Original Article

The early predictive effect of low expression of the *ITGA4* in colorectal cancer

Jiahao Mo[1]^, Jiahui Zhang[1], Haocheng Huang[1], Chang Liu[1], Yi Cheng[2], Yan Mo[1], Shaohua Wu[1], Yao Zhong[3], Cailing Zhong[2], Beiping Zhang[2]

[1]The Second Clinical Medical School, Guangzhou University of Chinese Medicine, Guangzhou, China; [2]Department of Gastroenterology, the Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, China; [3]Department of Traditional Chinese Medicine, School of Traditional Chinese Medicine, Chinese University of Hong Kong, Hong Kong, China

Contributions: (I) Conception and design: J Mo; (II) Administrative support: B Zhang; (III) Provision of study materials or patients: J Zhang, C Zhong; (IV) Collection and assembly of data: J Mo, H Huang, C Liu, Y Cheng; (V) Data analysis and interpretation: J Mo, J Zhang, Y Mo, S Wu, Y Zhong; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Cailing Zhong; Beiping Zhang. Department of Gastroenterology, the Second Affiliated Hospital of Guangzhou University of Chinese Medicine, 111 Dade Road, Yuexiu District, Guangzhou 510120, China. Email: TCMzhongcailing@163.com; doctorzbp@163.com.

Background: The early diagnosis of colorectal cancer (CRC) is very important for the prognosis of patients. It has been suggested that the cytosine-phosphate-guanine (CpG) island of itga4 is highly methylated in colorectal adenoma cell lines AA/C1, Vaco 235 and so on. So the purpose of our study is to explore the diagnostic accuracy and related mechanism of integrin alpha 4 (ITGA4) in early CRC.

Methods: The Cancer Genome Atlas (TCGA) database was used to analyze the relationship between the expression of *ITGA4* and the clinicopathological features and the overall survival rate of the disease. Then, the interaction protein and function enrichment region of *ITGA4* were analyzed. Finally, the infiltration of related immune cells was analyzed.

Results: Compared with normal tissues, the expression of *ITGA4* in colon adenocarcinoma and rectum adenocarcinoma (COAD-READ) tumor tissues was lower ($P<0.05$). The overall survival rate of COAD-READ patients with low *ITGA4* level was lower than that of patients with high *ITGA4* expression ($P<0.05$), and expression of *ITGA4* had a more significant predictive effect in the early stage of tumor development. The results of protein network and enrichment analysis suggested that *ITGA4* was closely related to *ITGB2* and might be involved in the inflammatory reaction and inflammatory tumor transformation process in the carcinogenesis of inflammatory bowel disease (IBD), which was verified by another independent sequence. In terms of immune infiltration, the expression level of *ITGA4* was positively correlated with the infiltration level of intestinal macrophages (Th17), immature dendritic cells (IDC), dendritic cells (DC), mast cells, and eosinophils in COAD-READ, and significantly negatively correlated with $CD56^{bright}$ natural killer (NK) cells.

Conclusions: The low expression of *ITGA4* was related to the poor prognosis of COAD-READ. Findings showed that *ITGA4* might participate in the inflammatory reaction and inflammatory tumor transformation process in the carcinogenesis of IBD, and that *ITGA4* was related to the infiltration of immune cells, macrophages, syndactyls, and $CD56^{bright}$ NK cells. The expression of *ITGA4* could be used as an early predictor of CRC. However, the mechanism of *ITGA4* promoting tumor progression in CRC still needs further research.

Keywords: Colon adenocarcinoma and rectum adenocarcinoma (COAD-READ); integrin alpha 4 (*ITGA4*); survival analysis; immune infiltration

Submitted Dec 17, 2021. Accepted for publication Feb 11, 2022.
doi: 10.21037/jgo-22-92
View this article at: https://dx.doi.org/10.21037/jgo-22-92

^ ORCID: 0000-0002-1869-7841.

Introduction

Colorectal cancer (CRC) is the third largest malignant tumor in the world. In 2018, there were more than 1.8 million newly confirmed cases of CRC globally, resulting in 881,000 deaths, ranking second among all cancers in the world. The incidence of CRC was affected by dietary habits, obesity, and lifestyle factors, and was related to national economic development (1,2). In China, with the rapid growth of the economy and a rising ageing population, CRC incidence rate and mortality rates also increased (1,3,4).

In the last 10 years, studies found that DNA methylation was widely involved in the occurrence and development of tumors (5-8). This was considered to be an important carcinogenic mechanism, therefore it could be used as a potential marker to indicate tumors. The discovery of many biomarkers related to CRC (such as *SEPT9*, *NDRG4*, *BMP3*, etc.) also provided new ideas for non-invasive screening of CRC. However, most studies focus on the predictors of cancer itself, and there are still few studies on the predictors of inflammation cancer transformation (9-11).

DNA methylation refers to the covalent bond of cytosine 5 carbon position of cytosine-phosphate-guanine (CpG) dinucleotide in the genome under the action of DNA methyltransferase, which leads to the changes of chromatin structure, DNA conformation, DNA stability and the interaction mode between DNA and protein, and finally controls gene expression. Therefore, DNA methylation mainly occurred at CpG sites. CpG sites are related to the effect of epigenetic transcriptional silencing of tumor suppressor genes on CpG islands in specific gene promoter regions and may play a very important role in the occurrence and development of CRC (11,12). Integrin alpha 4 (*ITGA4*) is a member of the integrin family. Methylation of this gene was found in a variety of primary tumors and was more common in colorectal tumors (13,14). Previous study found that the CpG island of *ITGA4* was highly methylated in adenoma cell lines AA/C1, Vaco 235, AA/C1/sb10, and Vaco 411 (15). The designed *ITGA4* qmsp detection method could detect a tumor's genomic DNA, and the diagnostic sensitivity for patients with colonic adenoma was 73%. Another study proposed the combined model of *ITGA4*, *SFMBT2*, *THBD*, and *ZNF304*, which improved the sensitivity of CRC screening to 96.1%, with 87% accuracy, and 100% sensitivity for advanced precancerous lesions (5). Attia's research proposed that the micro regulation of the *ITGA4* gene on disease could be better understood by detailed methylation analysis of specific CpG sites of the *ITGA4* gene (13). In addition, *ITGA4* played a role in cell surface adhesion and signal transduction. It was a common therapeutic target for Crohn's disease and inflammatory bowel disease (IBD). Therefore, these results suggested that *ITGA4* may be a potential biomarker of IBD and CRC, which was worthy of further mechanism research in the aspect of inflammation cancer transformation.

Therefore, this study comprehensively analyzed the expression, prognosis, and clinical significance of the *ITGA4* gene in CRC with the help of the Cancer Genome Atlas (TCGA) and string database, combined with data related to the functional pathway and gene set enrichment analysis (GSEA). GSEA is an analysis method for genome-wide expression profile microarray data. Genes are compared with predefined gene sets to build a molecular label database. In this database, known genes are grouped and classified according to multiple functional gene sets such as chromosome location, established gene set, sequence, tumor related gene set and go gene set, Then, by analyzing the gene expression profile data, we can understand their expression status in a specific functional gene set and whether this expression status has some statistical significance. This study also analyzed the immune cell infiltration to explore the immune-related mechanism of *ITGA4* to provide a theoretical basis for the early detection and mechanism of the *ITGA4* gene in CRC. We present the following article in accordance with the REMARK reporting checklist (available at https://jgo.amegroups.com/article/view/10.21037/jgo-22-92/rc).

Methods

Data source

Based on TCGA database (http://portal.gdc.cancer.gov) and the Gene Expression Omnibus (GEO) database, our study intends to analyze and summarize the expression differences of *ITGA4* between CRC individuals and normal individuals through inter group difference analysis of high and low expression groups, subgroup difference analysis of different clinical characteristics and survival analysis. The predicted effect of *ITGA4* is independent of clinical or other factors. TCGA database is a free data portal for a large-scale cancer genome project, which provides scholars and researchers with clinical and pathological information of 33 kinds of cancer. Through TCGA tools and its cancer browser, the expression data and matched

clinicopathological information of RNA corresponding to the target gene (processed by transcriptome sequencing technology) in patients with colon adenocarcinoma and rectum adenocarcinoma (COAD-READ) were obtained. The database was publicly accessible and available, so using data from it did not need to be approved by the local ethics committee. GEO database is a gene expression database created by National Center for Biotechnology Information (NCBI) and contains high-throughput gene expression data submitted by research institutions all over the world. The preliminary results obtained from the data of TCGA database will be verified by two independent data sources in geo database. As one of the largest gene chip databases in the world, GEO database is a comprehensive gene expression database from the NCBI (https://www.ncbi.nlm.nih.gov/geo/). All patients included in the study were from the online database. The inclusion criteria were CRC related studies, and the exclusion criteria were studies with selective reports or incorrect data. The final sample size was determined by the sample data in the online database after screening by inclusion and exclusion criteria. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Expression differential analysis

The microarray data included in the study are from the public online database. Gene chip data are collected, saved, sorted and uploaded in a strict and standardized manner. After converting and sorting out the data, the ggplot2 package of r3.6.3 software (R software is developed by the "R development core team" team of the University of Auckland, New Zealand. Now R has become the most popular language of data science and the most basic tool for data-driven companies such as Google and Facebook) was used to analyze the difference of *ITGA4* gene expression between normal people and COAD-READ patients. In the further clinical data comparison, the differences of gene expression in patients with different genders, different age groups, and different TNM stages were compared. TNM stage is a standard method for staging malignant tumors. In the TNM staging system: (I) T ("t" is the initial letter of the word "tumor") refers to the primary tumor. With the increase of tumor volume and the involvement range of adjacent tissues, it is expressed by T1–T4 in turn. (II) N ("n" is the first letter of the word "node") refers to the involvement of regional lymph nodes. When the lymph node is not involved, it is represented by N0. With the increase of the degree and scope of lymph node involvement, it is successively expressed by N1–N3. (III) m ("m" is the initial letter of the English word "metastasis") refers to distant metastasis (usually blood channel metastasis). Those without distant metastasis are expressed by M0 and those with distant metastasis are expressed by M1. On this basis, a specific stage is drawn with the grouping of the three indicators of TNM. Specifically, we first convert the rnaseq data in fpkm (frequencies per kilobase per million) format on the website into transcripts per million (TPM) reads format. Compared with normal tissues, the expression of *ITGA4* in COAD-READ patients is evaluated by statistical description map, in which the median method of *ITGA4* expression is selected as the cut-off value. Kruskal-Wallis test is used to analyze the expression difference of *ITGA4* in COAD-READ in different clinical features. In order to evaluate the predictive performance of *ITGA4*, after removing duplicate samples, we used the pROC package (version 1.17.0.1) of R software to analyze the included data, and used ggplot2 package (version 3.3.3) to visualize the data, and drew the receiver operating characteristic (ROC) curve by using sensitivity, specificity, positive predictive value and negative predictive value. Then, in order to verify the results of data analysis from TCGA database, we use another two groups of independent sample data from geo database for analysis and verification.

Survival and statistical analysis

In order to explore whether the expression level of *ITGA4* affected the clinical results of patients with COAD-READ, we used the survival package of R 3.6.3 software to statistically analyze the survival data of patients. We used the surviviner package for visualization to construct Kaplan-Meier survival curves to compare the survival differences between patients with high and low expression of *ITGA4*. As for the flow of Kaplan-Meier survival curves, it first calculates the probability that patients who have lived for a certain period will live for the next period (i.e., survival probability), and then multiplies the survival probability one by one, that is, the survival rate of the corresponding period, so as to estimate and depict the birth survival curve. The clinical endpoint of all examinations was survival or death. In addition, subgroup survival data of patients with different stages of TNM were analyzed.

Protein interaction analysis

We sorted the results of the differential analysis of two sets of data with high and low expression of *ITGA4* in section "Expression differential analysis" and extracted the differential genes with log2(fold change) ≥1. We input the results into the string database for analysis in order to obtain the protein-protein interaction (PPI) network of gene expression. (https://string-db.org/). We output the PPI network. The data analysis platform conducted a PPI network analysis of protein interaction, limited the species to *Homo sapiens*, set the confidence to ≥0.99 after pre-reading the data, and hid isolated proteins.

Enrichment analysis of GSEA gene set

GSEA (http://software.broadinstitute.org/gsea/index.jsp) is a computational method that can evaluate whether a given genome showed statistically significant consistent differences between two biological states. In our study, GSEA was used to elucidate the genomic significant differences observed between groups expressing high and low levels of *ITGA4*. We used the clusterprofiler package of R 3.6.3 software for GSEA analysis, set the object type as *Homo sapiens*, and the main reference gene set database in the analysis was msigdb collections (https://www.gsea-msigdb.org/gsea/msigdb). In the study, we assumed that, if the false discovery rate (FDR) <0.25 and the adjusted P value (P.adjust) <0.05, this indicated significant enrichment. Finally, by taking the expression profile of *ITGA4* as a phenotypic marker, we used the P value and normalized enrichment score (NES) to rank the pathways with *ITGA4* enrichment in the phenotype, and showed the top 10 pathways of NES.

Correlation analysis of immune infiltration

We used the GSVA package of R 3.6.3 software for immune infiltration correlation analysis. The specific algorithm was single sample GSEA (SSGSEA), the built-in algorithm of the GSVA package, and Spearman analysis was selected as the correlation analysis method. SSGSEA evaluated the genes in the genome of interest, and then determines whether the gene set was enriched by hypothesis test. GSVA package was an unsupervised classification of samples based on gene expression and multiple pathway information and changes in pathway activity. We analyzed the correlation between the expression of ITGA4 and 24 kinds of infiltrating immune cells. The immune cells included in the analysis were: activated dendritic cells (DC), B cells, CD8 T cells, cytotoxic cells, DC, eosinophils, immature dendritic cells (IDC), macrophages, mast cells, neutrophils, natural killer (NK) $CD56^{bright}$ cells, NK CD56dim cells, NK cells, plasmacytoid DC, T cells, T helper cells, T central memory, T effector memory, T follicular helper, T gamma delta, Th1 cells, Th17 cells, Th2 cells, regulatory T cells.

Results

Patient characteristics

RNA sequencing data and detailed clinical information from 622 CRC samples and 51 normal tissue samples from TCGA were included in our study. After removing duplicate samples, we summarized the clinical information of patients with high and low expression of *ITGA4* in COAD-READ patients in *Table 1*, including age, gender, and pathological stage (T, N or M).

Differential gene analysis results

The expression of *ITGA4* in tumor samples from patients with COAD-READ was significantly lower than that in normal tissues (the median difference between the two groups was −0.801 (−0.996 to 0.605), P<0.001). After differential analysis between the two groups of samples, there were 1,050 differential genes (excluding *ITGA4*), of which 998 were highly expressed and 52 were low expressed (*Figure 1*). The expression of protein products in the former was high, while the latter was on the contrary. When analyzing the correlation between the expression of *ITGA4* and its clinical parameters in patients with COAD-READ, the results showed that there was no significant difference between the messenger RNA (mRNA) level of *ITGA4* and gender (P=0.396), but results showed high expression in the group aged 65 and below, and there was significant difference between the group aged 65 and above (P=0.043). The above results are shown in *Figure 2*.

Then, we validated the results of the differential analysis. We used two independent queues from the GEO database (gse83889: P<0.001; gse23878: P<0.001) (*Figure 3*). In gse83889, there were 35 cases in the normal group and 101 cases in the tumor group. Tumor was lower than the average level of normal. The difference between the two groups was −0.348 (−0.422 to 0.275), and the difference was statistically significant (t=−9.344, P<0.001). In gse23878, there were 24 cases in the normal group and 35 cases in

Table 1 The clinical characteristics of COAD-READ patients in the ITGA4 high and low expression group

Characteristic	Low expression of ITGA4	High expression of ITGA4
N	309	310
T stage, n (%)		
T1	12 (1.9)	8 (1.3)
T2	53 (8.6)	52 (8.4)
T3	206 (33.4)	216 (35.0)
T4	36 (5.8)	34 (5.5)
N stage, n (%)		
N0	176 (28.6)	175 (28.3)
N1	79 (12.8)	71 (11.5)
N2	53 (8.6)	62 (10.0)
M stage, n (%)		
M0	224 (41.0)	235 (43.0)
M1	48 (8.8)	39 (7.1)
Age, n (%)		
≤65 years	127 (20.5)	142 (22.9)
>65 years	182 (29.4)	168 (27.1)
Gender, n (%)		
Female	134 (21.6)	155 (25.0)
Male	175 (28.3)	155 (25.0)
OS event, n (%)		
Alive	241 (38.9)	251 (40.5)
Dead	68 (11.0)	59 (9.5)
DSS event, n (%)		
Alive	249 (41.7)	270 (45.2)
Dead	39 (6.5)	39 (6.5)
PFI event, n (%)		
Alive	226 (36.5)	234 (37.8)
Dead	83 (13.4)	76 (12.3)

COAD-READ, colon adenocarcinoma-rectum adenocarcinoma; ITGA4, integrin alpha 4; OS, overall survival; DSS, disease-specific survival; PFI, progression-free interval.

the tumor group. Tumor was lower than the average level of normal. The difference between the two groups was −1.241 (−1.72 to 0.762), and the difference was statistically significant (t=−5.190, P<0.001). The results of ROC analysis suggest that the prediction ability of variable *ITGA4* has certain accuracy in predicting tumor and normal outcomes [area under the curve (AUC) =0.800, 95% confidence interval (CI) =0.749–0.851], as shown in Figure S1 (the area value under the ROC curve is generally between 0.5 and 1, and the closer the AUC is to 1, the better the diagnostic effect. AUC has lower accuracy when it is 0.5–0.7, and higher accuracy when it is above 0.7).

In subgroup analysis, the differential analysis was conducted between normal people and COAD-READ patients with different stages of TNM. The results showed that there were significant differences in the expression of *ITGA4* between normal people and COAD-READ patients with different stages of TNM. The expression of *ITGA4* was low in the samples of COAD-READ patients, as shown in *Figure 4*.

Survival analysis results

As shown in the Kaplan-Meier diagram, the overall survival rate of COAD-READ cases with low *ITGA4* mRNA expression in the experimental group was low. The median survival time of *ITGA4* low expression group was 61.6 months; the median survival time of high expression group was 101.4 months. The difference in survival between the two groups was statistically significant [hazard ratio (HR) =0.69, P=0.046], as shown in *Figure 5*. In the subgroup analysis of TNM stage, the patients with high expression of *ITGA4* in T1, T2, and T3 had better survival rates, and the difference was statistically significant (P=0.044), but there was no significant difference between the high and low expression groups in T4. Patients with high expression of the *ITGA4* gene in N0 and N1 (P=0.041) and M0 (P=0.028) had better survival rates than patients with low expression, and the difference was statistically significant. However, in patients with more advanced tumors in T4, N2, and M1, there was no significant difference between the high and low expression groups. The above results are shown in *Figure 6*.

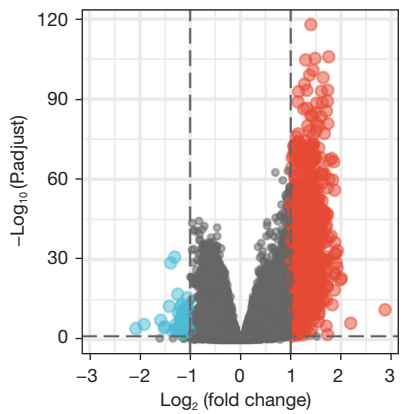

Figure 1 Differential expression map of ITGA4. P.adjust, adjusted P value; ITGA4, integrin alpha 4.

Protein interaction analysis and functional pathway enrichment

The *ITGA4* gene and the differential gene with log2(fold change) absolute value greater than or equal to 1 in the differential expression analysis were introduced into the string 11.0 database for PPI analysis. After pre-reading the data, the confidence was set to ≥0.99, the isolated protein was hidden, and the PPI protein network interaction diagram was output, as shown in *Figure 7*.

As shown in *Figure 8*, we analyzed the function and pathway enrichment of genes closely related to *ITGA4* in the PPI network, sorted them according to the number of enriched genes, and visualized the items with the previous number of genes. The bubble size in the figure represented

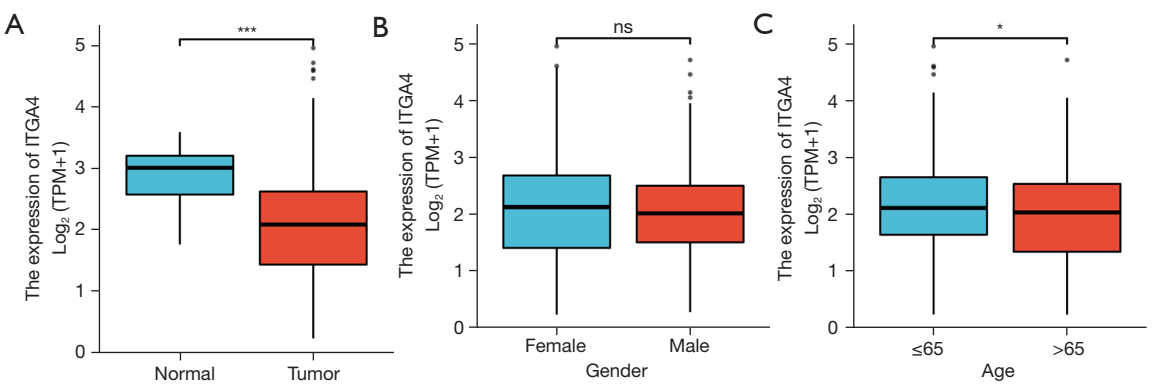

Figure 2 Analysis of the expression difference of ITGA4 between normal subjects and patients with COAD-READ. *, 0.01<P<0.05; ***, P<0.001. ns, no significance; COAD-READ, colon adenocarcinoma-rectum adenocarcinoma; ITGA4, integrin alpha 4; TPM, transcripts per million.

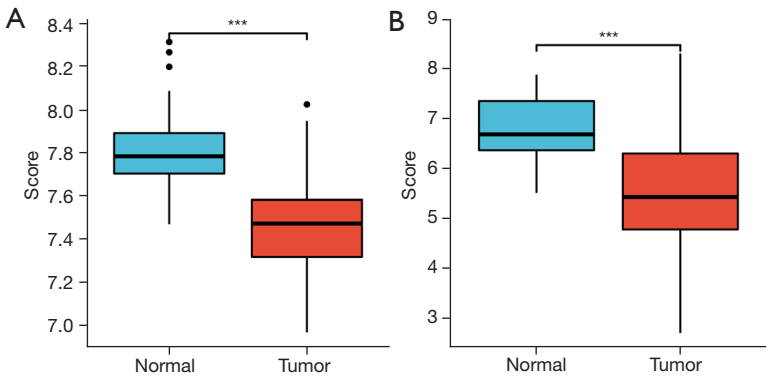

Figure 3 Differential analysis of independent cohort validation of ITGA4. ***, P<0.001. ITGA4, integrin alpha 4.

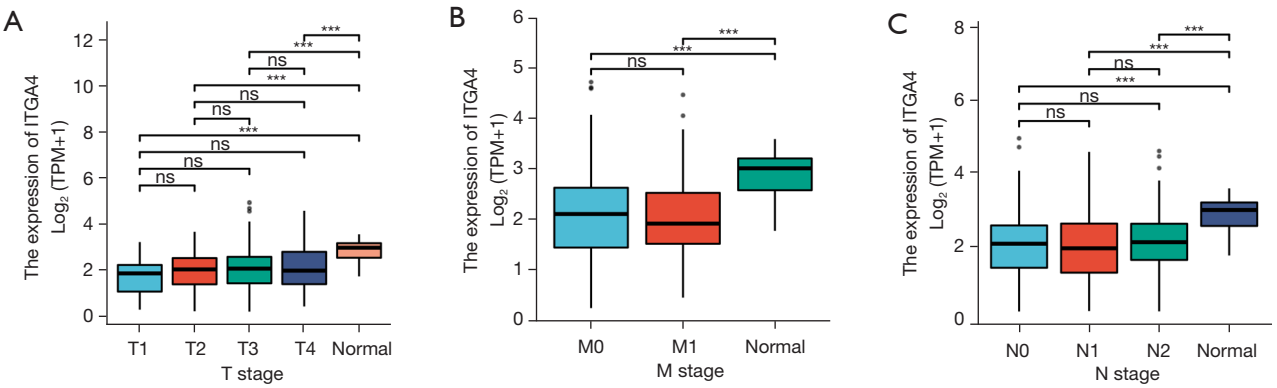

Figure 4 Analysis of TNM stage expression difference of ITGA4 between COAD-READ patients. ***, P<0.001. ns, no significance; COAD-READ, colon adenocarcinoma-rectum adenocarcinoma; ITGA4, integrin alpha 4; TPM, Transcripts per million.

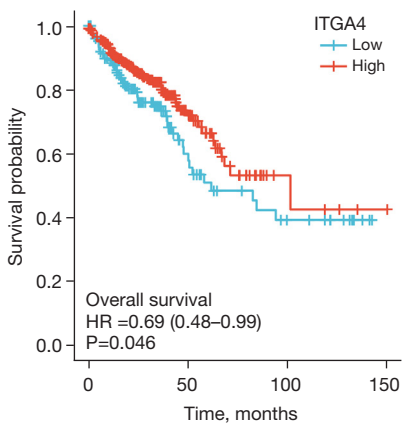

Figure 5 Survival curve of patients with high and low expression levels of ITGA4 and COAD-READ. HR, hazard ratio; COAD-READ, colon adenocarcinoma-rectum adenocarcinoma; ITGA4, integrin alpha 4.

the negative logarithm of corrected P based on 10. The larger the bubble area, the smaller the corresponding corrected P value, the closer the relationship between genes and functional pathways. The results of functional enrichment analysis showed that the corrected P values were small, including leukocyte migration, cell chemotaxis, leukocyte adhesion, cell calcium homeostasis and positive regulation of cytoplasmic calcium concentration, suggesting that the gene *ITGA4* was closely related to these functional pathways. Among the results of pathway enrichment analysis, the corrected P value of the chemokine signaling pathway and cell adhesion molecule pathway was the smallest, which means that it was most closely related to genes.

Enrichment analysis of GSEA gene set

As for the setting of threshold, it was generally considered

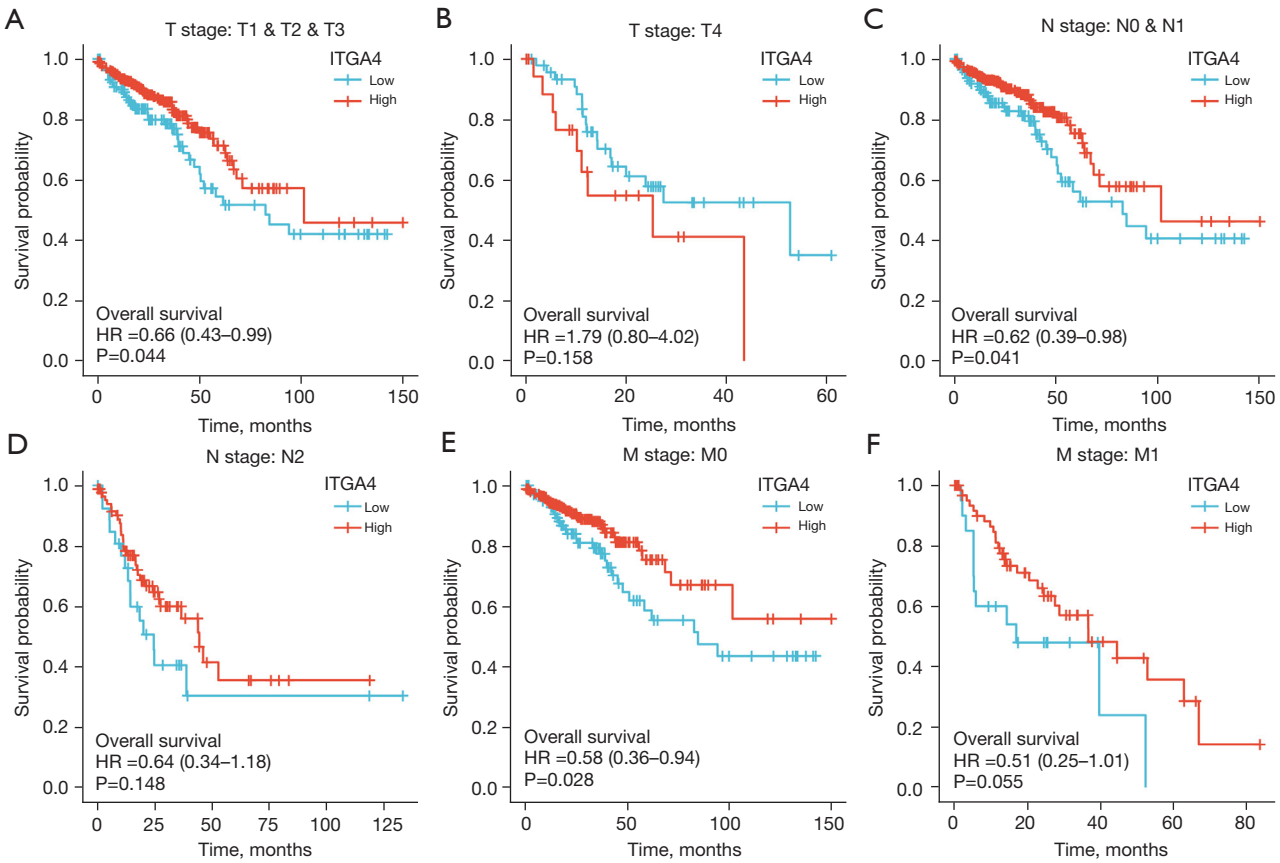

Figure 6 TNM subgroup survival analysis of ITGA4. HR, hazard ratio; ITGA4, integrin alpha 4.

that the gene set under the pathway of |NES| >1, FDR <0.25 and P.adjust <0.05 was meaningful. The greater the absolute value of NES, the smaller the FDR value, indicating the higher the reliability of the analysis results. Therefore, after setting the threshold and sorting, there were 309 data sets that meet FDR <0.25 and P.adjust <0.05. They were sorted in descending order according to the NES value, and the top 10 enrichment pathways were visually displayed, as shown in *Figure 9* and *Table 2*.

Correlation analysis between ITGA4 gene and invasive immune cells

Tumor associated infiltrating lymphocytes affect the survival rate of tumor patients. Therefore, we analyzed the correlation between the expression of *ITGA4* and 24 kinds of infiltrating immune cells. The results showed that *ITGA4* was closely related to IDC (correlation coefficient 0.665, P<0.001), DC (correlation coefficient 0.599, P<0.001), mast cells (correlation coefficient 0.586, P<0.001), eosinophils (correlation coefficient 0.583, P<0.001), and $CD56^{bright}$ NK cells (correlation coefficient –0.146, P<0.001). The first five showed a positive correlation, while last showed a negative correlation. The above results are shown in *Figure 10*.

Confirmatory subgroup inquiry analysis

From the results of section "Patient characteristics" to section "Correlation analysis between ITGA4 gene and invasive immune cells", we concluded that the action scene of *ITGA4* in colorectal tumors was mainly in the inflammatory environment. In order to verify the inflammatory tumor transformation effect of *ITGA4* in colorectal tumors, we did the second validation cohort analysis (gse4183). The cohort included 15 ulcerative colitis samples, 15 colorectal adenoma samples, and 15 CRC samples. Firstly, we did a differential analysis between the sample data of 15 ulcerative colitis and 30 tumor samples, and then made a differential analysis between the sample data of 15 ulcerative colitis and 15 colorectal adenomas

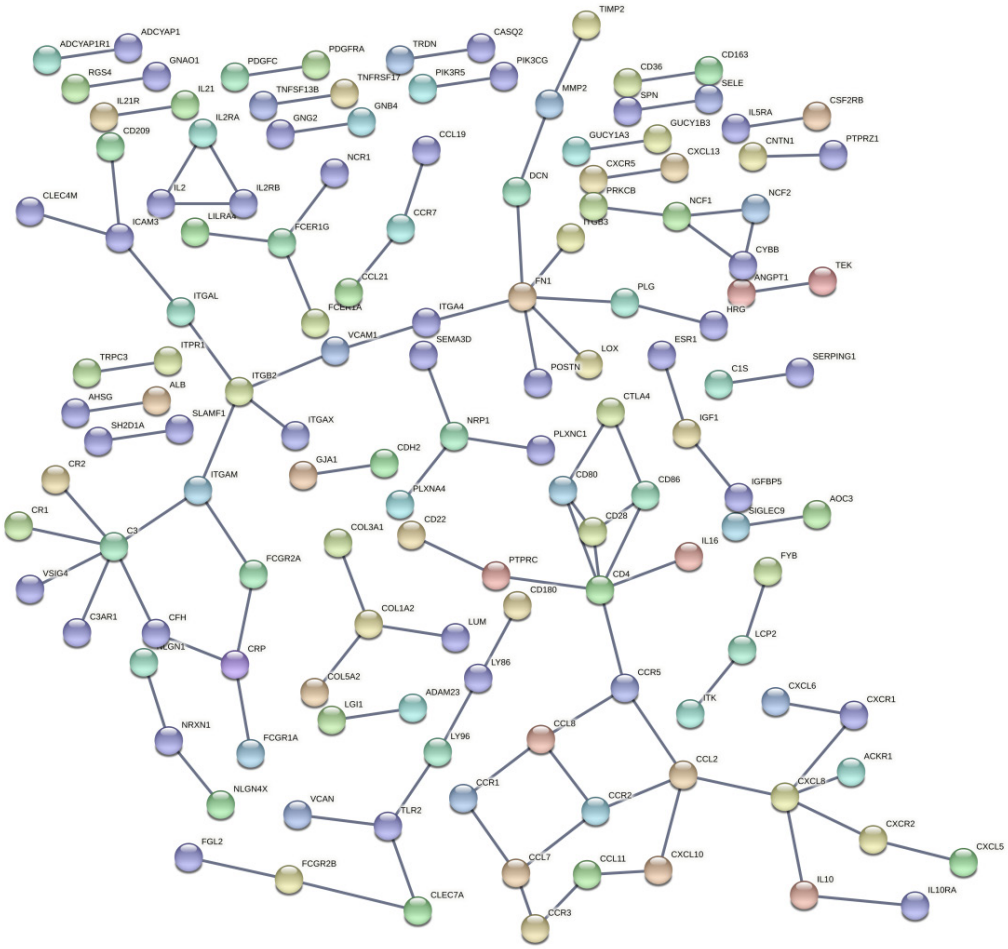

Figure 7 Interacting protein network. The database maps colors to distinguish them according to the score value interacting with them. Specific colors have no specific meaning.

to evaluate whether *ITGA4* played a role in the early inflammatory cancer transformation process. The results are shown in *Figure 11* (*Figure 11A*: ulcerative colitis tumor group, the mean expression was 9.01, P<0.001; *Figure 11B*: ulcerative colitis colorectal adenoma group, the mean expression was 8.93, P<0.001).

Discussion

Integrin is a member of the superfamily of transmembrane glycoproteins α and β heterodimer membrane receptor protein composed of two subunits, in which integrin alpha4 is a member of the integrin family encoded by *ITGA4* gene alpha 4 subunit. *ITGA4* is involved in cell proliferation, apoptosis, adhesion, and migration (16). Many studies have shown that methylated *ITGA4* was not only a valuable biomarker of early COAD-READ, but can also be used to monitor the therapeutic response of metastatic COAD-READ (5,17-19). In addition, compared with tissue biopsy, non-invasive methods such as *ITGA4* methylation detection in plasma and feces could also effectively detect CRC (9,20,21). Previous studies have shown that, after the intervention of *ITGA4* promoter demethylation, the re-expression of the *ITGA4* protein could be observed, indicating that the methylation of *ITGA4* in CRC cells was the main mechanism of epigenetic gene silencing (19). However, there was a lack of research on the correlation between the expression of *ITGA4* and the early prognosis of COAD-READ. Therefore, it was very important to clarify the expression of *ITGA4* in COAD-READ, the

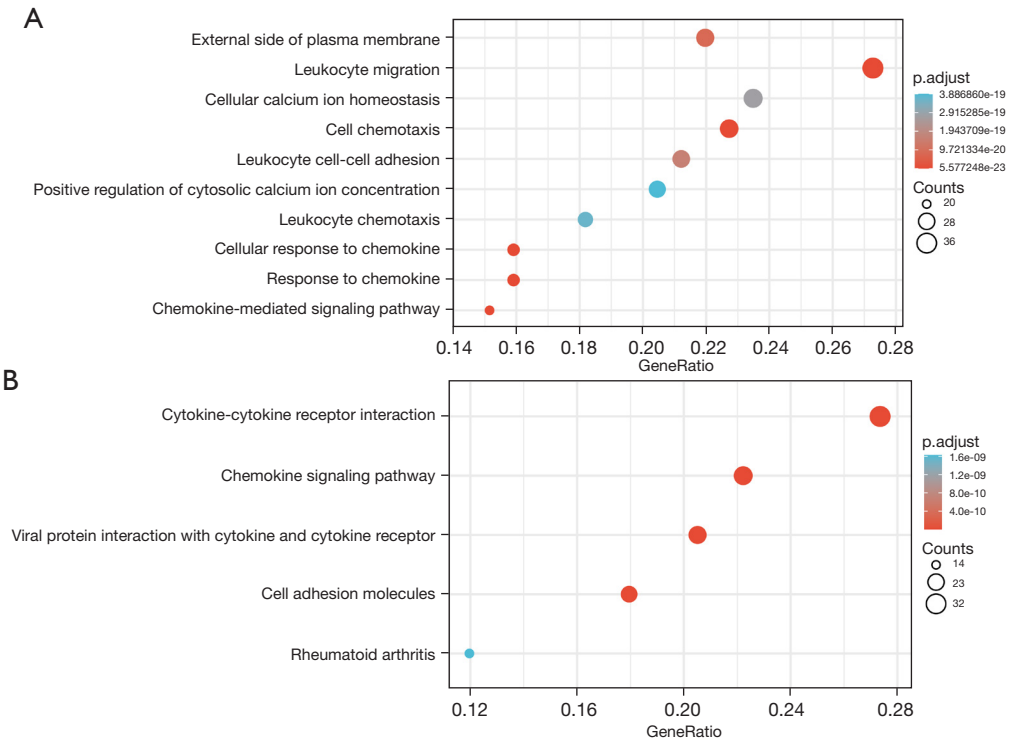

Figure 8 Enrichment analysis of GO (A) and KEGG (B). P.adjust, adjusted P value; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

immune infiltration mechanism regulated by *ITGA4*, and its prognostic value.

First, we studied the expression difference of *ITGA4* between normal sample tissues and COAD-READ sample tissues. The analysis of TCGA high-throughput RNA sequencing data showed that the expression level of *ITGA4* in different TNM stages of COAD-READ tissues was lower than that in normal colorectal tissues. Next, we conducted Kaplan-Meier survival analysis to explore the relationship between *ITGA4* expression and survival in patients with COAD-READ. When *ITGA4* showed low expression in the early and middle stages of COAD-READ, we found that the high expression of *ITGA4* was related to better overall survival. On the contrary, the difference of *ITGA4* expression did not affect the survival time of patients with T4, N2, and M1 tumors. In conclusion, low expression of *ITGA4* was associated with poor prognosis, and *ITGA4* could be used as a good predictor of early COAD-READ prognosis.

Mounting evidence shows that the inflammatory process is involved in all stages of tumor development, and the long-term proctitis process shapes the microenvironment of COAD-READ tumor cells (22). In this study, we found that *ITGA4* and *ITGB2* proteins acted on vascular cell adhesion molecule-1 (*VCAM-1*) through the PPI network. Previous studies have shown that, in COAD-READ, up regulating the expression of *VCAM-1* could promote the adhesion of tumor fibroblasts [cancer-associated fibroblast (CAFs)] to monocytes and the recruitment of macrophages, and the latter two could synergistically inhibit NK cells (23). Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was a database that systematically analyzed gene functions and related pathways. Combined with the gene ontology (GO)/KEGG enrichment analysis results, we found that *ITGA4* was enriched in the chemokine signal pathway and cell adhesion molecule pathway. Its functions were mainly to positively regulate leukocyte migration, cell chemotaxis, leukocyte adhesion, cell calcium homeostasis, and cytoplasmic calcium concentration. *ITGA4* was often methylated in COAD-READ. Therefore, according to the enrichment analysis results, it was speculated that *ITGA4* may not be an upstream gene mainly involved in chronic inflammation in COAD-READ patients. On the contrary, *ITGB2* may mainly act on *VCAM-1* and participate in the formation of an inflammatory microenvironment. Our second validation cohort (gse4183) also confirmed the

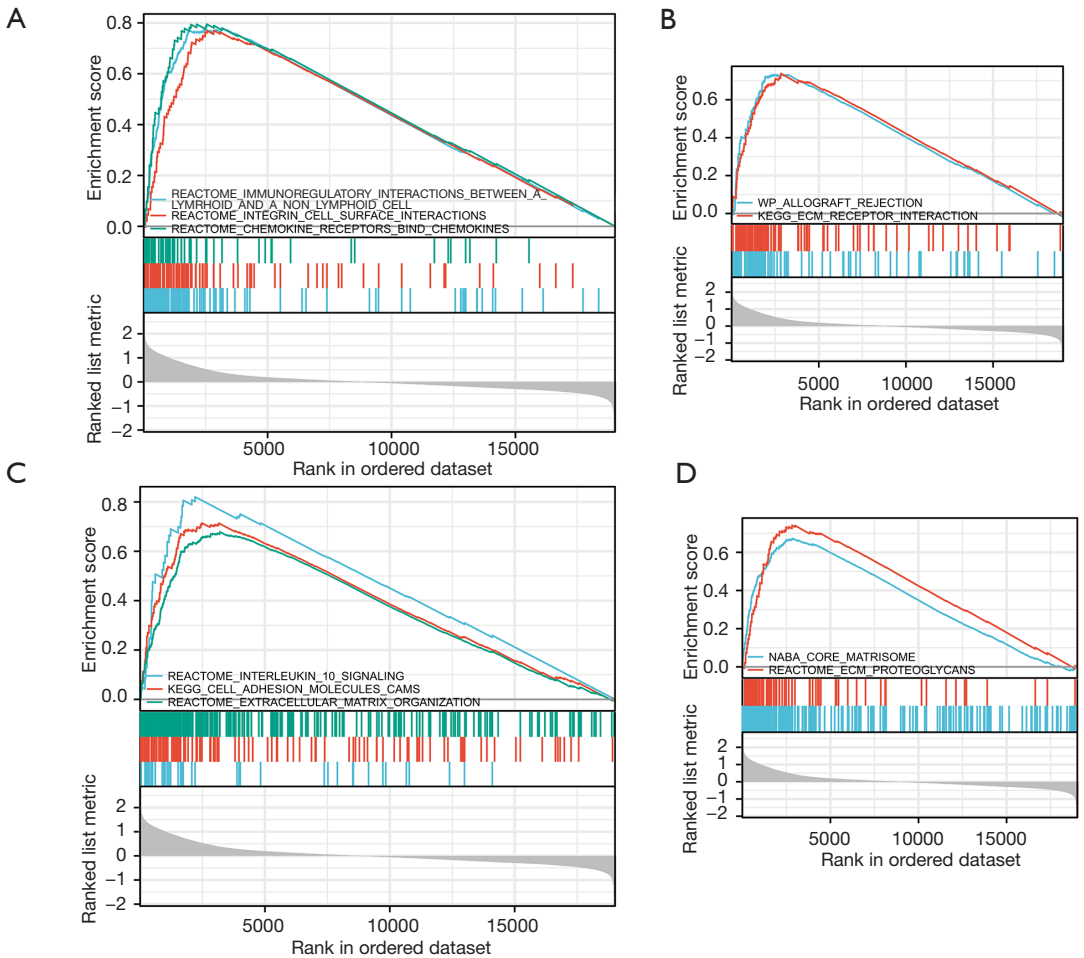

Figure 9 Results of enrichment analysis of GSEA gene set. GSEA, gene set enrichment analysis.

Table 2 Enrichment analysis results of GSEA (top 10 NES)

Description	Set size	NES
REACTOME_IMMUNOREGULATORY_INTERACTIONS_BETWEEN_A_LYMPHOID_AND_A_NON_LYMPHOID_CELL	126	2.68
REACTOME_INTEGRIN_CELL_SURFACE_INTERACTIONS	83	2.53
REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES	57	2.49
REACTOME_INTERLEUKIN_10_SIGNALING	45	2.47
KEGG_CELL_ADHESION_MOLECULES_CAMS	129	2.46
REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION	299	2.45
WP_ALLOGRAFT_REJECTION	89	2.44
KEGG_ECM_RECEPTOR_INTERACTION	82	2.42
NABA_CORE_MATRISOME	274	2.42

GSEA, Gene Set Enrichment Analysis; NES, normalized enrichment score.

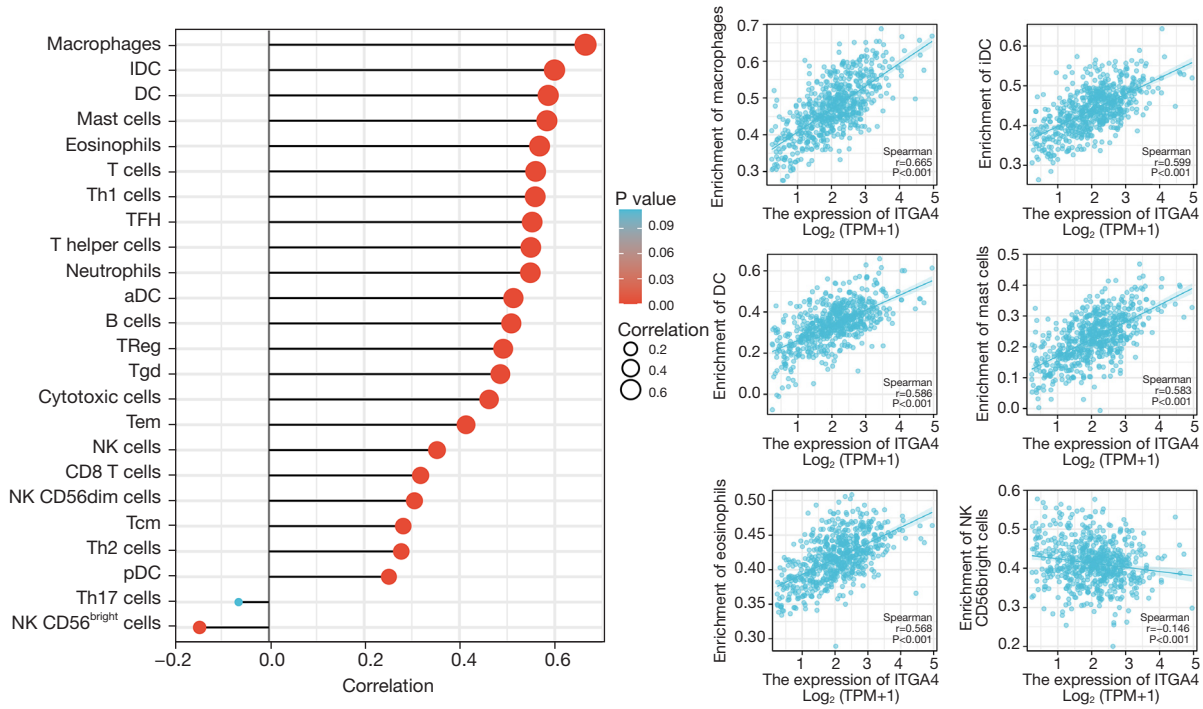

Figure 10 Immune-related molecular diagram of ITGA4. ITGA4, integrin alpha 4; IDC, immature dendritic cells; DC, dendritic cell; aDC, activated dendritic cells; pDC, plasmacytoid dendritic cells; TFH, T follicular helper; TReg, regulatory T cells; Tgd, T gamma delta; NK, natural killer; TPM, transcripts per million.

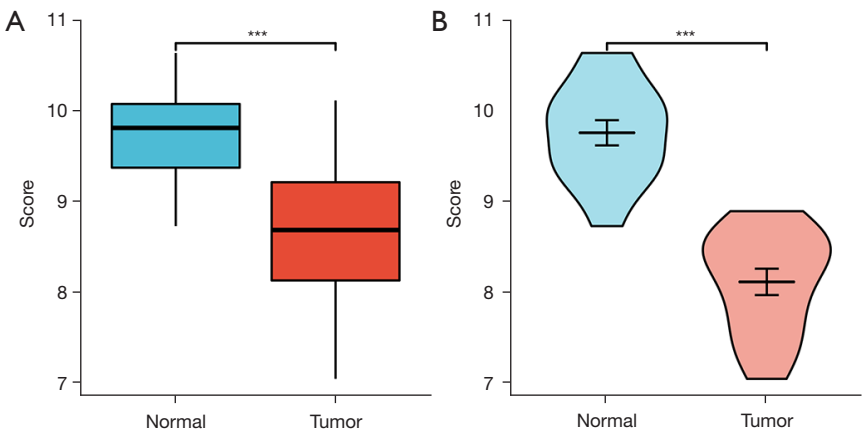

Figure 11 Analysis of inflammation-cancer subgroup of ITGA4. ***, P<0.001. ITGA4, integrin alpha 4.

difference in the expression of *ITGA4* between patients with ulcerative colitis and colorectal tumors, especially in early CRC (colorectal adenoma), which laterally confirmed that *ITGA4* could be a good predictor of the prognosis of early colorectal tumors. Studies have shown that *ITGB2*, which was related to the poor prognosis of COAD-READ patients, was involved in the proliferation, migration, and invasion of tumor cells (24). Interestingly, some studies have found that *ITGA4* was hypermethylated in IBD patients, which may be explained by the long-term high-level oxidative stress in inflammatory areas that induce DNA methylation (17). However, the results of the above studies seemed to be related to vedolizumab by selectively blocking the intestine the interaction between integrin and its natural ligand

mucosal addressin cell adhesion molecule 1 (*MAdCAM-1*) to treat moderate and severe IBD was controversial (25). Therefore, it was still necessary to further study the correlation between methylation frequency of related genes, the inflammatory mechanism, and inflammatory duration in IBD carcinogenesis.

In addition, we found that *ITGA4* may regulate tumor cell survival, proliferation, and apoptosis through the phosphatidylinositol 3-kinase (PI3K)/Akt pathway. PI3K activity was negatively correlated with apoptosis activity. A previous animal experiment on the mechanism of anaerobic promoting COAD-READ showed that the anaerobic surface protein PCWBR2 in the PI3K/Akt pathway could directly interact with the integrin expressed by CRC cells α2/β1 receptor interaction. On one hand, integrin α2/β1 complex could promote the adhesion of anaerobic bacteria in tumor tissue. On the other hand, the PI3K/Akt signaling pathway was activated and tumor cell proliferation was promoted (25). Therefore, we speculated that the low expression of the *ITGA4* protein and intestinal integrin in COAD-READ patients α2/β1 complex acted on the PI3K/Akt pathway to promote tumor cell proliferation and inhibit tumor cell apoptosis.

Conclusions

Through microarray data mining, independent cohort verification, functional pathway enrichment analysis, and immune infiltration correlation analysis of a large public database, this study found that the low expression of *ITGA4* was related to the poor prognosis of colorectal tumors. *ITGA4* could be used as a good predictor of the prognosis of early COAD-READ and may be a key gene in the conversion process of intestinal inflammation to colorectal tumors. However, this study had some limitations. The data set of this study came from multiple online databases, and further follow-up studies cannot be carried out. The conclusions need to be verified by further *in vivo* and *in vitro* experiments.

Acknowledgments

Funding: This work was supported by the double first-class and high-level university discipline collaborative innovation team of Guangzhou University of Traditional Chinese Medicine (2021xk58), and special support for clinical research from Guangdong Hospital of Traditional Chinese Medicine (YN10101914).

Footnote

Reporting Checklist: The authors have completed the REMARK reporting checklist. Available at https://jgo.amegroups.com/article/view/10.21037/jgo-22-92/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://jgo.amegroups.com/article/view/10.21037/jgo-22-92/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
2. Dekker E, Tanis PJ, Vleugels JLA, et al. Colorectal cancer. Lancet 2019;394:1467-80.
3. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66:115-32.
4. Fang EF, Scheibye-Knudsen M, Jahn HJ, et al. A research agenda for aging in China in the 21st century. Ageing Res Rev 2015;24:197-205.
5. Zhang X, Wan S, Yu Y, et al. Identifying potential DNA methylation markers in early-stage colorectal Cancer. Genomics 2020;112:3365-73.
6. Laird PW. The power and the promise of DNA methylation markers. Nat Rev Cancer 2003;3:253-66.
7. Mikeska T, Bock C, Do H, et al. DNA methylation biomarkers in cancer: progress towards clinical

implementation. Expert Rev Mol Diagn 2012;12:473-87.
8. Xie W, Kagiampakis I, Pan L, et al. DNA Methylation Patterns Separate Senescence from Transformation Potential and Indicate Cancer Risk. Cancer Cell 2018;33:309-321.e5.
9. Liu X, Wen J, Li C, et al. High-Yield Methylation Markers for Stool-Based Detection of Colorectal Cancer. Dig Dis Sci 2020;65:1710-9.
10. Draht MXG, Goudkade D, Koch A, et al. Prognostic DNA methylation markers for sporadic colorectal cancer: a systematic review. Clin Epigenetics 2018;10:35.
11. Lam K, Pan K, Linnekamp JF, et al. DNA methylation based biomarkers in colorectal cancer: A systematic review. Biochim Biophys Acta 2016;1866:106-20.
12. Long HK, King HW, Patient RK, et al. Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved. Nucleic Acids Res 2016;44:6693-706.
13. Attia HR, Ibrahim MH, El-Aziz SHA, et al. ITGA4 gene methylation status in chronic lymphocytic leukemia. Future Sci OA 2020;6:FSO583.
14. Yan J, Yang X, Jiao X, et al. Integrative transcriptomic and proteomic analysis reveals CD9/ITGA4/PI3K-Akt axis mediates trabecular meshwork cell apoptosis in human glaucoma. J Cell Mol Med 2020;24:814-29.
15. Ausch C, Kim YH, Tsuchiya KD, et al. Comparative analysis of PCR-based biomarker assay methods for colorectal polyp detection from fecal DNA. Clin Chem 2009;55:1559-63.
16. Zucchetto A, Vaisitti T, Benedetti D, et al. The CD49d/CD29 complex is physically and functionally associated with CD38 in B-cell chronic lymphocytic leukemia cells. Leukemia 2012;26:1301-12.
17. Gerecke C, Scholtka B, Löwenstein Y, et al. Hypermethylation of ITGA4, TFPI2 and VIMENTIN promoters is increased in inflamed colon tissue: putative risk markers for colitis-associated cancer. J Cancer Res Clin Oncol 2015;141:2097-107.
18. Ahmed D, Danielsen SA, Aagesen TH, et al. A tissue-based comparative effectiveness analysis of biomarkers for early detection of colorectal tumors. Clin Transl Gastroenterol 2012;3:e27.
19. Sun X, Chen D, Jin Z, et al. Genome-wide methylation and expression profiling identify methylation-associated genes in colorectal cancer. Epigenomics 2020;12:19-36.
20. Jin H, Pang L, Li H, et al. Value of combined detection of ITGA4 and SFRP2 gene methylation in stool DNA in diagnosis and prognostic evaluation of colorectal tumors. Nan Fang Yi Ke Da Xue Xue Bao 2021;41:891-7.
21. Zhang Y, Wu Q, Xu L, et al. Sensitive detection of colorectal cancer in peripheral blood by a novel methylation assay. Clin Epigenetics 2021;13:90.
22. Schmitt M, Greten FR. The inflammatory pathogenesis of colorectal cancer. Nat Rev Immunol 2021;21:653-67.
23. Zhang R, Qi F, Zhao F, et al. Cancer-associated fibroblasts enhance tumor-associated macrophages enrichment and suppress NK cells function in colorectal cancer. Cell Death Dis 2019;10:273.
24. Sun Z, Dang Q, Liu Z, et al. LINC01272/miR-876/ITGB2 axis facilitates the metastasis of colorectal cancer via epithelial-mesenchymal transition. J Cancer 2021;12:3909-19.
25. Christensen B, Colman RJ, Micic D, et al. Vedolizumab as Induction and Maintenance for Inflammatory Bowel Disease: 12-month Effectiveness and Safety. Inflamm Bowel Dis 2018;24:849-60.

(English Language Editor: C. Mullens)

Cite this article as: Mo J, Zhang J, Huang H, Liu C, Cheng Y, Mo Y, Wu S, Zhong Y, Zhong C, Zhang B. The early predictive effect of low expression of the *ITGA4* in colorectal cancer. J Gastrointest Oncol 2022;13(1):265-278. doi: 10.21037/jgo-22-92

Supplementary

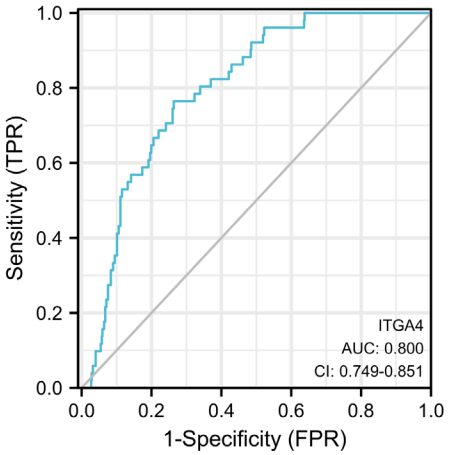

Figure S1 ROC curve of ITGA4 predicting colorectal cancer. AUC, area under the curve; TPR, true positive rate; FPR, false positive rate; ITGA4, integrin alpha 4; ROC, receiver operating characteristic.

https://dx.doi.org/10.21037/jgo-22-92