



Identification of gene biomarkers and immune cell infiltration characteristics in rectal cancer

Lina Wen^{1,2}, Zongqiang Han³, Yanlin Du⁴

¹Department of Clinical Nutrition, Beijing Shijitan Hospital, Capital Medical University, Beijing, China; ²Department of Oncology, Capital Medical University; Beijing Institute of Integrated Chinese and Western Medicine Oncology, Beijing, China; ³Department of Laboratory Medicine, Beijing Xiaotangshan Hospital, Beijing, China; ⁴Department of Oncology, Wangjing Hospital, China Academy of Chinese Medical Sciences, Beijing, China

Contributions: (I) Conception and design: L Wen; (II) Administrative support: Y Du; (III) Provision of study materials or patients: Y Du; (IV) Collection and assembly of data: L Wen, Z Han; (V) Data analysis and interpretation: L Wen, Y Du; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Yanlin Du. No. 6 Central South Road Wangjing, Beijing 100102, China. Email: lyddyl@163.com.

Background: Compared with colon cancer, the increase of morbidity is more significant for rectal cancer. The current study set out to identify novel and critical biomarkers or features that may be used as promising targets for early diagnosis and treatment monitoring of rectal cancer.

Methods: Microarray datasets of rectal cancer with a minimum sample size of 30 and RNA-sequencing datasets of rectal adenocarcinoma (READ) were downloaded from the Gene Expression Omnibus (GEO) database and The Cancer Genome Atlas (TCGA) database. The method of robust rank aggregation was utilized to integrate differentially expressed genes (DEGs). The protein-protein interaction (PPI) network of the DEGs was structured using the STRING platform, and hub genes were identified using the Cytoscape plugin cytoHubba and an UpSet diagram. R software was employed to perform functional enrichment analysis. Receiver operating characteristic (ROC) curves based on the GEO data and Kaplan-Meier curves based on the TCGA data were drawn to assess the diagnostic and prognostic values of the hub genes. Immune cell infiltration analysis was conducted with CIBERSORT, and the diagnostic value and correlations between prognostic genes and infiltrated immune cells were analyzed by principal component analysis (PCA), ROC curves, and correlation scatter plots.

Results: A total of 137 robust DEGs were obtained by integrating datasets in GEO. Twenty-four hub genes, including CHGA, TTR, SAA1, SPP1, MMP1, TGFBI, COL1A1, and PCK1, were identified as a diagnostic gene biomarker group for rectal cancer, and SAA1, SPP1, and SI were identified as potential novel prognostic biomarkers. Functionally, the hub genes were mainly involved in the rectal cancer related interleukin (IL)-17 and proximal tubule bicarbonate reclamation pathways. Twelve sensitive infiltrated immune cells were identified, and were correlated with prognostic genes.

Conclusions: The integrated gene biomarker group combined with immune cell infiltration can effectively indicate rectal cancer.

Keywords: Rectal cancer; diagnosis; prognosis; biomarker; immune cell infiltration

Submitted Apr 22, 2021. Accepted for publication Jun 04, 2021.

doi: 10.21037/jgo-21-255

View this article at: <https://dx.doi.org/10.21037/jgo-21-255>

Introduction

Colorectal cancer (CRC) is one of the highest incidence cancers in the world, and it ranks second in the United States for cancer-related mortality (1). CRC can be divided

into 2 major categories: rectal cancer and colon cancer. In patients with CRC, the rectum is the more frequently affected site compared to colon. In China, the incidence rate is significantly higher in males than females. Common

early detection methods for CRC involve colonoscopic, fecal, and blood sample examinations. According to the tumor size, location, and degree, as well as genetic changes and the patient's health status, CRC treatment can include surgery, chemotherapy, radiotherapy, molecular-targeted therapy, and immunotherapy (2-5).

It is important to note that rectal cancer and colon cancer are not identical, and appropriate treatment based on their own characteristics is of great importance. Rectal cancers account for about 30% of CRCs and the incidence rate increased more significantly compared to colon cancer (6). Early diagnosis or screening may reduce the morbidity of rectal cancer effectively. Currently, the standard treatment plan for locally advanced rectal cancer is neoadjuvant chemoradiotherapy (nCRT) followed by the total mesorectal excision, but there are a lack of specific biomarkers to monitor treatment effect (7,8). Thus, sensitive and reliable biomarkers are necessary to improve the early diagnosis and treatment of the disease.

Biomarkers based on tissue or blood samples may have a certain sensitivity and specificity for early prediction of disease occurrence or treatment response. To date, methods for the early diagnosis and treatment of CRC have been widely researched. For rectal cancer, the biomarkers that have been found, including gene expression profiles, DNA mutation and methylation, single-nucleotide polymorphism, circulating cell-free nucleic acids, microRNAs, long non-coding RNAs, antigens, enzymes, amino acids, lipids, circulating tumor cells (CTCs), tumor immune cell infiltration, and immune inflammatory cytokines (9). However, these biomarkers have not been fully applied in the clinic setting. A possible reason for this is the limitation of biomarker research. Firstly, the expression of a single biomarker can vary greatly depending on race, region, or study cohort (especially sample size), so a single biomarker does not have universal applicability. Secondly, the human body is a life system with multiple synergistic biological functions, and a single biomarker cannot fully reflect the state of the patient's disease or body; therefore, the integration of different biomarkers of the same category or even different biomarker categories is needed. An increasing bank of studies shows that integrated biomarkers are more significant compared with a single biomarker for the diagnosis and treatment of diseases.

Abnormal gene expression may play a key mechanistic role in cancer initiation. The occurrence of many types of cancers, results from the imbalance of expression of multiple genes. Therefore, variance in gene expression can provide

clues about cancer occurrence during the early stage. The screening of differentially expressed genes (DEGs) between cancer patients and healthy individuals can be an effective way to discover early biomarkers. Besides gene regulation, the tumor microenvironment (TME) also performs a vital role in the initiation and progress of cancer. Immune cells are important constituents of the TME. It has been recognized that the immune system plays an irreplaceable role in cancer. Genomic changes lead to the production of tumor antigens, which are recognized by the immune system as non self sources. In this case, cellular immune response is triggered. The immune system plays a role in immune surveillance through the way that immune cells infiltrated into the tumor microenvironment and regulated tumor progression. As one of the characteristics of cancer, immune cell infiltration is a key factor affecting the outcome of cancer immunotherapy. Immunotherapy can also affect the infiltration level of immune cells in tumor tissues. Previous studies have shown that the number of tumor infiltrating lymphocytes (TILs) is closely associated with the prognosis of CRC, for instance, a high density of CD8+ T lymphocytes is related to a better prognosis (10). The cytotoxic lymphocyte infiltration and its related immunogenomic pathways may be effective targets for CRC (11).

Considering the limitations of a single biomarker and the importance of tumor microenvironment, we propose gene biomarkers combined with tumor immune cell infiltration maybe a more effective indicator of disease progression. The reliability of biomarkers can be improved by clinical research based on different countries, multiple centers, and a large sample size. Therefore, in this investigation, rectal cancer microarray datasets with a sample size of more than 30 in the Gene Expression Omnibus (GEO) database were integrated using the robust rank aggregation (RRA) method to obtain the DEGs between normal and tumor tissue samples. The functions of robust DEGs were subsequently annotated through enrichment analyses of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (12,13). Then, the protein-protein interaction (PPI) network of the DEGs was set up to analyze the hub genes, and the survival analysis was also carried out to screen the prognostic genes based on a rectal adenocarcinoma (READ) dataset in The Cancer Genome Atlas (TCGA) database. Furthermore, the CIBERSORT algorithm was adopted to investigate the immune cell infiltration state of rectal cancer tissue samples in comparison with normal ones, in order to find immunology characteristics of rectal cancer. We aim to integrate genetic

and immunological features as precise biomarkers for rectal cancer.

We present the following article in accordance with the REMARK and MDAR reporting checklist (available at <https://dx.doi.org/10.21037/jgo-21-255>).

Methods

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Since this was a retrospective database study based on a public database, approval from the ethics committee was not required.

Data collection and processing

The platform annotation document and matrix files of microarray datasets of rectal cancer were selected and downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) according to the following inclusion criteria: (I) species were limited to *Homo sapiens*; (II) the minimum sample size was 30; (III) the experimental type of expression profiling was array; and (IV) both normal tissue samples and tumor tissue samples were available. Only 2 datasets GSE87211 and GSE90627 were included. There were 160 normal and 203 rectal cancer tissue samples in the GSE87211 dataset, and 96 normal and 32 rectal cancer tissue samples in the GSE90627 dataset. The platforms were GPL13497 and GPL17077, respectively. The DEGs between the normal and rectal cancer tissue samples in each dataset were identified with the limma package in R4.0.2 (R Foundation for Statistical Computing, Vienna, Austria). The cut-off criteria were: $|\log_2 \text{fold change (FC)}| > 1$ with P value < 0.05 . Additionally, the transcriptome profiles and corresponding clinical information of patients with READ were downloaded from the TCGA database (<https://portal.gdc.cancer.gov/>) for a survival analysis, which involved 2 normal and 84 READ tissue samples.

Robust DEG identification

The RRA method was used to integrate the 2 microarray datasets, and the robust DEGs were analyzed with R package Robust Rank Aggreg. Robust DEGs with $|\log_2 \text{FC}| > 1$ and P value < 0.05 were considered to be Significant.

GO and KEGG enrichment analyses

GO enrichment analyses including molecular function

(MF), biological process (BP), cellular component (CC), and KEGG enrichment analyses were further performed, and the results were visualized with R packages including “digest”, “GOplot”, “clusterProfiler”, “org.Hs.eg.db”, “enrichplot”, and “ggplot2”. P value < 0.05 was deemed to have statistical significance.

Hub gene identification in the DEG network

The robust DEGs were put into the STRING database to obtain the PPI network with the confidence score set at 0.4. The results were visualized with the Cytoscape v3.6.0 software, in which the MCODE plugin was applied for identification of the significant modules in the PPI network. In line with previous research (14), the cytoHubba plugin in Cytoscape was used to perform an integrated analysis of multiple algorithms for topological parameters including Maximal Clique Centrality (MCC), Density of Maximum Neighborhood Component (DMNC), Maximum Neighborhood Component (MNC), Degree, Edge Percolated Component (EPC), BottleNeck, EcCentricity, Closeness, Radiality, and Betweenness. Hub genes involved in the regulation of rectal cancer were screened based on the scores of above 10 topological parameters.

Biological value of hub genes

Expression distribution, correlation, and receiver operating characteristic (ROC) curve analyses of the hub genes were performed with R package. The clinical information and corresponding RNA-seq-Fragments Per Kilobase of Exon Per Million Fragments Mapped (FPKM) data of patients with READ in the TCGA were extracted for a survival analysis of the screened hub genes. The survival and survminer packages of R language were used for the analysis, with the level of statistical significance set at $P < 0.05$.

Immune infiltration analysis

The expression matrix of 22 types of immune cells was obtained using the CIBERSORT algorithm. The relative abundance of 22 types of immune cells was extracted for each sample, and the DEGs between normal and rectal cancer tissue samples were filtered by P value < 0.05 using R packages “ggpubr” and “cluster”. Then, the differences between normal and rectal cancer tissue samples were uncovered by performing principal component analysis (PCA). The correlation between each other of infiltrated

immune cells was analyzed using R package “corrplot”.

Association of prognostic hub gene expression with immune cell infiltration

To explore the relationships between prognostic hub genes and infiltrated immune cells, correlation scatter plots were drawn using R packages “ggplot2”, “ggpubr” and “ggpmisc”. The evaluation criteria were as follows: a correlation coefficient $|R|$ between 0.8–1.0 stood for a very strong correlation; 0.6–0.8 stood for a strong correlation; 0.4–0.6 stood for a moderate correlation; 0.2–0.4 stood for a weak correlation; and 0.0–0.2 stood for a very weak or no correlation.

Statistical analysis

The statistical analyses were performed using R 4.0.2 (<https://www.r-project.org/>) and ActivePerl (<https://www.activestate.com/>). Plots were drawn with the R package software. Correlation analyses were also performed using the R package software. All the statistical tests were 2-sided, and $P < 0.05$ was considered statistically significant.

Results

Clinical characteristics

Among the cases in the GSE87211 dataset, there were 248 (68.32%) males and 115 (31.68%) females, with ages ranging from 35.7 to 81.5 years old (62.90 ± 9.69 years old); for 2 cases, the sex and age were not recorded. Among the GSE90277 dataset cases, there were 76 (59.375%) males and 52 (40.625%) females, whose ages ranged from 27 to 79 years old (58.56 ± 12.53 years old). Among the cases from the TCGA database, there were 47 (56.63%) males and 36 (43.37%) females included, and the age range was from 33 to 90 years old (64.18 ± 11.62 years old); for 1 case, the sex and age were not recorded. The cases with incomplete information were excluded.

Screening of DEGs in each dataset

After an integrated analysis of the GSE87211 and GSE90277 microarray data using R language, 256 normal tissue samples and 235 rectal cancer tissue samples were included in this study. After filtering according to the cut-off criteria of $|\log_2 FC| > 1$ and P value < 0.05 , 2,897 DEGs

were identified in the GSE87211 dataset, including 1,398 upregulated and 1,499 downregulated genes, while there were 2,803 DEGs in GSE90277 dataset, including 1,404 upregulated and 1,399 downregulated genes. The distribution of statistical parameters of different genes in the 2 datasets is shown in separate volcano plots (*Figure 1A,B*).

Identification of robust DEGs

The RRA method can be applied for effective integration of different datasets and can minimize bias and errors among them. Through this approach, as screened using the same cut-off criteria as those for DEGs, 137 robust DEGs were distinguished, including 49 upregulated and 88 downregulated genes. The specific information of all robust DEGs is listed in [Table S1](#), and among them the top 20 upregulated and downregulated ones are visualized in the heatmap (*Figure 1C*). A consistent change trend of the same genes in different datasets can be observed.

Functional analysis of robust DEGs

The functions of the screened robust DEGs were further explored through GO enrichment and KEGG pathway enrichment analyses. The GO enrichment results revealed that the robust DEGs were mainly enriched in MFs of receptor ligand activity, signaling receptor activator activity and carbonate dehydratase activity; the most significant BPs were bicarbonate transport, chloride transport and antimicrobial humoral response; the most significant CCs included the apical part of cell, apical plasma membrane and zymogen granule (*Figure 2A*, [Table S2](#)). The KEGG pathway enrichment analysis indicated that the top 5 pathways through which the robust DEGs regulated the rectal cancer were: nitrogen metabolism, proximal tubule bicarbonate reclamation, cytokine-cytokine receptor interaction, pyruvate metabolism, and pancreatic secretion (*Figure 2B*, [Table S3](#)).

Construction of the PPI Network

The PPI interaction network of the robust DEGs was constructed using the STRING database and visualized with Cytoscape 3.6.0. There were 86 nodes and 156 edges, with 34 upregulated and 52 downregulated genes, after the removal of the disconnected nodes (*Figure 3A*). The whole PPI network was modularized into 5 key parts using the MCODE plugin; the scores of modules 1, 2, 3, 4 and

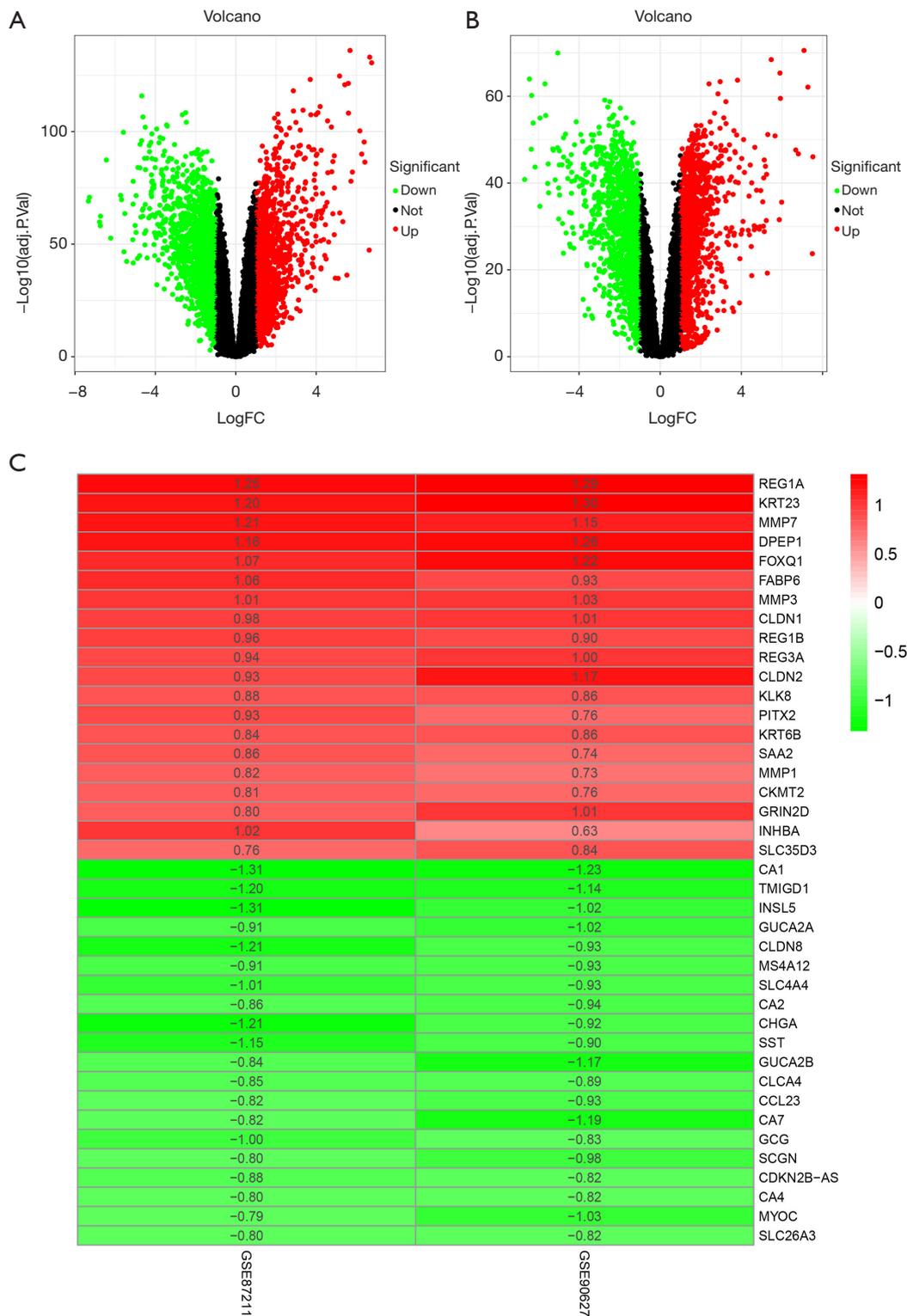


Figure 1 DEGs of selected datasets. (A) Volcano plots of DEGs in GSE87211. (B) Volcano plots of DEGs in GSE90627. (C) Heatmap of the top 20 upregulated and downregulated DEGs. Red represents upregulated DEGs, while green represents downregulated DEGs. DEG, differentially expressed gene.

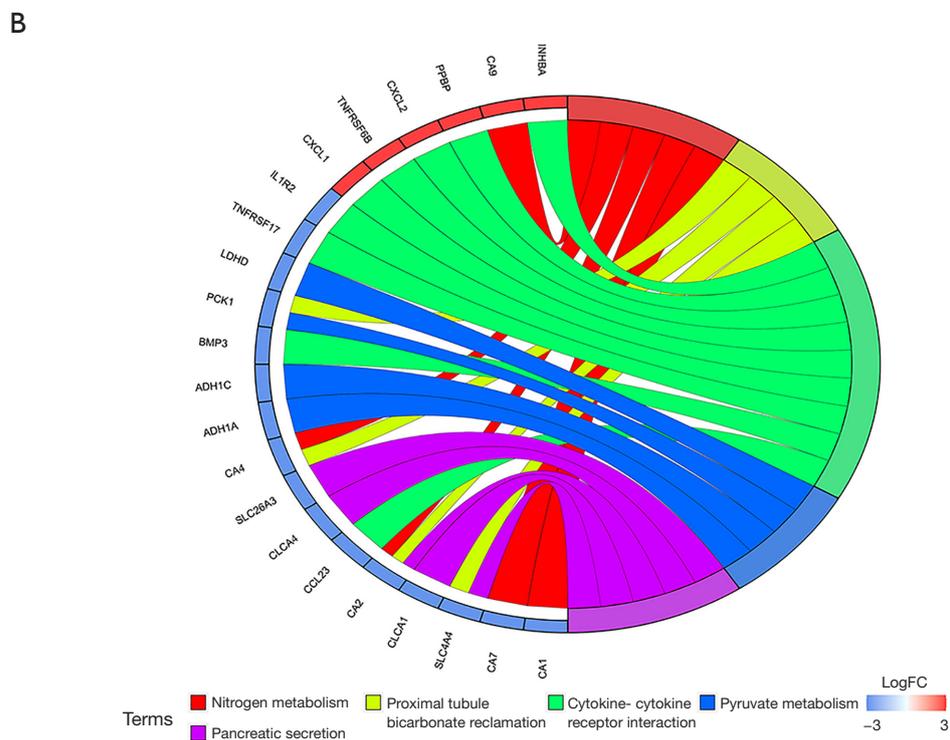
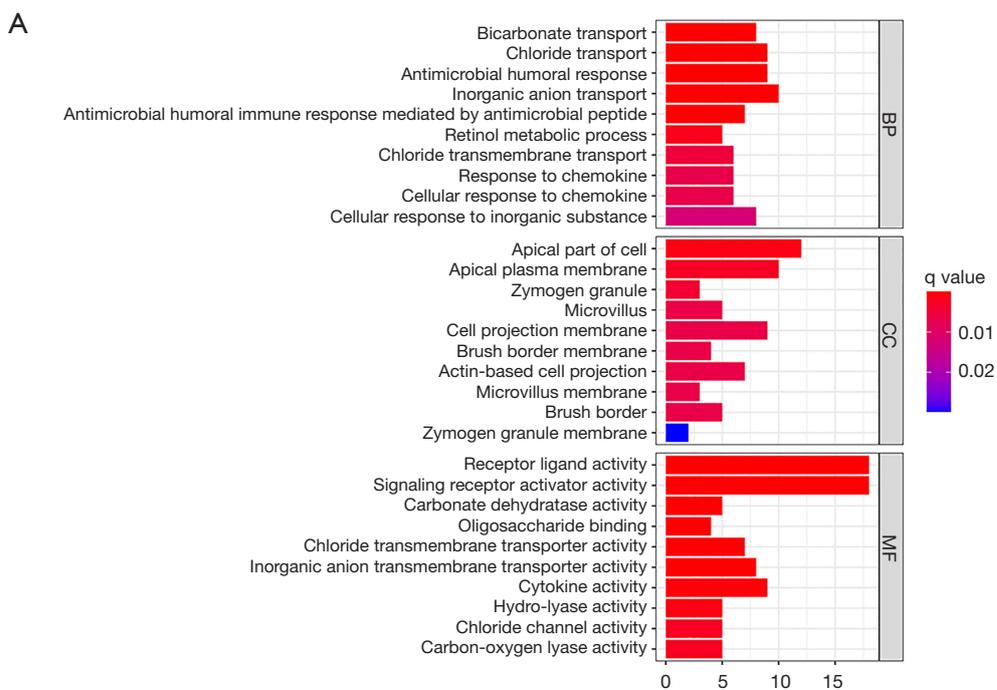


Figure 2 DEG functional enrichment analysis. (A) Bar plot of GO enrichment analysis of DEGs. (B) Circle graph of KEGG pathway analysis of DEGs. DEG, differentially expressed gene; GO, Gene Ontology; BP, biological process; CC, cellular component; MF, molecular function; KEGG, Kyoto Encyclopedia of Genes and Genomes.

5 were 7.429, 6, 4, 3.333 and 3, respectively. According to the enrichment analyses of GO and KEGG (Tables S2,S3), module 1 was mainly related to the pancreatic secretion pathway (Figure 3B); module 2 was mainly related to the cytokine-cytokine receptor interaction pathway (Figure 3C); module 3 was mainly related to the BPs of antimicrobial humoral response and humoral immune response and the MFs of peptidoglycan binding and oligosaccharide binding (Figure 3D); module 4 was mainly related to the BP of the collagen catabolic process and the MFs of metalloproteinase activity and serine-type endopeptidase activity (Figure 3E); and module 5 was related to the BP of calcium-independent cell-cell adhesion via plasma membrane cell-adhesion molecules (Figure 3F).

Hub gene identification

To comprehensively analyze the hub genes in the PPI network, the scores of 10 topological parameters were integrated using the Cytoscape 3.6.0 cytoHubba plugin. The top 50 genes according to each algorithm were ranked, and the intersection of the ranked genes of 10 algorithms were taken and exhibited as UpSet diagram in order to obtain the hub genes. Finally, 24 common genes were identified as hub genes (Figure 3G). The detailed information of the 24 hub genes is given in Table 1.

Diagnostic and prognostic value of hub genes

The expression distribution of the 24 hub genes in normal and rectal cancer tissue samples is illustrated as a boxplot (Figure 4A). Significant differences in these genes existed between the 2 tissue groups. GO terms for the BPs of the 24 hub genes were concentrated in the migration and chemotaxis of leukocytes and neutrophils, while the main rectal cancer-related KEGG pathways were the interleukin (IL)-17 signaling pathway and proximal tubule bicarbonate reclamation (Figure 4B,C, Tables S4,S5). The other functions of the hub genes were similar to those of the robust DEGs.

Correlation analysis revealed that all the hub genes were related to each other (Figure 5A). ROC curve analysis by R package indicated that all 24 hub genes had a high diagnostic value for identifying rectal cancer: GUCA2B (full name of genes were listed in Table 1) area under the curve (AUC) =0.970, GUCA2A AUC =0.965, TMIGD1 AUC =0.966, SLC26A3 AUC =0.945, MS4A12 AUC =0.938, SPP1 AUC =0.910, PPBP AUC =0.911, CLCA4 AUC =0.919, SI AUC

=0.855, CLCA1 AUC =0.860, SST AUC =0.955, CXCL2 AUC =0.936, SAA1 AUC =0.928, CXCL1 AUC =0.948, AQP8 AUC =0.954, MMP1 AUC =0.927, COL1A1 AUC =0.909, SLC4A4 AUC =0.975, SLC30A10 AUC =0.946, PCK1 AUC =0.927, TGFBI AUC =0.955, TTR AUC =0.880, CHGA AUC =0.941, and GCG AUC =0.947 (Figure 5B).

The prognostic value of hub genes was verified using the TCGA database. READ tissue samples in the TCGA dataset were divided into groups with high and low expression of hub genes according to the best-separation cutoff value, Kaplan-Meier (K-M) survival curves, and the correlation between hub genes and overall survival (OS) were analyzed using R package. The hub DEGs SAA1 (P=0.036), SPP1 (P=0.001), SI (P=0.02), CLCA1 (P=0.039), AQP8 (P=0.018), COL1A1 (P=0.025), MMP1 (P=0.016), and TGFBI (P=0.01) were found to be significantly associated with OS in the READ population. However, only the correlations of SAA1, SPP1, SI, and CLCA1 with prognosis were consistent with the clinical significance of gene overexpression (Figure 6); the correlations between AQP8, COL1A1, MMP1, and TGFBI and prognosis were not in accordance with the clinical significance of gene overexpression (Figure S1). The diagnostic and prognostic value of CLCA1 in colon cancer and rectal cancer has been reported previously.

Analysis of immune cell infiltration

The 22 types of infiltrated immune cells in normal tissues and rectal cancer tissues were analyzed using the CIBERSORT algorithm. The distribution of the 22 types of immune cells in each sample showed there to be immunological differences between normal and rectal cancer tissue samples (Figure 7A). The abundance of immune cells in each sample is illustrated as a heatmap in Figure 7B, and the violin plot in Figure 7C visualizes the differences in each type of immune cell between the 2 groups. Compared with normal tissue samples, rectal cancer tissue samples displayed significantly increased abundance of T cells CD4 naive, T cells CD4 memory activated, natural killer (NK) cells resting, monocytes, macrophages M0, dendritic cells activated, mast cells activated, and neutrophils. Immune cells with significantly reduced abundance levels included B cells memory, plasma cells, T cells CD8, T cells CD4 memory resting, T cells follicular helper, macrophages M2, and mast cells resting. Immune cells such as B cells naive, T cells regulatory (Tregs), T cells gamma delta, NK cells activated, macrophages M1, dendritic cells resting, and eosinophils showed no difference between the 2 groups.

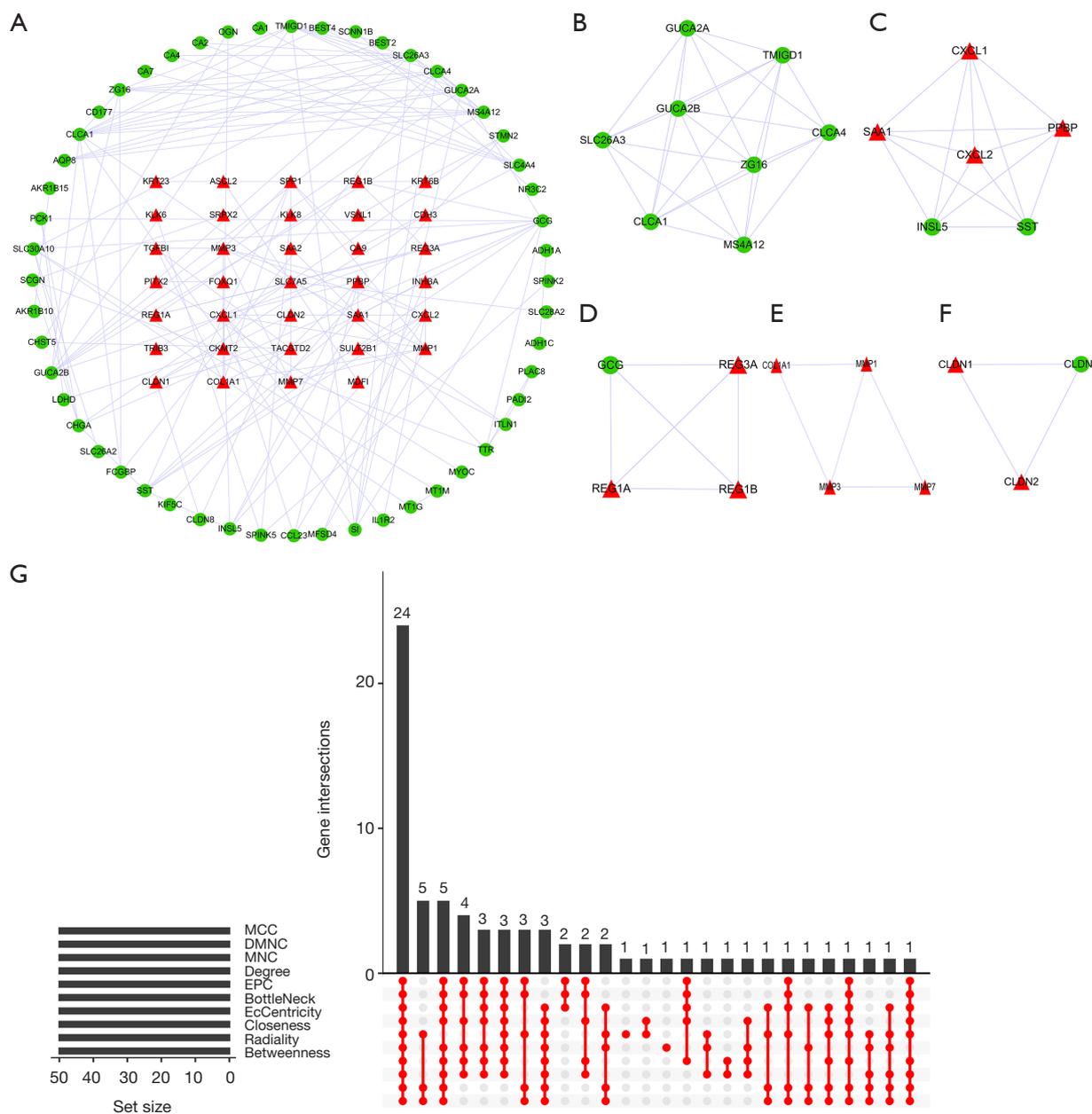


Figure 3 PPI network and hub genes. (A) PPI network of the DEGs. PPI network was divided into different modules: (B) module 1, (C) module 2, (D) module 3, (E) module 4, (F) module 5. (G) Hub genes were identified by intersecting the top 50 genes of every algorithm. Red triangles represent upregulated genes, while green circles represent downregulated genes. PPI, protein-protein interaction; DEG, differentially expressed gene.

The 15 differentially expressed immune cells may be closely related to rectal cancer occurrence and development. The PCA results of immune cell infiltration implied that there were differences between the normal and rectal cancer tissue groups, although in some individual samples, no

difference was found between the 2 groups (Figure 7D).

Correlation analysis showed that 12 of the 15 infiltrated immune cells were related to each other, with the exception of T cells follicular helper, NK cells resting, and dendritic cells activated (Figure 8A). The sensitivity and specificity of these 15 differential

Table 1 Description of the 24 hub genes

Gene	Full name	Change
<i>GUCA2B</i>	Guanylate cyclase activator 2B	Down
<i>TMIGD1</i>	Transmembrane and immunoglobulin domain-containing protein 1	Down
<i>SLC26A3</i>	Chloride anion exchanger	Down
<i>MS4A12</i>	Membrane-spanning 4-domains subfamily A member 12	Down
<i>SPP1</i>	Osteopontin	Up
<i>PPBP</i>	Platelet basic protein	Up
<i>GUCA2A</i>	Guanylin	Down
<i>CLCA4</i>	Calcium-activated chloride channel regulator 4	Down
<i>SI</i>	Sucrase-isomaltase, intestinal	Down
<i>CLCA1</i>	Calcium-activated chloride channel regulator 1	Down
<i>SST</i>	Somatostatin	Down
<i>CXCL2</i>	C-X-C motif chemokine 2	Up
<i>SAA1</i>	Serum amyloid A-1 protein	Up
<i>CXCL1</i>	Growth-regulated alpha protein	Up
<i>AQP8</i>	Aquaporin-8	Down
<i>MMP1</i>	Interstitial collagenase	Up
<i>COL1A1</i>	Collagen alpha-1(I) chain	Up
<i>SLC4A4</i>	Electrogenic sodium bicarbonate cotransporter 1	Down
<i>SLC30A10</i>	Zinc transporter 10	Down
<i>PCK1</i>	Phosphoenolpyruvate carboxykinase, cytosolic (GTP)	Down
<i>TGFB1</i>	Transforming growth factor-beta-induced protein ig-h3	Up
<i>TTR</i>	Transthyretin	Down
<i>CHGA</i>	Chromogranin-A	Down
<i>GCG</i>	Glucagon	Down

immune cells for the diagnosis of rectal cancer were further verified by ROC curve analysis (*Figure 8B*). The results showed that except for T cells follicular helper (AUC =0.422), NK cells resting (AUC =0.551), and dendritic cells activated (AUC =0.621), the other 12 types of infiltrated immune cells were all sensitive for the identification of rectal cancer, with AUCs >0.65. Therefore, the characteristics of immune cell infiltration may be an effective indicator of rectal cancer.

Relationship between prognostic genes and immune cell infiltration

The relationship between prognostic genes and immune

infiltration in rectal cancer was further analyzed. According to relationship evaluation criteria, SAA1 exhibited a strong correlation with macrophages M0, mast cells resting, and macrophages M2; a moderate correlation with T cells CD4 memory activated, plasma cells, T cells CD4 memory resting, neutrophils, mast cells activated, T cells CD4 naive, and T cells CD8; and a weak correlation with B cells memory and monocytes. SPP1 exhibited a strong correlation with macrophages M0 and M2, mast cells resting, and mast cells activated; a moderate correlation with T cells CD4 memory resting, neutrophils, plasma cells, T cells CD4 naive, T cells CD8, and B cells memory; and a weak correlation with T cells CD4 memory activated and monocytes. SI exhibited a

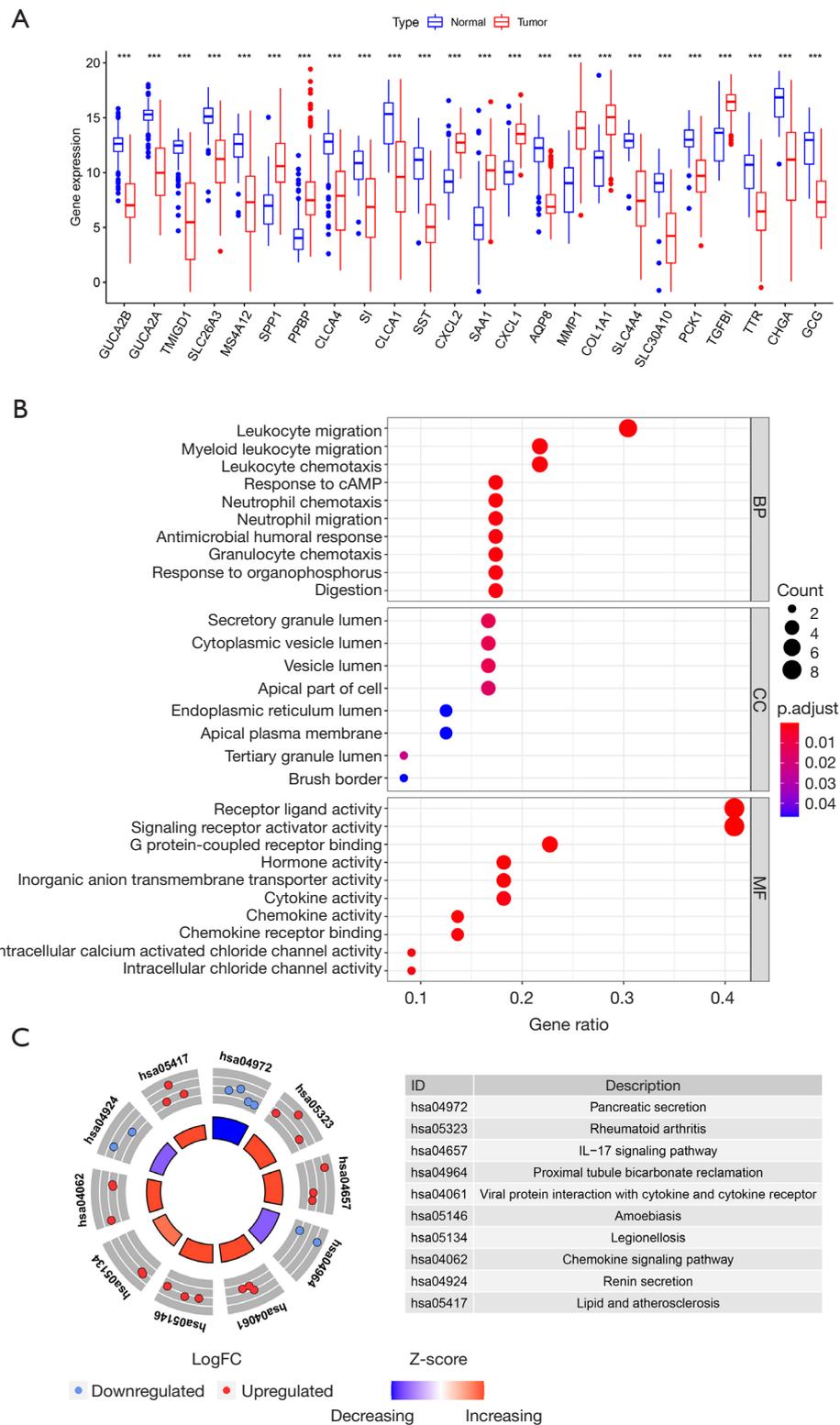


Figure 4 Expression and function of hub genes. (A) Expression of 24 hub genes. (B) Bubble chart of GO enrichment analysis. (C) Circle graph of KEGG pathway enrichment analysis. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

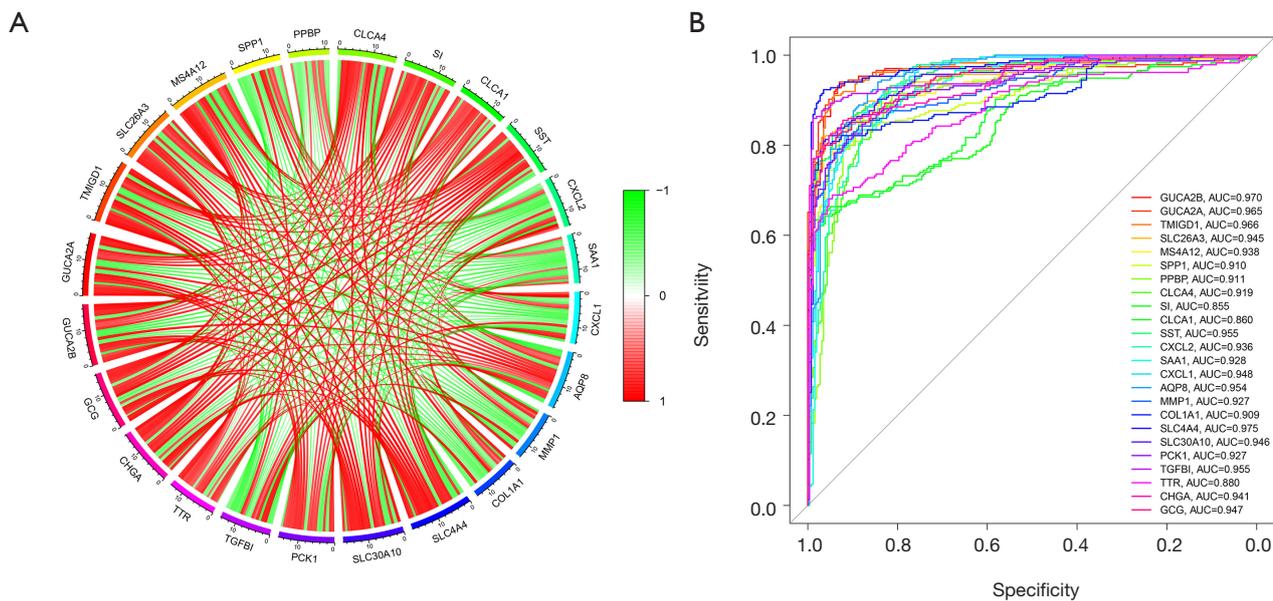


Figure 5 Evaluation of diagnostic value of hub genes. (A) Correlations exist between each other of the 24 hub genes. (B) ROC curves of the 24 hub genes. ROC, receiver operating characteristic.

strong correlation with macrophages M2, neutrophils, and plasma cells; a moderate correlation with mast cells resting, T cells CD4 memory resting, B cells memory, macrophages M0, T cells CD4 naive, mast cells activated, and T cells CD8; and a weak correlation with T cells CD4 memory activated and monocytes (Figure 9). CLCA1 also exhibited correlations with the 12 types of infiltrated immune cells (Figure S2).

Discussion

Due to the complexity of pathogenesis and heterogeneity of cancer, the use of a single biomarker is not reliable or suitable for diagnosing or prognosticating the disease. It is necessary to integrate multiple biomarkers or assessment indices to obtain a more accurate indicator of cancer. Open public data platforms such as GEO and TCGA provide an opportunity for the discovery and validation of tumor biomarkers. For instance, a study in China showed that tissue expression of myc proto-oncogene protein (MYC), proliferating cell nuclear antigen (PCNA), and Metalloproteinase inhibitor 1 (TIMP1) protein combining with MRI-detected extramural vascular invasion could provide additional prognostic details for the preoperative treatment of rectal cancer (7). Another study, also from China, integrated 3 GEO microarray datasets to analyze gene biomarkers of CRC. In the present work, the included

2 microarray datasets of rectal cancer from different countries, each of which had a minimum sample size of 32, which may be helpful in reducing the deviation caused by racial, regional, and individual differences in biomarker research. The same upward or downward trend for each gene in the 2 datasets suggested the reliability of the results.

The 137 robust DEGs including 49 upregulated genes and 88 downregulated genes were identified using a standard and robust RRA method, and 24 hub genes were further obtained. Analyses of GO enrichment and KEGG pathway indicated that the DEGs were mainly involved in bicarbonate transport and material metabolism-related BPs, MFs and signaling pathways. Besides, hub genes mainly took part in proximal tubule bicarbonate reclamation and the IL-17 signaling pathway. Previous studies have shown that the transport of bicarbonate plays an important role in the diagnosis and treatment of multiple cancers, and the expression levels of bicarbonate transporter in patients with colon cancer have changed widely compared with that in healthy bodies. The results of this work confirm that bicarbonate transport also plays an important role in rectal cancer.

In the PPI network, 24 hub genes were discovered through the integration of 10 algorithms to improve the accuracy. GUCA2B binding with GUCA2A can trigger the activation of a transmembrane receptor expressed on intestinal epithelial cells, guanylyl cyclase C (GUCY2C),

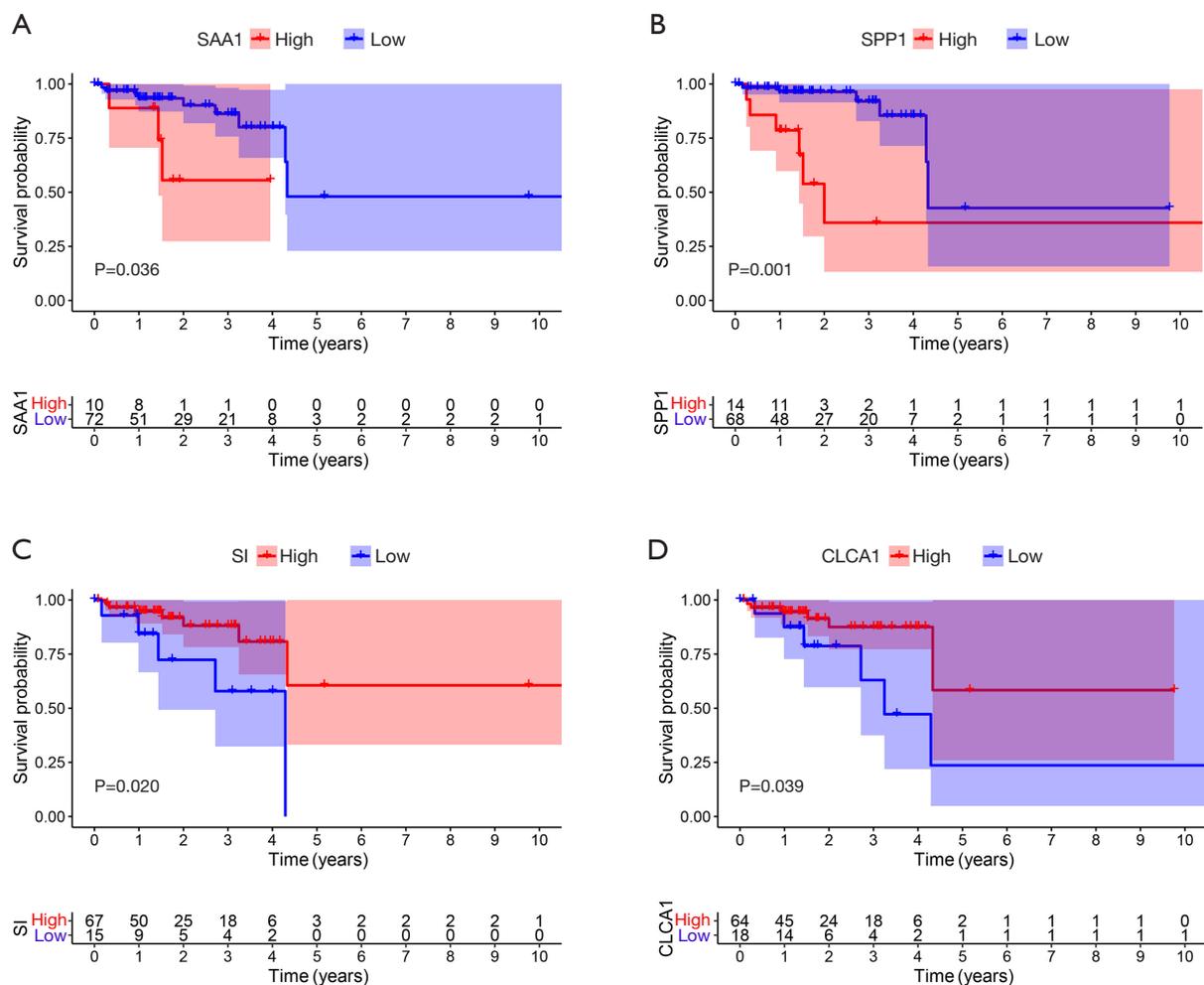


Figure 6 Survival analysis of the 24 hub genes. Gene changes of *SAA1* (A), *SPP1* (B), *SI* (C), and *CLCA1* (D) were obviously associated with the overall survival of patients with rectal cancer ($P < 0.05$).

which participates in regulatory mechanism of intestinal homeostasis, and its deletion may be related to the occurrence of CRC (15,16). Downregulation of *GUCA2B* and *GUCA2A* confirmed their important roles in rectal cancer development. In renal cancer, *TMIGD1* is a tumor suppressor via regulation of p21Cip1/p27Kip1, but its relationship with other cancers is unclear (17). Attention has increasingly been paid to the role of plasma membrane transporters in cancer; for instance, the representative *SLC26A3* is a tumor suppressor for colon cancer. As a Cl^-/HCO_3^- exchanger, *SLC26A3* can promote the outflow of HCO_3^- and possibly play an anti-tumor effect through the regulation of intracellular pH (18). As a member of the membrane-spanning 4-domains subfamily, *MS4A12* is specifically expressed in colon epithelium, which can affect

the colon cancer cell proliferation and cell cycle, and is a potential target for colon cancer immunotherapy (19). The calcium-activated chloride channel protein *CLCA4* and the cyclic tetradeca peptide hormone *SST* can inhibit the proliferation and invasion of CRCs (20,21). *SI* deficiency can lead to dyspepsia, but its correlation with CRC is unclear (22). The calcium-activated chloride channel protein *CLCA1* has been proposed as a diagnostic and prognostic biomarker for both colon and rectal cancer (23). Over expression of *AQP8*, a member of aquaporins family, can inhibit the proliferation and invasion of colon cancer cells (24). The decreased expression of *CHGA* in the early stage of colon cancer may be a novel biomarker for colon cancer diagnosis (25). Downregulated *SLC4A4* can suppress progression of CRC (26). The relationship between *TTR*,

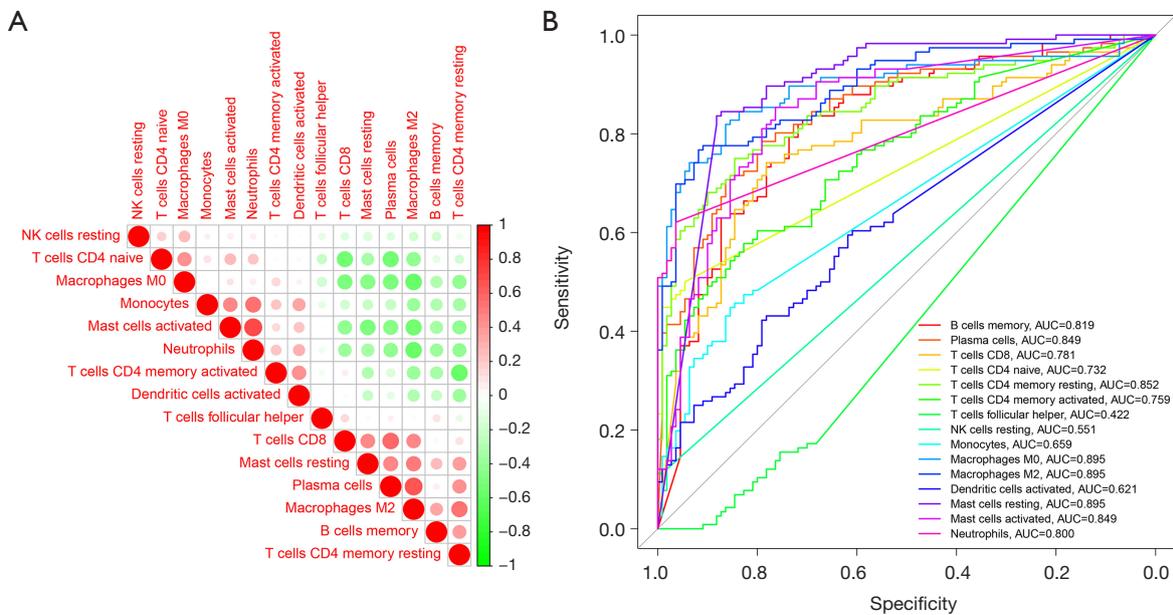
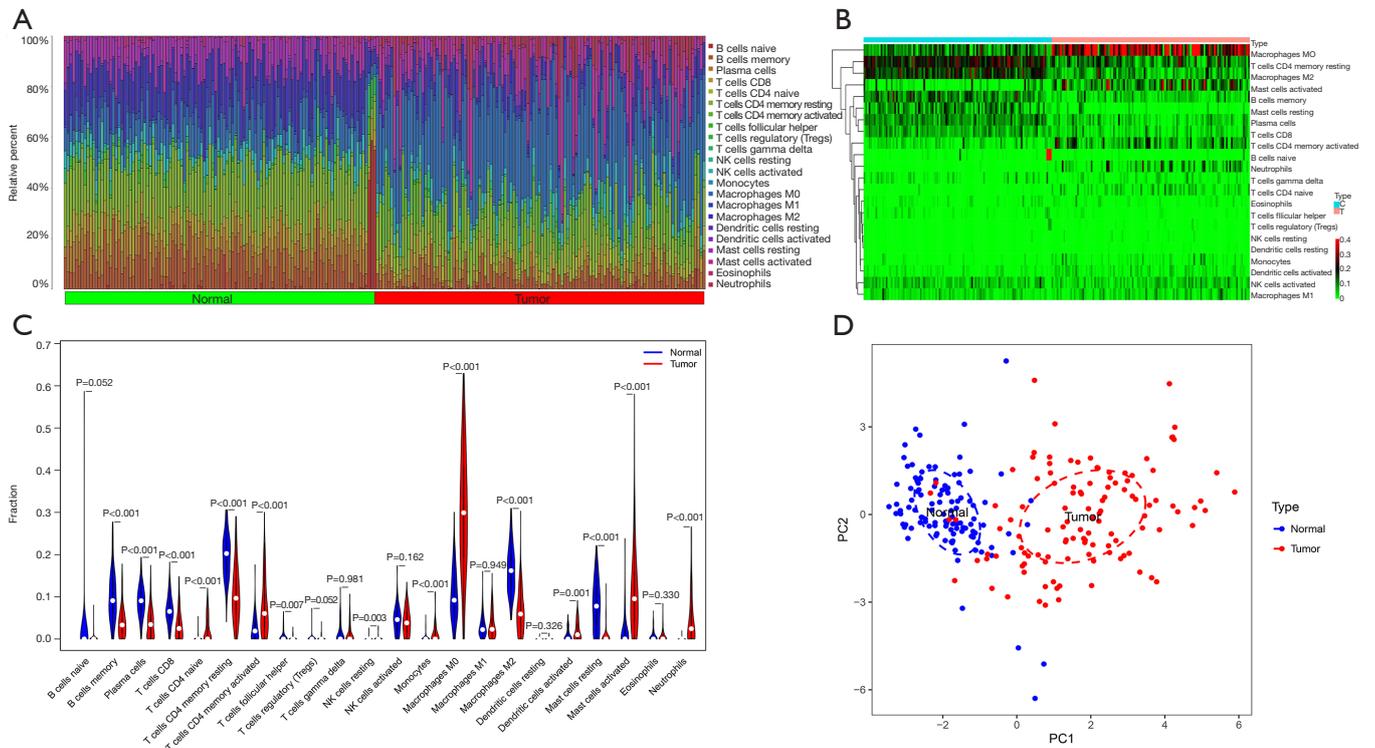


Figure 8 Diagnostic value of infiltrated immune cells. (A) Correlation between each of the infiltrated immune cell types. (B) ROC curves of infiltrated immune cells. ROC, receiver operating characteristic.

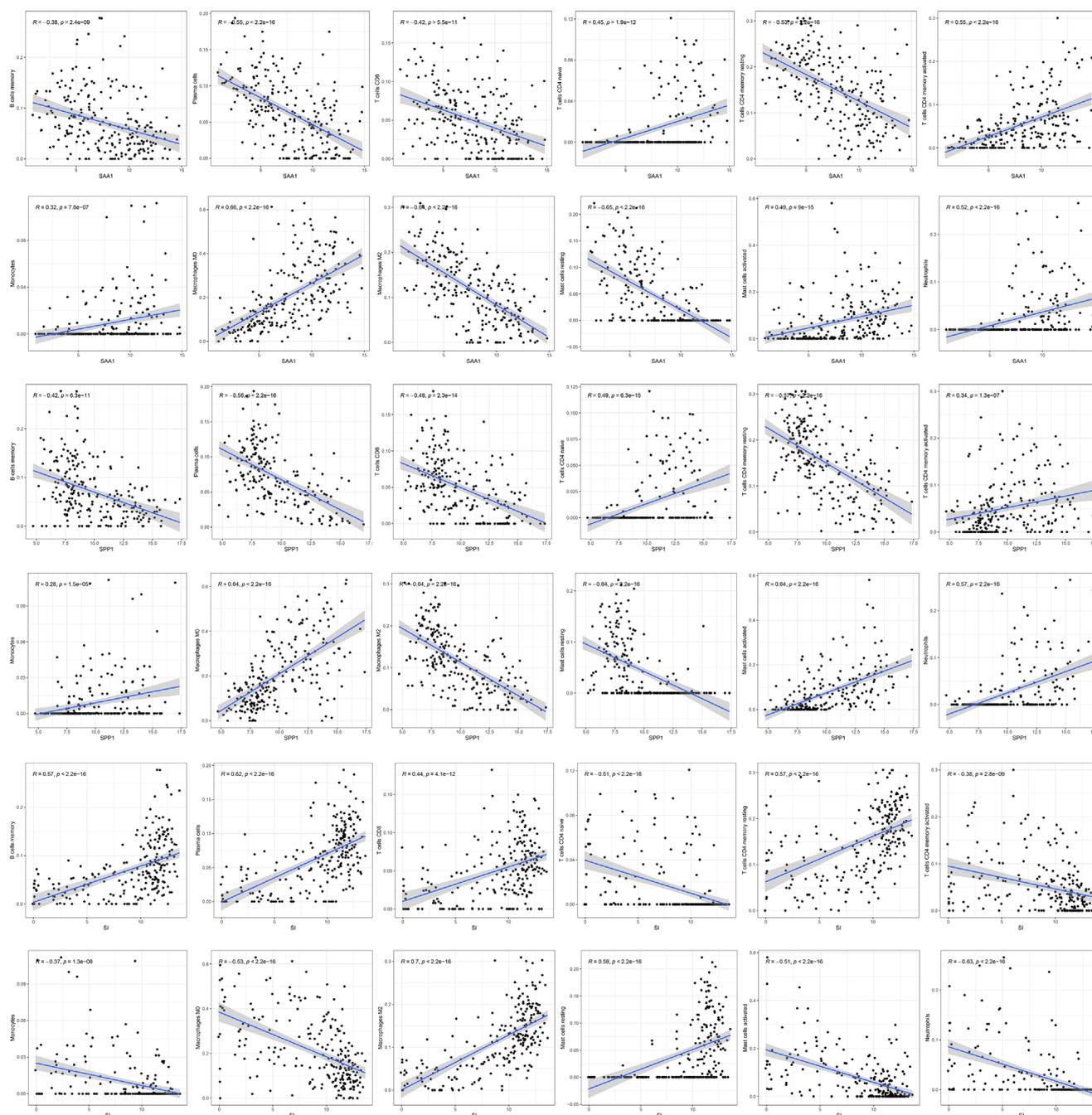


Figure 9 Relationship between the prognostic genes of *SAA1*, *SPP1*, and *SI*, and immune infiltration.

GCG, and CRC has not yet been illuminated. A single-chip-based study identified PPBP as the core gene of rectal cancer (27). Chemokine (C-X-C motif) ligands including CXCL1 and CXCL2 are the only diagnostic and prognostic biomarkers for colon cancer (28). SPP1 is related to a

poor prognosis of colon cancer, and MMP1 is related to a poor prognosis of CRC (20,29). SAA1 can promote breast cancer metastasis via immune cell infiltration, and it is highly expressed in colon cancer as well as TGFBI (30-32). The effects of downregulation or upregulation of

the above genes demonstrate that they also play a major role in the development of rectal cancer. Additionally, COL1A1, SLC30A10, and PCK1 can promote CRC metastasis via regulation of the WNT/PCP pathway, the miR-21c/APC axis, and nucleotide synthesis (33-35). Furthermore, the expression levels of cancer promoting SLC30A10 and PCK1 were decreased in rectal cancer samples based on the 2 GEO datasets, which needs to be further verified. In brief, GUCA2B, GUCA2A, TMIGD1, SLC26A3, MS4A12, CLCA4, SST, SI, CLCA1, AQP8, SLC4A4, GCG, CXCL1, and SLC30A10 which have been reported to be diagnostic biomarkers in CRC by other microarray dataset studies, PPBP and CXCL2 which were found to be biomarkers of rectal cancer in a single microarray dataset study, and CHGA, TTR, SAA1, SPP1, MMP1, TGFBI, COL1A1, and PCK1, may constitute a gene biomarker group for the identification of rectal cancer.

The correlation analysis between these gene markers and survival time based on the TCGA database showed that 8 genes were significantly correlated with survival time ($P < 0.05$). Among them, high expression levels of SAA1 and SPP1 are associated with a poor prognosis, while high expression levels of SI and CLCA1 are associated with an improved prognosis of rectal cancer, and the prognostic significance of CLCA1 has also been evidenced by research of other datasets, thus confirming the reliability of the study (23). For AQP8, MMP1, COL1A1, and TGFBI, the relationship between genes and prognosis is not consistent with the clinical significance of gene overexpression, which needs to be further verified using a larger sample. Thus, SAA1, SPP1, and SI could be new prognostic biomarkers of rectal cancer.

A variety of treatment methods based on immune classification exhibit certain advantages against cancer. Tumor immune infiltrating can directly or indirectly interfere with the progress of tumor through the release of cytokines or cytokine receptors and interaction with other components in the TME (36). The prognostic value of different immune cells depends on the type of cancer, and the type and abundance of infiltrated immune cells are important factors for tumor immune infiltration affecting the clinical prognosis. An understanding of the characteristics of immune cells and the body's immune state is helpful for evaluating the pathological status or therapeutic effects. Analysis of the distribution of the 22 types of infiltrated immune cells using the CIBERSORT algorithm showed that 15 types of immune cells differed significantly between normal and rectal cancer tissues, including 1 type of B lymphocyte, 1 type of plasma cell, 5 types of T lymphocytes, 1 type of NK cell, 1 type of monocyte,

2 types of macrophages, 1 type of dendritic cell, 2 types of mast cells, and neutrophils. Previous studies have shown that Tregs have both positive and negative effects in CRC, which may depend on the location of invasion and the definition of "Tregs" used (37). However, there was no difference of Tregs in this study, which may be related to the above factors, or the region, race and number of samples. The increased abundance of T cells CD4 naive, T cells CD4 memory activated, NK cells resting, monocytes, macrophages M0, dendritic cells activated, mast cells activated, and neutrophils implies that these cells may be positively correlated with rectal cancer, whereas the decreased abundance of B cells memory, plasma cells, T cells CD8, T cells CD4 memory resting, T cells follicular helper, macrophages M2, and mast cells resting implies that these cells may be negatively associated with rectal cancer. Among the 12 sensitive immune cells, the prognostic genes SAA1, SPP1, and SI all have the closest correlation with macrophages M2. As a reported diagnostic and prognostic signature of colon cancer and rectal cancer, CLCA1 is also closely correlated with the 12 immune cells. Other studies have shown that in CRC, infiltrated T helper 1 cells, T cells follicular helper, macrophages M1, dendritic cells, and NK cells are associated with a good prognosis, while macrophages M2, myeloid-derived suppressor cells, T helper 17 cells, and B cells are associated with a poor prognosis (38). Because features of rectal cancer may be not entirely identical to those of colon cancer, the correlation between these infiltrated immune cells and the prognosis of rectal cancer needs to be further studied.

In conclusion, a biomarker group of 24 differential genes integrating with immune cell infiltration characteristics was found to be conducive to obtaining a comprehensive reflection of the body state and effectively indicating rectal cancer. Functional enrichment analysis indicated that the hub genes were closely associated with the process of rectal cancer. Among them, SAA1, SPP1, and SI may be novel prognostic biomarkers of rectal cancer. Of course, due to the limitation resulting from sample sizes in the GEO and TCGA databases, validation of our results in clinical cohort studies based on multiple centers and a large sample size is needed in the future. This study may provide reference for early diagnosis and treatment monitoring of rectal cancer at molecular and immune levels.

Acknowledgments

Funding: This work was supported by National Natural Science Foundation of China (No. 81803773), Incubating

Program of Beijing Municipal Administration of Hospitals (PX2020078), and Open Project of Beijing Key Laboratory of Urine Cell Molecular Diagnosis (No. 2020-KF18).

Footnote

Reporting Checklist: The authors have completed the REMARK and MDAR reporting checklist. Available at <https://dx.doi.org/10.21037/jgo-21-255>

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at <https://dx.doi.org/10.21037/jgo-21-255>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are responsible for all aspects of the work to ensure that issues relating to the accuracy or completeness of any part of the work are properly investigated and resolved. Since this was a retrospective database study based on a public database, approval from the ethics committee was not required. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Siegel RL, Miller KD, Goding Sauer A, et al. Colorectal cancer statistics, 2020. *CA Cancer J Clin* 2020;70:145-64.
2. Modest DP, Pant S, Sartore-Bianchi A. Treatment sequencing in metastatic colorectal cancer. *Eur J Cancer* 2019;109:70-83.
3. Piawah S, Venook AP. Targeted therapy for colorectal cancer metastases: A review of current methods of molecularly targeted therapy and the use of tumor biomarkers in the treatment of metastatic colorectal cancer. *Cancer* 2019;125:4139-47.
4. Vassos N, Piso P. Metastatic Colorectal Cancer to the Peritoneum: Current Treatment Options. *Curr Treat Options Oncol* 2018;19:49.
5. Jiao Q, Ren Y, Ariston Gabrie AN, et al. Advances of immune checkpoints in colorectal cancer treatment. *Biomed Pharmacother* 2020;123:109745.
6. Bailey CE, Hu CY, You YN, et al. Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975-2010. *JAMA Surg* 2015;150:17-22.
7. Dayde D, Tanaka I, Jain R, et al. Predictive and Prognostic Molecular Biomarkers for Response to Neoadjuvant Chemoradiation in Rectal Cancer. *Int J Mol Sci* 2017;18:573.
8. De Felice F, Crocetti D, Maiuri V, et al. Locally Advanced Rectal Cancer: Treatment Approach in Elderly Patients. *Curr Treat Options Oncol* 2020;21:1.
9. Oh HH, Joo YE. Novel biomarkers for the diagnosis and prognosis of colorectal cancer. *Intest Res* 2020;18:168-83.
10. Galon J, Angell HK, Bedognetti D, et al. The continuum of cancer immunosurveillance: prognostic, predictive, and mechanistic signatures. *Immunity* 2013;39:11-26.
11. Shen Y, Guan Y, Hummel JJ, et al. Immunogenomic pathways associated with cytotoxic lymphocyte infiltration and survival in colorectal cancer. *BMC Cancer* 2020;20:124.
12. Kolde R, Laur S, Adler P, et al. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* 2012;28:573-80.
13. Niu J, Yan T, Guo W, et al. Identification of Potential Therapeutic Targets and Immune Cell Infiltration Characteristics in Osteosarcoma Using Bioinformatics Strategy. *Front Oncol* 2020;10:1628.
14. Chin CH, Chen SH, Wu HH, et al. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* 2014;8 Suppl 4:S11.
15. Xu H, Ma Y, Zhang J, et al. Identification and Verification of Core Genes in Colorectal Cancer. *Biomed Res Int* 2020;2020:8082697.
16. Pattison AM, Merlino DJ, Blomain ES, et al. Guanylyl cyclase C signaling axis and colon cancer prevention. *World J Gastroenterol* 2016;22:8070-7.
17. Meyer RD, Zou X, Ali M, et al. TMIGD1 acts as a tumor suppressor through regulation of p21Cip1/p27Kip1 in renal cancer. *Oncotarget* 2018;9:9672-84.
18. Bhutia YD, Babu E, Ramachandran S, et al. SLC transporters as a novel class of tumour suppressors: identity, function and molecular mechanisms. *Biochem J* 2016;473:1113-24.
19. Koslowski M, Sahin U, Dhaene K, et al. MS4A12 is a colon-selective store-operated calcium channel promoting malignant cell processes. *Cancer Res* 2008;68:3458-66.
20. Chen L, Lu D, Sun K, et al. Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. *Gene*

- 2019;692:119-25.
21. Chen H, Liu Y, Jiang CJ, et al. Calcium-Activated Chloride Channel A4 (CLCA4) Plays Inhibitory Roles in Invasion and Migration Through Suppressing Epithelial-Mesenchymal Transition via PI3K/AKT Signaling in Colorectal Cancer. *Med Sci Monit* 2019;25:4176-85.
 22. Kim SB, Calmet FH, Garrido J, et al. Sucrase-Isomaltase Deficiency as a Potential Masquerader in Irritable Bowel Syndrome. *Dig Dis Sci* 2020;65:534-40.
 23. Wei FZ, Mei SW, Wang ZJ, et al. Differential Expression Analysis Revealing CLCA1 to Be a Prognostic and Diagnostic Biomarker for Colorectal Cancer. *Front Oncol* 2020;10:573295.
 24. Chow PH, Bowen J, Yool AJ. Combined Systematic Review and Transcriptomic Analyses of Mammalian Aquaporin Classes 1 to 10 as Biomarkers and Prognostic Indicators in Diverse Cancers. *Cancers (Basel)* 2020;12:1911.
 25. Zhang X, Zhang H, Shen B, et al. Chromogranin-A Expression as a Novel Biomarker for Early Diagnosis of Colon Cancer Patients. *Int J Mol Sci* 2019;20:2919.
 26. Yang H, Lu Y, Lan W, et al. Down-regulated Solute Carrier Family 4 Member 4 Predicts Poor Progression in Colorectal Cancer. *J Cancer* 2020;11:3675-84.
 27. Liu BX, Huang GJ, Cheng HB. Comprehensive Analysis of Core Genes and Potential Mechanisms in Rectal Cancer. *J Comput Biol* 2019;26:1262-77.
 28. Zhuo C, Wu X, Li J, et al. Chemokine (C-X-C motif) ligand 1 is associated with tumor progression and poor prognosis in patients with colorectal cancer. *Biosci Rep* 2018;38:BSR20180580.
 29. Choe EK, Yi JW, Chai YJ, et al. Upregulation of the adipokine genes ADIPOR1 and SPP1 is related to poor survival outcomes in colorectal cancer. *J Surg Oncol* 2018;117:1833-40.
 30. Gutfeld O, Prus D, Ackerman Z, et al. Expression of serum amyloid A, in normal, dysplastic, and neoplastic human colonic mucosa: implication for a role in colonic tumorigenesis. *J Histochem Cytochem* 2006;54:63-73.
 31. Hansen MT, Forst B, Cremers N, et al. A link between inflammation and metastasis: serum amyloid A1 and A3 induce metastasis, and are targets of metastasis-inducing S100A4. *Oncogene* 2015;34:424-35.
 32. Zhu J, Chen X, Liao Z, et al. TGFBI protein high expression predicts poor prognosis in colorectal cancer patients. *Int J Clin Exp Pathol* 2015;8:702-10.
 33. Zhang Z, Wang Y, Zhang J, et al. COL1A1 promotes metastasis in colorectal cancer by regulating the WNT/PCP pathway. *Mol Med Rep* 2018;17:5037-42.
 34. Hou L, Liu P, Zhu T. Long noncoding RNA SLC30A10 promotes colorectal tumor proliferation and migration via miR-21c/APC axis. *Eur Rev Med Pharmacol Sci* 2020;24:6682-91.
 35. Yamaguchi N, Weinberg EM, Nguyen A, et al. PCK1 and DHODH drive colorectal cancer liver metastatic colonization and hypoxic growth by promoting nucleotide synthesis. *Elife* 2019;8:52135.
 36. Li W, Jin X, Guo S, et al. Comprehensive analysis of prognostic immune-related genes in the tumor microenvironment of colorectal cancer. *Aging (Albany NY)* 2021;13:5506-24.
 37. Jochems C, Schlom J. Tumor-infiltrating immune cells and prognosis: the potential link between conventional cancer therapy and immunity. *Exp Biol Med (Maywood)* 2011;236:567-79.
 38. Roelands J, Kuppen PJK, Vermeulen L, et al. Immunogenomic Classification of Colorectal Cancer and Therapeutic Implications. *Int J Mol Sci* 2017;18:2229.
- (English Language Editor: J. Reynolds)

Cite this article as: Wen L, Han Z, Du Y. Identification of gene biomarkers and immune cell infiltration characteristics in rectal cancer. *J Gastrointest Oncol* 2021;12(3):964-980. doi: 10.21037/jgo-21-255

Table S1 The specific information of all robust DEGs

Name	P value	fdr	logFC
REG1A	3.60E-08	0.000804685	7.067253816
KRT23	9.99E-08	0.001072913	6.955363898
DPEP1	1.96E-07	0.001095266	6.72476472
MMP7	1.44E-07	0.001072913	6.549577091
FOXQ1	4.00E-07	0.001788189	6.384793419
CLDN2	1.60E-06	0.003251253	5.857558238
MMP3	8.99E-07	0.002873875	5.654452334
CLDN1	1.16E-06	0.00321874	5.542775508
FABP6	5.76E-07	0.002145827	5.538320689
REG3A	1.44E-06	0.003227681	5.409012457
REG1B	1.30E-06	0.00321874	5.184640207
GRIN2D	7.74E-06	0.009616483	5.048970125
KLK8	2.91E-06	0.005431624	4.835866166
KRT6B	4.62E-06	0.007301772	4.756270403
PITX2	4.09E-06	0.007042714	4.705183898
INHBA	9.60E-06	0.011298531	4.595602263
CDH3	2.31E-05	0.021517874	4.473172533
SLC35D3	1.17E-05	0.013035898	4.47180081
SAA2	4.90E-06	0.007301772	4.465944253
KLK6	1.96E-05	0.0190481	4.448137186
CPNE7	5.47E-05	0.034969313	4.382447635
CKMT2	6.72E-06	0.008841017	4.364945696
MMP1	5.47E-06	0.007650096	4.331675048
C17orf96	4.66E-05	0.031819246	4.198304502
FAM3B	1.74E-05	0.017703071	4.173726857
SLC7A5	6.35E-05	0.037354327	4.163926198
SAA1	4.66E-05	0.031819246	3.999204346
H19	2.43E-05	0.021758684	3.99540576
CA9	1.54E-05	0.016366187	3.970272462
SNAR-A3	8.17E-05	0.040629641	3.720662917
TACSTD2	4.16E-05	0.031007197	3.679116364
SULT2B1	4.84E-05	0.031819246	3.677548617
TESC	3.46E-05	0.027617941	3.62325825
LY6G6D	7.95E-05	0.040629641	3.601946959
PPBP	2.56E-05	0.02200848	3.568882823
VSNL1	5.85E-05	0.035379561	3.548623247
ASCL2	6.55E-05	0.037561139	3.399986192
TRIB3	3.68E-05	0.028413707	3.379930813
MDF1	4.75E-05	0.031819246	3.358477233
TUBB3	3.17E-05	0.026230084	3.322225549
SRPX2	7.28E-05	0.039743591	3.228786088
HULC	0.000106197	0.049489994	3.155162765
CXCL2	8.99E-05	0.043732883	3.153721806
TNFRSF6B	6.86E-05	0.03835889	3.149277493
SPP1	5.85E-05	0.035379561	3.134587104
CXCL1	7.50E-05	0.03995538	3.098551712
TGFB1	0.000108819	0.049676986	3.04600235
COL1A1	8.06E-05	0.040629641	2.958479939
TNS4	9.60E-05	0.045703448	2.627302161
PLAC8	0.000184762	0.047505194	-2.680929554
STMN2	0.000172927	0.046604945	-2.731111464
LOC401022	0.000136798	0.040800513	-2.795494485
NR3C2	0.000141271	0.041040099	-2.800447818
IL1R2	0.000114159	0.036480333	-2.909253911
BCAS1	0.000119627	0.03716577	-2.913712432
CAPN13	0.000125223	0.038371352	-2.928341249
BTNL3	0.000148885	0.042157122	-2.934733402
TNFRSF17	0.000184762	0.047505194	-2.959291194
HOXB13	0.000116877	0.036822841	-2.984102
C11orf86	9.11E-05	0.032881046	-3.022635275
SPINK5	0.000107504	0.035891889	-3.036781464
AKR1B15	9.60E-05	0.033997944	-3.043733368
AKR1B10	9.73E-05	0.033997944	-3.088296286
TMEM236	0.000101049	0.034680029	-3.108926689
PADI2	8.40E-05	0.030816946	-3.124531267
CHST5	0.000110142	0.036231865	-3.127271725
ST6GALNAC6	0.000144293	0.041380528	-3.153655015
MT1G	0.00016634	0.045936589	-3.216222322
HS3ST6	0.000155121	0.043373642	-3.21895104
TRPM6	0.000171269	0.046604945	-3.236563862
LDHD	7.61E-05	0.029279322	-3.252022588
FCGBP	0.000139772	0.041040099	-3.25376384
MFSD4	5.85E-05	0.025667525	-3.261631342
PDE6A	7.18E-05	0.028587495	-3.275371138
FAM55D	0.000176269	0.046830566	-3.29641639
CR2	6.15E-05	0.026437687	-3.340510097
SLC28A2	5.47E-05	0.024978081	-3.351177076
KIF5C	5.66E-05	0.025322545	-3.4006025
SCGB2A1	0.000102324	0.034680029	-3.404421705
C4orf7	8.17E-05	0.030472231	-3.421311474
TEX11	0.000177952	0.046830566	-3.442335511
PCK1	4.00E-05	0.020792895	-3.541396683
ANO5	6.35E-05	0.026705278	-3.552878636
C2orf88	4.58E-05	0.023264745	-3.581756146
KRTAP13-1	2.69E-05	0.015793377	-3.62724934
LGALS2	4.92E-05	0.02336567	-3.683600924
GLDN	4.75E-05	0.02336567	-3.693286819
GREM2	7.28E-05	0.028587495	-3.707330371
VSIG2	2.37E-05	0.015145961	-3.781985895
SLC26A2	6.45E-05	0.026705278	-3.799521108
SPINK2	2.01E-05	0.014400181	-3.804981368
PKIB	2.75E-05	0.015793377	-3.805917067
BEST2	3.03E-05	0.01611286	-3.807699947
BMP3	7.72E-05	0.029279322	-3.817950304
CHP2	2.19E-05	0.014400181	-3.830129937
MT1M	0.000132397	0.040021603	-3.855859584
CD177	5.01E-05	0.02336567	-3.923270978
ITLN1	0.000112812	0.036480333	-3.938716882
SCNN1B	1.49E-05	0.011881877	-3.952652682
SI	7.18E-05	0.028587495	-3.953383453
PI16	0.000193456	0.049175198	-4.005962822
TTR	2.96E-05	0.01611286	-4.076434414
ADH1C	2.19E-05	0.014400181	-4.077470972
HEPACAM2	2.56E-05	0.015793377	-4.086945719
LOC646627	3.03E-05	0.01611286	-4.092311352
ADH1A	2.07E-05	0.014400181	-4.150075466
SPIB	4.84E-05	0.02336567	-4.242971012
CWH43	1.74E-05	0.012982252	-4.371309279
CA4	6.08E-06	0.007157462	-4.447447023
SLC26A3	7.05E-06	0.007868032	-4.4496308
DAO	9.21E-06	0.00894467	-4.486186892
CDKN2B-AS	5.77E-06	0.007157462	-4.672630413
SLC30A10	1.21E-05	0.010402446	-4.680812871
LOC389023	1.12E-05	0.010046046	-4.720048388
CLCA4	2.70E-06	0.005036732	-4.776568993
CCL23	3.36E-06	0.005784104	-4.805062895
OGN	1.59E-05	0.012236762	-4.836822995
ZG16	7.74E-06	0.007868032	-4.845514424
SCGN	5.18E-06	0.007157462	-4.899460196
C2orf40	9.60E-06	0.00894467	-4.939166165
CA2	1.93E-06	0.004681804	-4.962759686
MYOC	6.08E-06	0.007157462	-5.00701001
GCG	5.18E-06	0.007157462	-5.010953842
AQP8	1.30E-05	0.010758937	-5.032363824
BEST4	2.62E-05	0.015793377	-5.052843
MS4A12	1.60E-06	0.004681804	-5.061831438
CLCA1	7.39E-06	0.007868032	-5.272034511
GUCA2A	8.99E-07	0.004681804	-5.31427127
SLC4A4	1.76E-06	0.004681804	-5.326196622
CA7	3.84E-06	0.00613732	-5.531522238
GUCA2B	2.30E-06	0.004681804	-5.560135294
SST	2.30E-06	0.004681804	-5.638536208
CHGA	2.11E-06	0.004681804	-5.68823246
CLDN8	1.44E-06	0.004681804	-5.871240528
INSL5	6.75E-07	0.004681804	-6.414954285
TMIGD1	1.44E-07	0.00160937	-6.44848406
CA1	1.60E-08	0.000357638	-6.980474323

Table S2 GO enrichment analysis of robust DEGs

ONTOLOGY	ID	Description	Gene ratio	Bg ratio	P value	p.adjust	q value	Gene ID	Count
BP	GO:0015701	bicarbonate transport	8/120	42/18670	2.26E-10	4.05E-07	3.75E-07	CA1/SLC4A4/CA2/CA7/CA4/SLC26A3/CA9/SLC26A2	8
BP	GO:0006821	chloride transport	9/120	108/18670	3.17E-08	2.84E-05	2.63E-05	CA2/CLCA4/CA7/SLC26A3/CLCA1/BEST4/BEST2/ANO5/SLC26A2	9
BP	GO:0019730	antimicrobial humoral response	9/120	122/18670	9.18E-08	5.47E-05	5.07E-05	REG1A/REG1B/REG3A/CHGA/PPBP/SPINK5/ITLN1/CXCL1/CXCL2	9
BP	GO:0015698	inorganic anion transport	10/120	169/18670	1.39E-07	6.24E-05	5.78E-05	SLC4A4/CA2/CLCA4/CA7/SLC26A3/CLCA1/BEST4/BEST2/ANO5/SLC26A2	10
BP	GO:0061844	antimicrobial humoral immune response mediated by antimicrobial peptide	7/120	73/18670	4.36E-07	0.000156015	0.000144501	REG1A/REG1B/REG3A/PPBP/SPINK5/CXCL1/CXCL2	7
BP	GO:0042572	retinol metabolic process	5/120	41/18670	6.28E-06	0.001874325	0.001735997	ADH1A/ADH1C/TTR/AKR1B15/AKR1B10	5
BP	GO:1902476	chloride transmembrane transport	6/120	88/18670	2.19E-05	0.00560664	0.005192859	CLCA4/SLC26A3/CLCA1/BEST4/BEST2/SLC26A2	6
BP	GO:1990868	response to chemokine	6/120	97/18670	3.81E-05	0.00758666	0.007026751	REG1A/CCL23/PPBP/PADI2/CXCL1/CXCL2	6
BP	GO:1990869	cellular response to chemokine	6/120	97/18670	3.81E-05	0.00758666	0.007026751	REG1A/CCL23/PPBP/PADI2/CXCL1/CXCL2	6
BP	GO:0071241	cellular response to inorganic substance	8/120	215/18670	7.56E-05	0.013221823	0.012246028	DPEP1/MMP3/CLDN1/SLC30A10/CHP2/CPNE7/MT1M/MT1G	8
BP	GO:0098661	inorganic anion transmembrane transport	6/120	111/18670	8.13E-05	0.013221823	0.012246028	CLCA4/SLC26A3/CLCA1/BEST4/BEST2/SLC26A2	6
BP	GO:0006959	humoral immune response	10/120	356/18670	0.000101419	0.015128279	0.014011784	REG1A/REG1B/REG3A/CHGA/PPBP/CR2/SPINK5/ITLN1/CXCL1/CXCL2	10
BP	GO:1990266	neutrophil migration	6/120	118/18670	0.000114011	0.015698396	0.014539826	CCL23/PPBP/CD177/SAI1/CXCL1/CXCL2	6
BP	GO:0016999	antibiotic metabolic process	6/120	122/18670	0.000136973	0.017512945	0.016220457	DPEP1/MMP3/ADH1A/ADH1C/PCK1/AKR1B10	6
BP	GO:0034754	cellular hormone metabolic process	6/120	129/18670	0.000185856	0.021834692	0.020223252	ADH1A/ADH1C/TTR/AKR1B15/AKR1B10/SPP1	6
BP	GO:0008544	epidermis development	11/120	464/18670	0.000202022	0.021834692	0.020223252	KRT23/FOXQ1/REG3A/PITX2/KRT6B/INHBA/KRTAP131/CDH3/SULT2B1/SPINK5/HOXB13	11
BP	GO:0071248	cellular response to metal ion	7/120	188/18670	0.000214281	0.021834692	0.020223252	DPEP1/CLDN1/SLC30A10/CHP2/CPNE7/MT1M/MT1G	7
BP	GO:0034308	primary alcohol metabolic process	5/120	85/18670	0.000219567	0.021834692	0.020223252	ADH1A/ADH1C/TTR/AKR1B15/AKR1B10	5
BP	GO:0030574	collagen catabolic process	4/120	47/18670	0.000233841	0.022030308	0.020404431	MMP7/MMP3/MMP1/KLK6	4
BP	GO:0070098	chemokine-mediated signaling pathway	5/120	88/18670	0.000258251	0.023113498	0.02140768	CCL23/PPBP/PADI2/CXCL1/CXCL2	5
BP	GO:0097530	granulocyte migration	6/120	141/18670	0.000300792	0.025638924	0.023746725	CCL23/PPBP/CD177/SAI1/CXCL1/CXCL2	6
BP	GO:0050900	leukocyte migration	11/120	499/18670	0.000375622	0.029276683	0.02711601	CHGA/CCL23/MMP1/PPBP/CD177/PADI2/SAI1/SLC7A5/CXCL1/COL1A1/CXCL2	11
BP	GO:0043588	skin development	10/120	419/18670	0.000376181	0.029276683	0.02711601	KRT23/FOXQ1/CLDN1/REG3A/KRT6B/INHBA/KRTAP13-1/CDH3/SPINK5/COL1A1	10
BP	GO:0016338	calcium-independent cell-cell adhesion via plasma membrane cell-adhesion molecules	3/120	23/18670	0.00041749	0.030007083	0.027792505	CLDN1/CLDN2/CLDN8	3
BP	GO:0097529	myeloid leukocyte migration	7/120	210/18670	0.000419093	0.030007083	0.027792505	CHGA/CCL23/PPBP/CD177/SAI1/CXCL1/CXCL2	7
BP	GO:0001523	retinoid metabolic process	5/120	104/18670	0.000558758	0.037043618	0.034309731	ADH1A/ADH1C/TTR/AKR1B15/AKR1B10	5
BP	GO:0030593	neutrophil chemotaxis	5/120	104/18670	0.000558758	0.037043618	0.034309731	CCL23/PPBP/SAI1/CXCL1/CXCL2	5
BP	GO:0098656	anion transmembrane transport	8/120	291/18670	0.000588718	0.037635932	0.034858332	SLC4A4/CLCA4/SLC26A3/CLCA1/BEST4/BEST2/SLC26A2/SLC7A5	8
BP	GO:0030595	leukocyte chemotaxis	7/120	224/18670	0.000615612	0.037998147	0.035193815	CHGA/CCL23/PPBP/PADI2/SAI1/CXCL1/CXCL2	7
BP	GO:0016101	diterpenoid metabolic process	5/120	110/18670	0.000721278	0.043036266	0.039860111	ADH1A/ADH1C/TTR/AKR1B15/AKR1B10	5
BP	GO:0042445	hormone metabolic process	7/120	233/18670	0.000776471	0.043786311	0.040554802	ADH1A/ADH1C/TTR/CLK6/AKR1B15/AKR1B10/SPP1	7
BP	GO:0060326	cell chemotaxis	8/120	304/18670	0.000782772	0.043786311	0.040554802	CHGA/CCL23/SAI1/PPBP/PADI2/SAI1/CXCL1/CXCL2	8
BP	GO:0048545	response to steroid hormone	9/120	383/18670	0.000827663	0.044894456	0.041581164	CLDN1/CA2/SST/PCK1/PADI2/SCGB2A1/SPP1/COL1A1/NR3C2	9
BP	GO:0032963	collagen metabolic process	5/120	115/18670	0.000881756	0.046421855	0.042995837	MMP7/MMP3/MMP1/KLK6/COL1A1	5
CC	GO:0045177	apical part of cell	12/126	384/19717	6.64E-06	0.001089398	0.000922982	DPEP1/CLDN1/CA2/CLCA4/CA4/SLC26A3/AQP8/SCNN1B/SLC26A2/SI/SLC7A5/TRPM6	12
CC	GO:0016324	apical plasma membrane	10/126	318/19717	3.82E-05	0.003133227	0.002854595	DPEP1/CLDN1/CLCA4/CA4/SLC26A3/SCNN1B/SLC26A2/SI/SLC7A5/TRPM6	10
CC	GO:0042588	zymogen granule	3/126	14/19717	8.81E-05	0.004816077	0.004080373	REG1A/CLCA1/ZG16	3
CC	GO:0005902	microvillus	5/126	83/19717	0.00019182	0.007864638	0.006663237	DPEP1/CA2/CLCA1/CA9/SLC26A2	5
CC	GO:0031253	cell projection membrane	9/126	326/19717	0.000250332	0.007938419	0.006725747	REG1A/DPEP1/CA4/SLC26A3/CA9/SLC26A2/TESC/ITLN1/TRPM6	9
CC	GO:0031526	brush border membrane	4/126	53/19717	0.00036542	0.007938419	0.006725747	CA4/SLC26A3/ITLN1/TRPM6	4
CC	GO:0098858	actin-based cell projection	7/126	208/19717	0.000384553	0.007938419	0.006725747	DPEP1/CA2/CLCA1/CA9/TUBB3/SLC26A2/LY6G6D	7
CC	GO:0031528	microvillus membrane	3/126	23/19717	0.00041099	0.007938419	0.006725747	DPEP1/CA9/SLC26A2	3
CC	GO:0005903	brush border	5/126	99/19717	0.000435645	0.007938419	0.006725747	CA4/SLC26A3/SI/ITLN1/TRPM6	5
CC	GO:0042589	zymogen granule membrane	2/126	11/19717	0.002145814	0.03519135	0.029815521	CLCA1/ZG16	2
CC	GO:0016323	basolateral plasma membrane	6/126	217/19717	0.002732176	0.038048876	0.032236531	CLDN1/CLDN8/SLC4A4/CA2/CA9/TACSTD2	6
CC	GO:0098862	cluster of actin-based cell projections	5/126	150/19717	0.002784064	0.038048876	0.032236531	CA4/SLC26A3/SI/ITLN1/TRPM6	5
CC	GO:0031225	anchored component of membrane	5/126	170/19717	0.004744703	0.055636741	0.047137675	DPEP1/CA4/CD177/ITLN1/LY6G6D	5
CC	GO:0060205	cytoplasmic vesicle lumen	7/126	325/19717	0.004923828	0.055636741	0.047137675	GCG/ZG16/TTR/PPBP/PADI2/CXCL1/PLAC8	7
CC	GO:0031983	vesicle lumen	7/126	327/19717	0.005088726	0.055636741	0.047137675	GCG/ZG16/TTR/PPBP/PADI2/CXCL1/PLAC8	7
MF	GO:0048018	receptor ligand activity	18/117	482/17696	2.74E-09	4.10E-07	3.18E-07	REG1A/INSL5/GUCA2A/SST/CCL23/GCG/SAI1/INHBA/OGN/TTR/FAM3B/PPBP/GREM2/BMP3/SAI1/SPP1/CXCL1/CXCL2	18
MF	GO:0030546	signaling receptor activator activity	18/117	487/17696	3.23E-09	4.10E-07	3.18E-07	REG1A/INSL5/GUCA2A/SST/CCL23/GCG/SAI1/INHBA/OGN/TTR/FAM3B/PPBP/GREM2/BMP3/SAI1/SPP1/CXCL1/CXCL2	18
MF	GO:0004089	carbonate dehydratase activity	5/117	15/17696	3.30E-08	2.80E-06	2.17E-06	CA1/CA2/CA7/CA4/CA9	5
MF	GO:0070492	oligosaccharide binding	4/117	15/17696	2.34E-06	0.000148713	0.000115248	REG1A/REG1B/REG3A/ITLN1	4
MF	GO:0015108	chloride transmembrane transporter activity	7/117	100/17696	4.44E-06	0.000195	0.000151119	CLCA4/SLC26A3/CLCA1/BEST4/BEST2/ANO5/SLC26A2	7
MF	GO:0015103	inorganic anion transmembrane transporter activity	8/117	142/17696	4.61E-06	0.000195	0.000151119	SLC4A4/CLCA4/SLC26A3/CLCA1/BEST4/BEST2/ANO5/SLC26A2	8
MF	GO:0005125	cytokine activity	9/117	220/17696	1.56E-05	0.000564869	0.000437756	CCL23/INHBA/FAM3B/PPBP/GREM2/BMP3/SPP1/CXCL1/CXCL2	9
MF	GO:0016836	hydro-lyase activity	5/117	56/17696	3.38E-05	0.001074314	0.00083256	CA1/CA2/CA7/CA4/CA9	5
MF	GO:0005254	chloride channel activity	5/117	75/17696	0.000138359	0.003108513	0.002409001	CLCA4/CLCA1/BEST4/BEST2/ANO5	5
MF	GO:0016835	carbon-oxygen lyase activity	5/117	75/17696	0.000138359	0.003108513	0.002409001	CA1/CA2/CA7/CA4/CA9	5
MF	GO:0005229	intracellular calcium activated chloride channel activity	3/117	16/17696	0.000148141	0.003108513	0.002409001	CLCA4/CLCA1/ANO5	3
MF	GO:0061778	intracellular chloride channel activity	3/117	16/17696	0.000148141	0.003108513	0.002409001	CLCA4/CLCA1/ANO5	3
MF	GO:0005179	hormone activity	6/117	122/17696	0.000159097	0.003108513	0.002409001	INSL5/GUCA2A/SST/GCG/INHBA/TTR	6
MF	GO:0042834	peptidoglycan binding	3/117	17/17696	0.000179019	0.003247921	0.002517038	REG1A/REG1B/REG3A	3
MF	GO:0015106	bicarbonate transmembrane transporter activity	3/117	19/17696	0.000252653	0.004278256	0.003315515	SLC4A4/SLC26A3/CLCA1/BEST4/BEST2/ANO5/SLC26A2	3
MF	GO:0008009	chemokine activity	4/117	49/17696	0.000305659	0.004619625	0.003580065	CCL23/PPBP/CXCL1/CXCL2	4
MF	GO:0005253	anion channel activity	5/117	89/17696	0.000309187	0.004619625	0.003580065	CLCA4/CLCA1/BEST4/BEST2/ANO5	5
MF	GO:0008509	anion transmembrane transporter activity	9/117	336/17696	0.000394355	0.005564781	0.004312533	SLC4A4/CLCA4/SLC26A3/CLCA1/BEST4/BEST2/ANO5/SLC26A2/SLC7A5	9
MF	GO:0001664	G protein-coupled receptor binding	8/117	280/17696	0.000547332	0.007316963	0.005670419	INSL5/CCL23/GCG/MYOC/PPBP/SAI1/CXCL1/CXCL2	8
MF	GO:0008083	growth factor activity	6/117	163/17696	0.000750509	0.009531459	0.007386585	REG1A/INHBA/OGN/PPBP/BMP3/CXCL1	6
MF	GO:0005539	glycosaminoglycan binding	7/117	229/17696	0.000824171	0.00996855	0.007725316	REG1A/MMP7/REG1B/REG3A/CCL23/GREM2/SAI1	7
MF	GO:0042379	chemokine receptor binding	4/117	66/17696	0.000953788	0.011011912	0.008533889	CCL23/PPBP/CXCL1/CXCL2	4
MF	GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	5/117	119/17696	0.001162877	0.012842207	0.009952311	ADH1A/ADH1C/LDHD/AKR1B15/AKR1B10	5
MF	GO:0008237	metallopeptidase activity	6/117	181/17696	0.001290424	0.01365699	0.010583743	MMP7/DPEP1/MMP3/CLCA4/MMP1/CLCA1	6
MF	GO:0016829	lyase activity	6/117	187/17696	0.001524149	0.015485349	0.012000664	CA1/CA2/CA7/CA4/CA9/PCK1	6
MF	GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	5/117	128/17696	0.001607255	0.015701644	0.012168286	ADH1A/ADH1C/LDHD/AKR1B15/AKR1B10	5
MF	GO:0018455	alcohol dehydrogenase [NAD(P)+] activity	2/117	10/17696	0.001883955	0.017090162	0.013244344	ADH1A/ADH1C	2
MF	GO:0019531	oxalate transmembrane transporter activity	2/117	10/17696	0.001883955	0.017090162	0.013244344	SLC26A3/SLC26A2	2
MF	GO:0030246	carbohydrate binding	7/117	271/17696	0.002167334	0.018982857	0.014711124	REG1A/REG1B/REG3A/ZG16/LGALS2/SI/ITLN1	7
MF	GO:0052650	NADP-retinol dehydrogenase activity	2/117	11/17696	0.002292675	0.019411312	0.015043163	AKR1B15/AKR1B10	2
MF	GO:0022839	ion gated channel activity	3/117	43/17					

Table S3 KEGG pathway enrichment analysis of robust DEGs

ID	Description	Gene ratio	Bg ratio	P value	p.adjust	q value	Gene ID	Count
hsa00910	Nitrogen metabolism	5/59	17/8102	9.97E-08	1.16E-05	9.97E-06	CA1/CA2/CA7/CA4/CA9	5
hsa04964	Proximal tubule bicarbonate reclamation	4/59	23/8102	2.03E-05	0.001174916	0.001012859	SLC4A4/CA2/CA4/PCK1	4
hsa04060	Cytokine-cytokine receptor interaction	9/59	295/8102	0.000251876	0.00973919	0.008395854	CCL23/INHBA/PPBP/BMP3/IL1R2/TNFRSF6B/CXCL1/CXCL2/TNFRSF17	9
hsa00620	Pyruvate metabolism	4/59	47/8102	0.000358211	0.010388126	0.008955281	ADH1A/ADH1C/PCK1/LDHD	4
hsa04972	Pancreatic secretion	5/59	102/8102	0.000837155	0.016559945	0.014275815	SLC4A4/CA2/CLCA4/SLC26A3/CLCA1	5
hsa04978	Mineral absorption	4/59	59/8102	0.000856549	0.016559945	0.014275815	SLC26A3/MT1M/MT1G/TRPM6	4

Table S4 GO enrichment analysis of hub genes

ONTOLOGY	ID	Description	Gene ratio	Bg ratio	P value	p.adjust	q value	Gene ID	Count
BP	GO:0050900	leukocyte migration	7/23	499/18670	1.58E-06	0.000867623	0.000552656	CHGA/MMP1/PPBP/SAA1/CXCL1/COL1A1/CXCL2	7
BP	GO:0097529	myeloid leukocyte migration	5/23	210/18670	4.90E-06	0.000867623	0.000552656	CHGA/PPBP/SAA1/CXCL1/CXCL2	5
BP	GO:0051591	response to cAMP	4/23	97/18670	5.62E-06	0.000867623	0.000552656	SLC26A3/AQP8/PCK1/COL1A1	4
BP	GO:0030595	leukocyte chemotaxis	5/23	224/18670	6.71E-06	0.000867623	0.000552656	CHGA/PPBP/SAA1/CXCL1/CXCL2	5
BP	GO:0030593	neutrophil chemotaxis	4/23	104/18670	7.42E-06	0.000867623	0.000552656	PPBP/SAA1/CXCL1/CXCL2	4
BP	GO:1990266	neutrophil migration	4/23	118/18670	1.22E-05	0.001054195	0.000671498	PPBP/SAA1/CXCL1/CXCL2	4
BP	GO:0019730	antimicrobial humoral response	4/23	122/18670	1.40E-05	0.001054195	0.000671498	CHGA/PPBP/CXCL1/CXCL2	4
BP	GO:0071621	granulocyte chemotaxis	4/23	123/18670	1.44E-05	0.001054195	0.000671498	PPBP/SAA1/CXCL1/CXCL2	4
BP	GO:0046683	response to organophosphorus	4/23	134/18670	2.02E-05	0.001312959	0.000836325	SLC26A3/AQP8/PCK1/COL1A1	4
BP	GO:0007586	digestion	4/23	139/18670	2.33E-05	0.001312959	0.000836325	GUCA2A/SST/GUCA2B/SI	4
BP	GO:0097530	granulocyte migration	4/23	141/18670	2.47E-05	0.001312959	0.000836325	PPBP/SAA1/CXCL1/CXCL2	4
BP	GO:0060326	cell chemotaxis	5/23	304/18670	2.93E-05	0.001379569	0.000878754	CHGA/PPBP/SAA1/CXCL1/CXCL2	5
BP	GO:0014074	response to purine-containing compound	4/23	149/18670	3.07E-05	0.001379569	0.000878754	SLC26A3/AQP8/PCK1/COL1A1	4
BP	GO:0071320	cellular response to cAMP	3/23	53/18670	3.67E-05	0.001535611	0.000978149	SLC26A3/AQP8/PCK1	3
BP	GO:0015698	inorganic anion transport	4/23	169/18670	5.02E-05	0.00195619	0.001246048	SLC4A4/CLCA4/SLC26A3/CLCA1	4
BP	GO:0031284	positive regulation of guanylate cyclase activity	2/23	10/18670	6.49E-05	0.002374256	0.001512346	GUCA2A/GUCA2B	2
BP	GO:0030198	extracellular matrix organization	5/23	368/18670	7.27E-05	0.002393714	0.001524741	MMP1/TTR/SPP1/TGFBI/COL1A1	5
BP	GO:0043062	extracellular structure organization	5/23	369/18670	7.37E-05	0.002393714	0.001524741	MMP1/TTR/SPP1/TGFBI/COL1A1	5
BP	GO:0031282	regulation of guanylate cyclase activity	2/23	12/18670	9.51E-05	0.002808445	0.001788915	GUCA2A/GUCA2B	2
BP	GO:0061844	antimicrobial humoral immune response mediated by antimicrobial peptide	3/23	73/18670	9.60E-05	0.002808445	0.001788915	PPBP/CXCL1/CXCL2	3
BP	GO:0070098	chemokine-mediated signaling pathway	3/23	88/18670	0.000167377	0.004450694	0.00283499	PPBP/CXCL1/CXCL2	3
BP	GO:1902476	chloride transmembrane transport	3/23	88/18670	0.000167377	0.004450694	0.00283499	CLCA4/SLC26A3/CLCA1	3
BP	GO:1990868	response to chemokine	3/23	97/18670	0.000223266	0.0054421	0.003466493	PPBP/CXCL1/CXCL2	3
BP	GO:1990869	cellular response to chemokine	3/23	97/18670	0.000223266	0.0054421	0.003466493	PPBP/CXCL1/CXCL2	3
BP	GO:0006821	chloride transport	3/23	108/18670	0.000306431	0.006609567	0.004210143	CLCA4/SLC26A3/CLCA1	3
BP	GO:0031281	positive regulation of cyclase activity	2/23	22/18670	0.000330357	0.006609567	0.004210143	GUCA2A/GUCA2B	2
BP	GO:0098661	inorganic anion transmembrane transport	3/23	111/18670	0.000332136	0.006609567	0.004210143	CLCA4/SLC26A3/CLCA1	3
BP	GO:0043200	response to amino acid	3/23	113/18670	0.000350022	0.006609567	0.004210143	SST/PCK1/COL1A1	3
BP	GO:0010447	response to acidic pH	2/23	23/18670	0.000361549	0.006609567	0.004210143	SST/PCK1	2
BP	GO:0030810	positive regulation of nucleotide biosynthetic process	2/23	23/18670	0.000361549	0.006609567	0.004210143	GUCA2A/GUCA2B	2
BP	GO:0051349	positive regulation of lyase activity	2/23	23/18670	0.000361549	0.006609567	0.004210143	GUCA2A/GUCA2B	2
BP	GO:1900373	positive regulation of purine nucleotide biosynthetic process	2/23	23/18670	0.000361549	0.006609567	0.004210143	GUCA2A/GUCA2B	2
BP	GO:0098656	anion transmembrane transport	4/23	291/18670	0.000405183	0.007182789	0.004575272	SLC4A4/CLCA4/SLC26A3/CLCA1	4
BP	GO:0071377	cellular response to glucagon stimulus	2/23	26/18670	0.000463397	0.007973163	0.005078722	PCK1/GCG	2
BP	GO:0009914	hormone transport	4/23	317/18670	0.000559542	0.009352338	0.005957225	CHGA/TTR/SPP1/GCG	4
BP	GO:0032496	response to lipopolysaccharide	4/23	330/18670	0.000650676	0.01057348	0.006735064	PPBP/PCK1/CXCL1/CXCL2	4
BP	GO:0002237	response to molecule of bacterial origin	4/23	343/18670	0.000751923	0.011888518	0.007572713	PPBP/PCK1/CXCL1/CXCL2	4
BP	GO:0006959	humoral immune response	4/23	356/18670	0.000863899	0.013299493	0.008471472	CHGA/PPBP/CXCL1/CXCL2	4
BP	GO:0033762	response to glucagon	2/23	36/18670	0.000891575	0.013373619	0.008518688	PCK1/GCG	2
BP	GO:0009268	response to pH	2/23	40/18670	0.001100553	0.015703012	0.010002459	SST/PCK1	2
BP	GO:1900371	regulation of purine nucleotide biosynthetic process	2/23	40/18670	0.001100553	0.015703012	0.010002459	GUCA2A/GUCA2B	2
BP	GO:0048545	response to steroid hormone	4/23	383/18670	0.001133428	0.015728687	0.010018813	SST/SPP1/PCK1/COL1A1	4
BP	GO:0030808	regulation of nucleotide biosynthetic process	2/23	41/18670	0.001156126	0.015728687	0.010018813	GUCA2A/GUCA2B	2
BP	GO:0015701	bicarbonate transport	2/23	42/18670	0.001213024	0.0161277	0.010272975	SLC4A4/SLC26A3	2
BP	GO:0031279	regulation of cyclase activity	2/23	43/18670	0.001271243	0.016526165	0.010526788	GUCA2A/GUCA2B	2
BP	GO:0045981	positive regulation of nucleotide metabolic process	2/23	46/18670	0.001453805	0.017718246	0.011286116	GUCA2A/GUCA2B	2
BP	GO:0051339	regulation of lyase activity	2/23	46/18670	0.001453805	0.017718246	0.011286116	GUCA2A/GUCA2B	2
BP	GO:1900544	positive regulation of purine nucleotide metabolic process	2/23	46/18670	0.001453805	0.017718246	0.011286116	GUCA2A/GUCA2B	2
BP	GO:0043434	response to peptide hormone	4/23	436/18670	0.001826249	0.021803172	0.01388812	SLC30A10/PCK1/COL1A1/GCG	4
BP	GO:0071222	cellular response to lipopolysaccharide	3/23	205/18670	0.001964707	0.022987067	0.014642234	PPBP/CXCL1/CXCL2	3
BP	GO:0019935	cyclic-nucleotide-mediated signaling	3/23	212/18670	0.002161811	0.024320372	0.015491519	CHGA/GUCA2B/GCG	3
BP	GO:0071219	cellular response to molecule of bacterial origin	3/23	212/18670	0.002161811	0.024320372	0.015491519	PPBP/CXCL1/CXCL2	3
BP	GO:0007584	response to nutrient	3/23	219/18670	0.002370845	0.026168759	0.016668899	SST/SPP1/COL1A1	3
BP	GO:0007588	excretion	2/23	63/18670	0.002708599	0.029343156	0.018690917	GUCA2B/SLC26A3	2
BP	GO:0071216	cellular response to biotic stimulus	3/23	236/18670	0.002929681	0.031143503	0.019837697	PPBP/CXCL1/CXCL2	3
BP	GO:0031667	response to nutrient levels	4/23	499/18670	0.002981258	0.031143503	0.019837697	SST/SPP1/PCK1/COL1A1	4
BP	GO:0071300	cellular response to retinoic acid	2/23	69/18670	0.003239085	0.03324324	0.021175181	PCK1/COL1A1	2
BP	GO:0046883	regulation of hormone secretion	3/23	262/18670	0.003930896	0.039647826	0.025254756	CHGA/SPP1/GCG	3
BP	GO:0051453	regulation of intracellular pH	2/23	84/18670	0.004759422	0.047190879	0.030059507	SLC4A4/SLC26A3	2
CC	GO:0034774	secretory granule lumen	4/24	321/19717	0.000566363	0.011535838	0.00745622	TTR/PPBP/CXCL1/GCG	4
CC	GO:0060205	cytoplasmic vesicle lumen	4/24	325/19717	0.000593329	0.011535838	0.00745622	TTR/PPBP/CXCL1/GCG	4
CC	GO:0031983	vesicle lumen	4/24	327/19717	0.000607149	0.011535838	0.00745622	TTR/PPBP/CXCL1/GCG	4
CC	GO:0045177	apical part of cell	4/24	384/19717	0.001105281	0.01575026	0.010180223	CLCA4/SLC26A3/AQP8/SI	4
CC	GO:1904724	tertiary granule lumen	2/24	55/19717	0.002027219	0.023110296	0.014937403	PPBP/CXCL1	2
CC	GO:0005788	endoplasmic reticulum lumen	3/24	309/19717	0.00604535	0.046625468	0.030136498	SPP1/COL1A1/GCG	3
CC	GO:0005903	brush border	2/24	99/19717	0.00640979	0.046625468	0.030136498	SLC26A3/SI	2
CC	GO:0016324	apical plasma membrane	3/24	318/19717	0.006543925	0.046625468	0.030136498	CLCA4/SLC26A3/SI	3
MF	GO:0048018	receptor ligand activity	9/22	482/17696	2.78E-09	1.38E-07	6.89E-08	GUCA2A/SST/TTR/PPBP/SAA1/SPP1/CXCL1/GCG/CXCL2	9
MF	GO:0030546	signaling receptor activator activity	9/22	487/17696	3.04E-09	1.38E-07	6.89E-08	GUCA2A/SST/TTR/PPBP/SAA1/SPP1/CXCL1/GCG/CXCL2	9
MF	GO:0005179	hormone activity	4/22	122/17696	1.43E-05	0.000433406	0.000215575	GUCA2A/SST/TTR/GCG	4
MF	GO:0001664	G protein-coupled receptor binding	5/22	280/17696	2.02E-05	0.000449057	0.00022336	PPBP/SAA1/CXCL1/GCG/CXCL2	5
MF	GO:0015103	inorganic anion transmembrane transporter activity	4/22	142/17696	2.60E-05	0.000449057	0.00022336	SLC4A4/CLCA4/SLC26A3/CLCA1	4
MF	GO:0008009	chemokine activity	3/22	49/17696	2.96E-05	0.000449057	0.00022336	PPBP/CXCL1/CXCL2	3
MF	GO:0042379	chemokine receptor binding	3/22	66/17696	7.25E-05	0.000942998	0.000469044	PPBP/CXCL1/CXCL2	3
MF	GO:0005125	cytokine activity	4/22	220/17696	0.000142613	0.001594252	0.000792977	PPBP/SPP1/CXCL1/CXCL2	4
MF	GO:0005229	intracellular calcium activated chloride channel activity	2/22	16/17696	0.000175193	0.001594252	0.000792977	CLCA4/CLCA1	2
MF	GO:0061778	intracellular chloride channel activity	2/22	16/17696	0.000175193	0.001594252	0.000792977	CLCA4/CLCA1	2
MF	GO:0015106	bicarbonate transmembrane transporter activity	2/22	19/17696	0.000249086	0.00189131	0.000940733	SLC4A4/SLC26A3	2
MF	GO:0015108	chloride transmembrane transporter activity	3/22	100/17696	0.000249404	0.00189131	0.000940733	CLCA4/SLC26A3/CLCA1	3
MF	GO:0008509	anion transmembrane transporter activity	4/22	336/17696	0.00071252	0.00498764	0.002480839	SLC4A4/CLCA4/SLC26A3/CLCA1	4
MF	GO:0022839	ion gated channel activity	2/22						

Table S5 KEGG pathway enrichment analysis of hub genes

ID	Description	Gene ratio	Bg ratio	P value	p.adjust	q value	Gene ID	Count
hsa04972	Pancreatic secretion	4/16	102/8104	3.83E-05	0.002031543	0.001291149	SLC4A4/CLCA4/SLC26A3/CLCA1	4
hsa05323	Rheumatoid arthritis	3/16	93/8104	0.000735285	0.008501102	0.005402885	MMP1/CXCL1/CXCL2	3
hsa04657	IL-17 signaling pathway	3/16	94/8104	0.000758609	0.008501102	0.005402885	MMP1/CXCL1/CXCL2	3
hsa04964	Proximal tubule bicarbonate reclamation	2/16	23/8104	0.000902567	0.008501102	0.005402885	SLC4A4/PCK1	2
hsa04061	Viral protein interaction with cytokine and cytokine receptor	3/16	100/8104	0.000908527	0.008501102	0.005402885	PPBP/CXCL1/CXCL2	3
hsa05146	Amoebiasis	3/16	102/8104	0.000962389	0.008501102	0.005402885	CXCL1/COL1A1/CXCL2	3

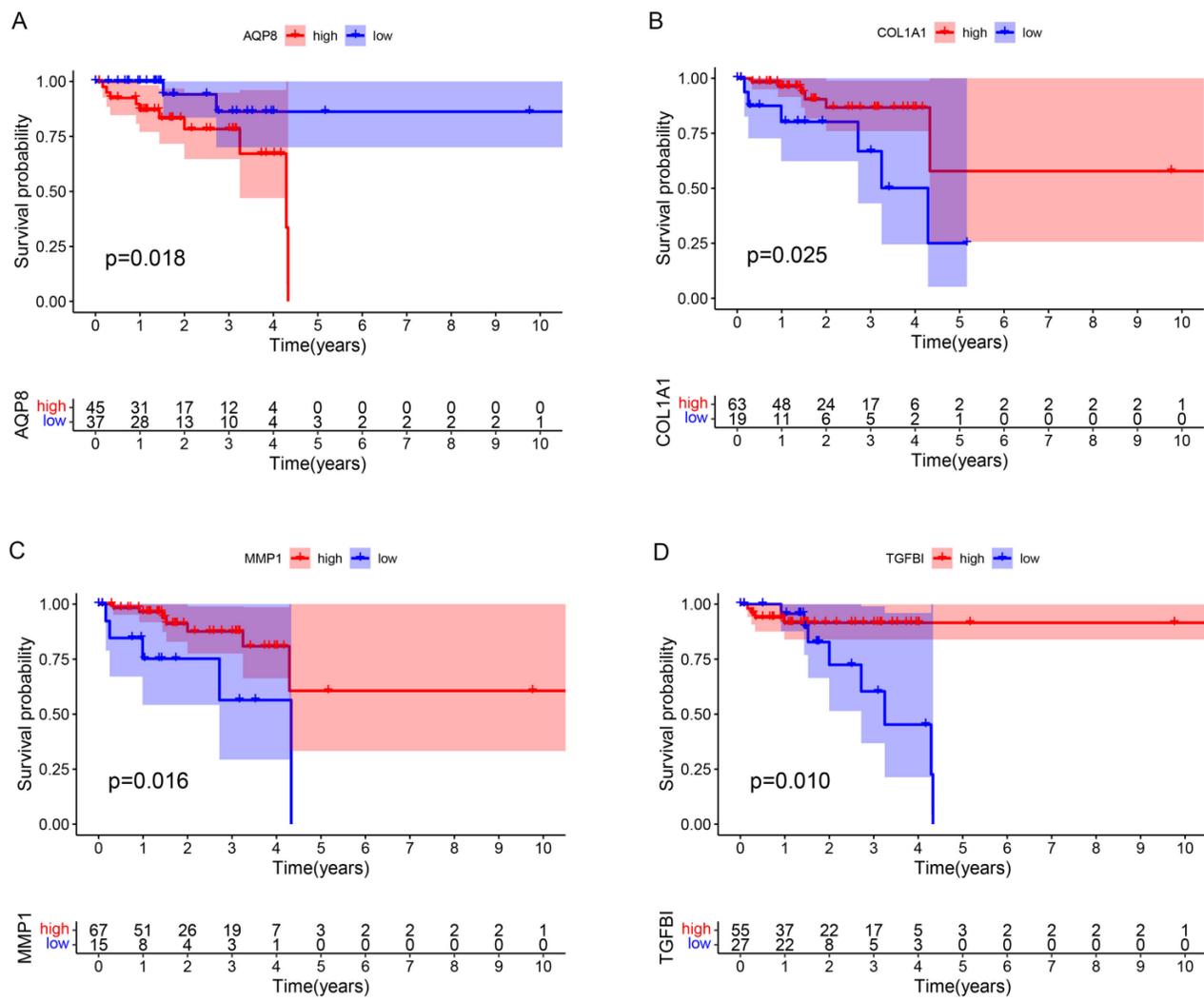


Figure S1 Survival analysis of the 24 hub genes. Gene changes of AQP8 (A), COL1A1 (B), MMP1 (C), and TGFBI (D) were significantly correlated with the overall survival of patients with rectal cancer ($P < 0.05$).

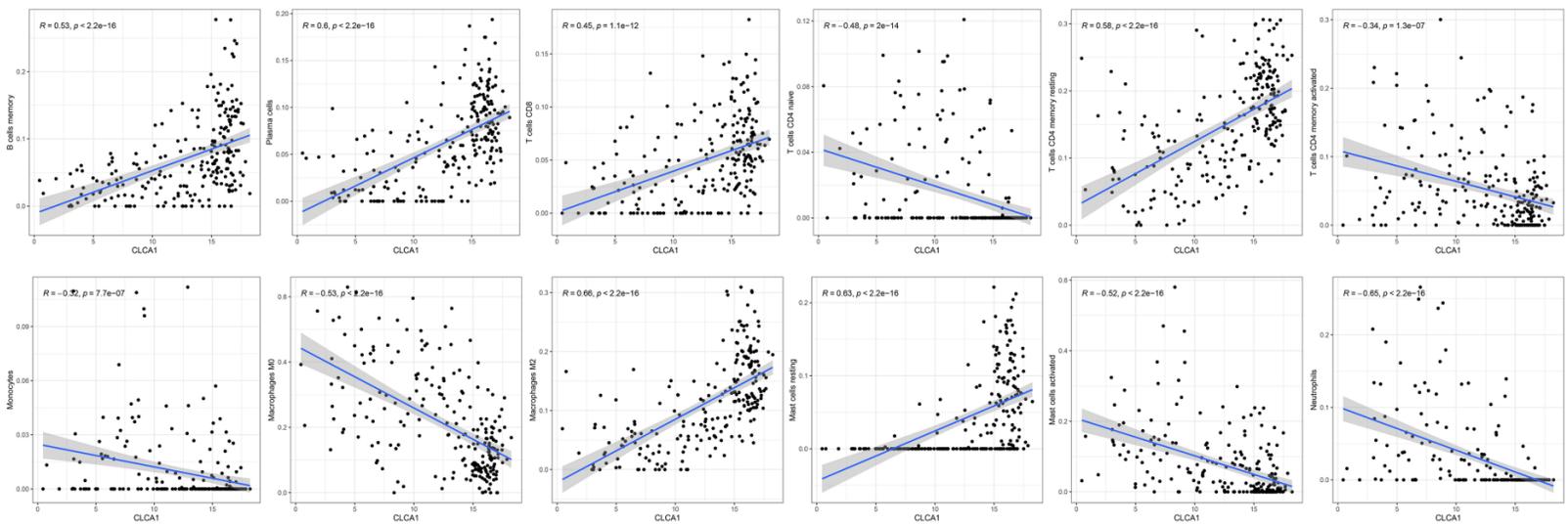


Figure S2 Relationship between the prognostic gene CLCA1 and immune cell infiltration.