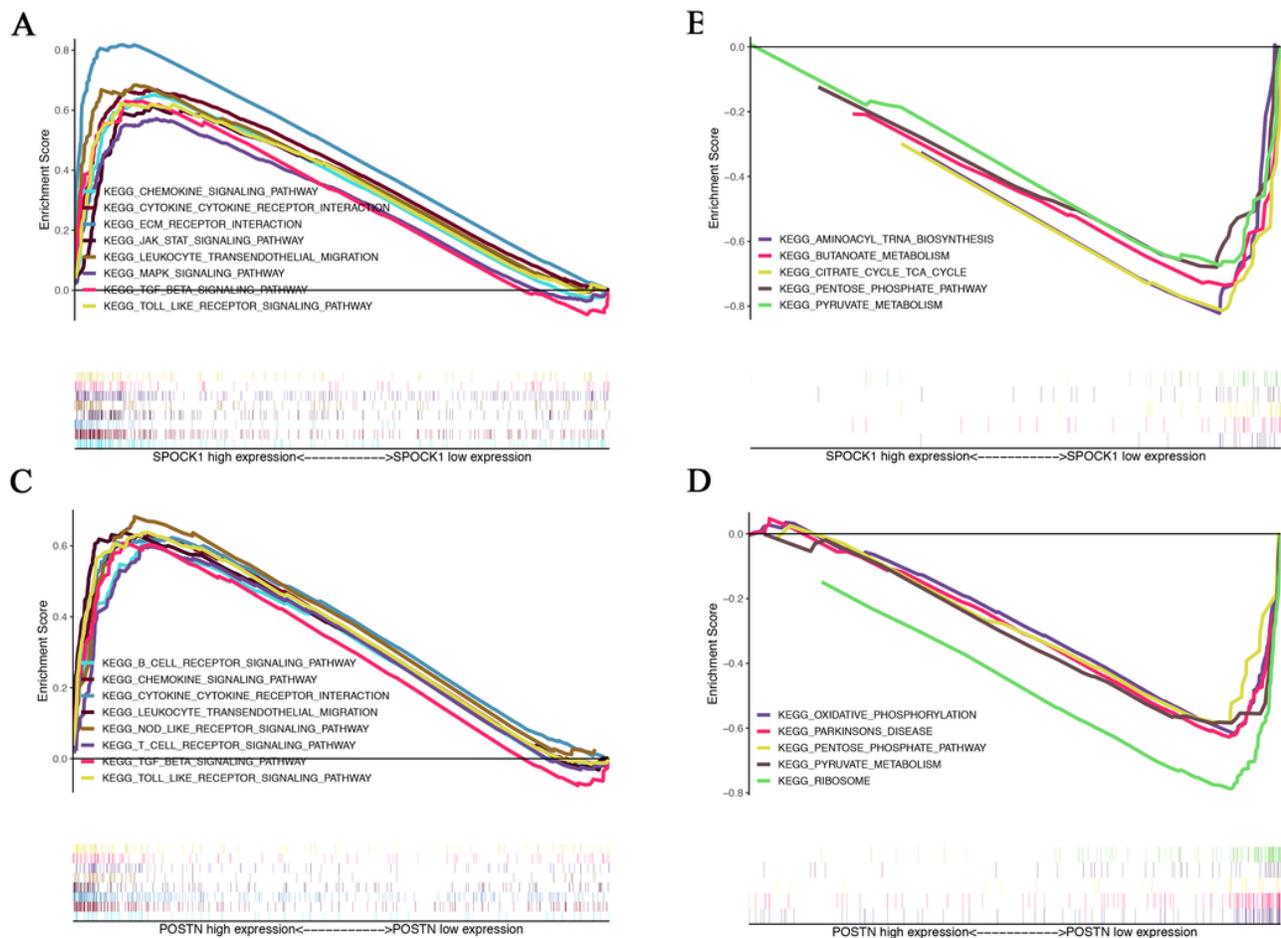


Figure 6

KEGG pathway enrichment analysis of SPOCK1 and POSTN. (A) The enriched gene sets in the SPOCK1 high expression group by GSEA. (B) The enriched gene sets in the SPOCK1 low expression group by GSEA. (C) The enriched gene sets in the POSTN high expression group by GSEA. (D) The enriched gene sets in the POSTN low expression group by GSEA. Only gene sets with NOM $p < 0.05$ and FDR $q < 0.25$ were considered significant. GSEA, gene set enrichment analysis.

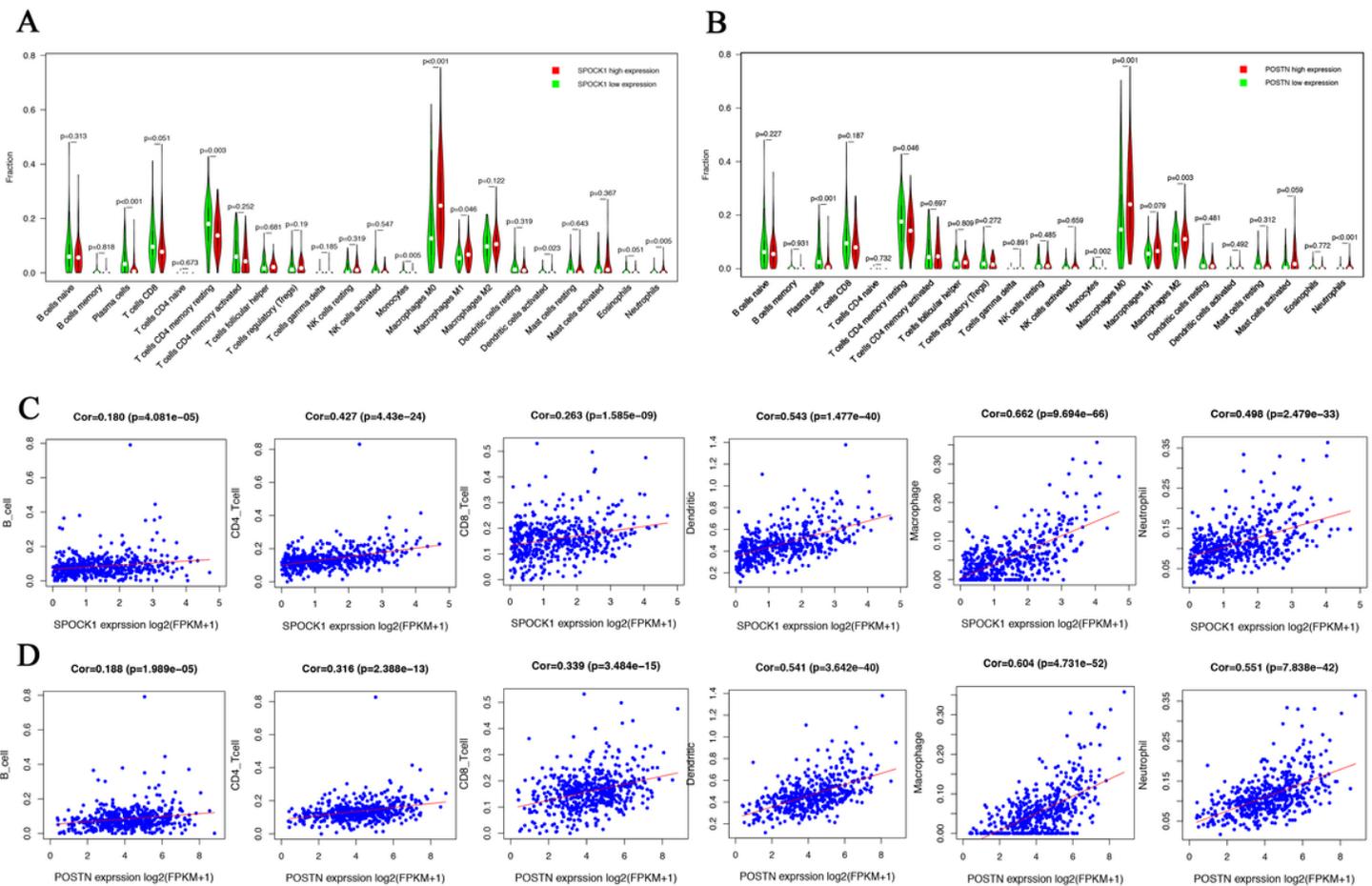


Figure 7

Correlation of infiltration immune cells with BTK and POSTN expression in CRC. (A) The violin plot shows the differences of 22 CIBERSORT immune cells between CRC samples with low or high SPOCK1 expression. (B) The violin plot shows the differences of 22 CIBERSORT immune cells between CRC samples with low or high POSTN expression. (C) The scatter plot shows the correlation of SPOCK1 expression with 6 kinds of immune cells in TIMER. (D) The scatter plot shows the correlation of POSTN expression with 6 kinds of immune cells in TIMER.

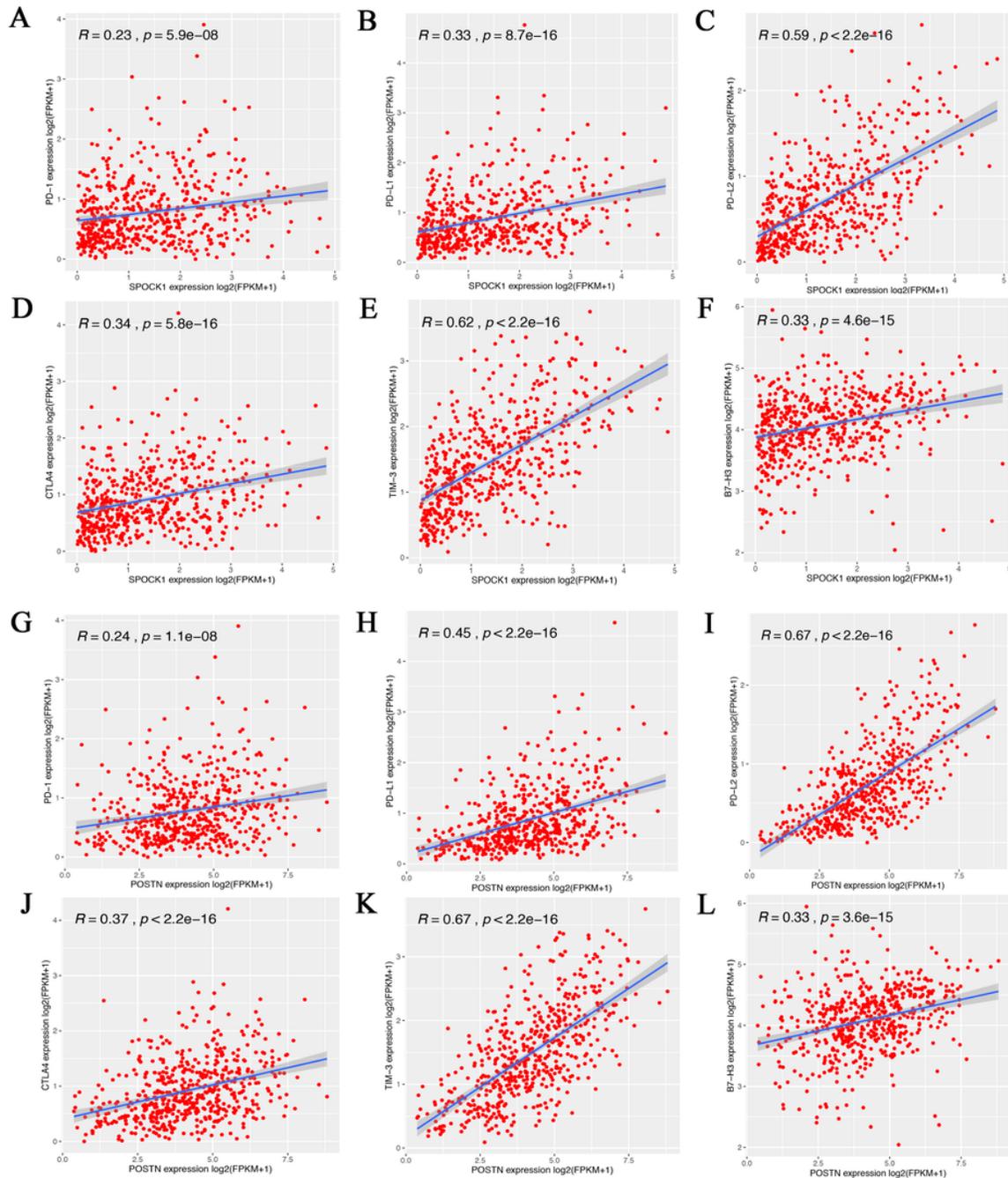


Figure 8

Correlation of immune checkpoints with SPOCK1 and POSTN expression in CRC. (A-F) SPOCK1 expression is positively correlated with the expression of PD-1, PD-L1, PD-L2, CTLA4, TIM-3, and B7-H3. (G-L) POSTN expression is positively correlated with the expression of PD-1, PD-L1, PD-L2, CTLA4, TIM-3, and B7-H3 in TCGA.