

Hub Genes Related to the Pathogenesis and Progression of Colorectal Cancer and Adenoma by Integrative Bioinformatics Approaches

Yuxuan HUANG

Huaqiao University

Ge CUI (✉ dym@zjhu.edu.cn)

The First Affiliated Hospital of Huzhou University <https://orcid.org/0000-0002-9763-5701>

Primary research

Keywords: colorectal cancer (CRC), colorectal adenoma (CA), differential genes (DEGs), protein-protein interaction (PPI), function enrichment analysis, prognostic values

Posted Date: June 25th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-637295/v1>

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

[Read Full License](#)

Abstract

Aims: To utilize the bioinformatics to analyze the differentially expressed genes (DEGs), interaction proteins, perform gene enrichment analysis, protein-protein interaction network (PPI) and map the hub genes between colorectal cancer(CRC) and colorectal adenocarcinomas(CA).

Methods: We analyzed a microarray dataset (GSE32323 and GSE4183) from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) in tumor tissues and non-cancerous tissues were identified using the dplyr and Venn diagram packages of the R Studio software. Functional annotation of the DEGs was performed using the Gene Ontology (GO) website. Pathway enrichment (KEGG) used the WebGestalt to analyze the data and R Studio to generate the graph. We constructed a protein–protein interaction (PPI) network of DEGs using STRING and Cytoscape software was used for visualization. Survival analysis of the hub genes and was performed using the online platform GEPIA to determine the prognostic value of the expression of hub genes in cell lines from CRC patients. The expression of molecules with prognostic values was validated on the UALCAN database. The expression of hub genes was examined using the Human Protein Atlas.

Results: Applying the GEO2R analysis and R studio, we identified a total of 471 upregulated and 278 downregulated DEGs. By using the online database WebGestalt, we identified the most relevant biological networks involving DEGs with statistically significant differences in expression were mainly associated with biological processes involved in the cell proliferation, cell cycle transition, cell homeostasis and indicated the role of each DEGs in cell cycle regulation pathways. We found 10 hub genes with prognostic values were overexpressed in the CRC and CA samples.

Conclusion: we found out ten hub genes and three core genes closely associated with the pathogenesis and prognosis of CRC and CA, which is of great significance for colorectal tumor early detection and prognosis evaluation.

Highlights

- Ten hub genes (ADCY9, CCL19, CXCL12, CXCL13, GNG2, LPAR1, CCL5, PYY, P2RY14 and SST) closely associated with the pathogenesis of colorectal cancer and colorectal adenoma with cancerous potential.
- Three core genes (CCL5, CCL19 and CXCL13) have a close relationship with prognosis of colorectal cancer.

Introduction

Colorectal cancer (CRC) is a worldwide health problem, and has become the third most common cancer in the world.¹ Clinically, initial treatment includes surgery,² with minimally invasive surgery (MIS) being increasingly offered.³ For stage-II and stage-III colorectal cancer patients, surgical treatment is usually

supplemented by neoadjuvant radiation therapy,^{4,5} and for high-risk stage-II and stage-III colon cancer is commonly treated by adjuvant chemotherapy, immunotherapy.^{6,7} Relative survival has increased steadily over recent decades. However, low survival rates are still observed in many parts of world. Due to the high incidence and mortality of colorectal cancer, the exploration of early detection markers, new treatment targets and prognostic biomarkers for colorectal cancer has always been the focus of attention.⁸

Research found that CRC are sporadic and the vast majority of CRC develops from the adenoma-carcinoma sequence.⁹ Study testifies that patients with an advanced adenoma detected at colonoscopy were at remarkably increased risk of developing colorectal cancer compared with those with no adenoma,¹⁰ but the relationship between CA and long-term CRC incidence is still unclear.

The most commonly used high-throughput to analyze complicated disease pathogenesis is microarray-based gene expression assessment. However, studies performed that utilize human CRC and CA gene expression profilings are very uncommon. In this study, we aimed at exploring the differentially expressed genes (DEGs), gene network, pathways, and protein interactions between CRC and CA. To detect the DEGs between CRC (GSE32323) and CA (GSE4183) cell lines, we adopted a bioinformatics approach to analyze DEG data retrieved from the Gene Expression Omnibus (GEO) database and worked out on R Studio to perform Venndiagram. For the screened DEGs, functional annotation assessment with Gene Ontology (GO) and pathway enrichment assessment with the Kyoto Encyclopedia of Genes and Genomes (KEGG) were carried out using the online database of WebGestalt. Ultimately, we found 10 potential hub genes and three core genes that were strongly linked to OC. Moreover, the DEGs were utilized to construct a protein-protein interaction (PPI) network, and utilized to identify 10 hub genes. Further, we performed survival analysis on all the hub genes on GEPIA. Eventually, we found out 10 potential hub genes and 3 core genes that were closely linked to CRC.

Materials And Methods

Data Preprocessing and Screening of DEGs

The expression profile was performed on the CRC gene dataset (GSE32323) and CA gene dataset (GSE4183), which was obtained from GEO (Gene Expression Omnibus database, <http://www.ncbi.nlm.nih.gov/geo/>). “Colorectal Cancer” AND “Homo sapiens”, “Colorectal Adenoma” AND “Homo sapiens” were the keywords used to search CRC, CA-related expression profiles within the GEO datasets respectively.

The GSE32323 expression profiling was conducted in arrays that included 17 human CRC cell lines, 17 human normal cell lines and other 10 sources of cell lines (COLO320 cell line, HCT116 cell line, HT29 cell line, RKO cell line, SW480 cell line). We selected some of these datasets (GSM800742 to GSM800775) to utilize the GEO2R (<http://www.ncbi.nlm.nih.gov/geo/geo2r/>) statistical tool to recalculate and assess the genes that were expressed differently between the human CRC cell lines and the human normal cell lines.

The GSE4183 expression profiling was conducted in arrays that included 15 human CRC cell lines, 8 human normal cell lines, 15 human CA cell lines, and 15 human inflammatory bowel diseases (IBD) cell lines. We selected some of these datasets (GSM95473 to GSM95495) to utilize the GEO2R (<http://www.ncbi.nlm.nih.gov/geo/geo2r/>) statistical tool to recalculate and assess the genes that were expressed differently between the human CA cell lines and the human normal cell lines.

The Benjamini and Hochberg (false discovery rate) and t-test methods were utilized with the GEO2R tool to calculate the FDR and p-values, respectively, to identify the DEGs. We considered $p < 0.05$ and $|\log(\text{fold change})| > 1$ to be statistically significant for the DEGs, and $\log\text{FC} \geq 1$ and $\log\text{FC} \leq -1$ were considered to indicate upregulated and downregulated DEGs, respectively.¹¹

Utilizing all of the DEGs identified in the CRC and CA cell lines, two volcano plots were constructed by using the Volcano Plot (<https://paolo.shinyapps.io/ShinyVolcanoPlot/>) online server (Fig. 1A.1B). The resultant DEG dataset was collected and used for further analysis.

We designed a series of codes in the R Studio to analyze the differential expression genes between CRC and CA, and analyze the up-regulated genes and down-regulated genes to find out the genes that have common influence, and further produced Venn diagram of those DEGs (Fig. 1C).

Construction of PPI Network and Screening for the Hub Genes

In order to evaluate the protein-protein interaction network (PPI) between the targeted DEGs, we used the online database STRING (v11.0, <http://www.string-db.org/>) to visualize the PPIs between the statistically significant DEG-encoded proteins in the resultant dataset.¹² To ensure the accuracy of PPI network, we applied a cutoff score of ≥ 0.7 (high confidence interaction score) to obtain the significant PPIs. Then, the resulting PPI networks were exported from the website and imported into Cytoscape software v3.8.2 for visualization, which based on the log fold change values, the PPI networks were plotted for both the up-regulated and down-regulated DEGs.¹³ (Fig. 2A.2B)

In order to simplify the PPI network, we excluded the non-interacting genes. The top 10 genes with the highest degree of connection to the others were considered as hub genes based on the analysis using CytoHubba from Cytoscape (Fig. 2C).

Analysis of Gene Enrichment

Some statistically target DEGs were further imported into the online analysis tool WebGestalt (<http://www.webgestalt.org/option.php>) to conduct and perform the GO function and pathway enrichment analysis. Importing the obtained DEGs and protein list into WebGestalt, which can quickly analyze and output protein networks, metabolic pathways and maps. We applied the Gene Ontology and KEGG Pathway Function Database to identify the enriched pathways involving DEGs in terms of the hypergeometric distribution, and the p-values were calculated by using the default database, which based

on an FDR $p < 0.005$ and significant p -value < 0.05 . Eventually, the graphical descriptions of the molecular interactions between the DEGs were output.

Genetic Alterations of Hub Genes

The datasets of colorectal/colon adenocarcinoma in cBioPortal, including the data of 3037 samples, was selected for the calculation of the genetic alterations in hub genes using cBioPortal. This platform allows for the rapid, intuitive, and high-quality access to molecular profiles and transforms these datasets into a biologic visualization and clinical applications.¹⁴ The genomic alterations included gene mutations, copy number variations, mRNA expression z-scores with a z-score threshold of ± 2.0 and protein expression z-scores.

Survival Analysis of Hub Genes

An interactive web server platform called Gene Expression Profiling Interactive Analysis (GEPIA2, <http://gepia.cancer-pku.cn/>) provides comprehensive and customized analysis of functionalities based on TCGA (The Cancer Genome Atlas) and genotype-tissue expression (GTEx) data. GEPIA performs overall survival (OS) or disease free survival (DFS, also called relapse-free survival and RFS) analysis based on gene expression. It utilizes Log-rank test, a.k.a the Mantel–Cox test, for hypothesis test, and the 95% confidence interval information can also be included in the survival plot.

Results

Identification of DEGs

The GSE32323 and GSE4183 were obtained on GEO database. The GSE 32323 dataset was obtained from 44 patient cell lines that comprised 34 samples, including 17 CRC cell lines and 17 normal cell lines, and the GSE 4183 dataset was obtained from 53 patient cell lines that comprised 23 samples, including 15 CA cell lines and 8 normal cell lines. To identify the DEGs from these two groups (CRC and Normal cell line) we conducted GEO2R web-server analysis to calculate the p -values and $|\log_{2}FC|$ values and imported the results into R Studio. According to the results analyzed by R package, 749 DEGs (we eliminated mRNAs and lncRNAs) were differentially expressed ($|\log_{2}FC| \geq 1$ and adjusted $P > 0.05$). Out of these, 471 DEGs were over-expressed while 278 DEGs were downregulated (Table 1).

We constructed a Venn diagram using the package downloaded in R Studio to verify the same overlapping genes (Fig. 1C).

Construction of the PPI Network

In order to assess the PPIs between the DEGs, we applied the STRING to identify the PPI networks for both the up- and downregulated genes. A combined score of ≥ 0.7 for the nodes was considered to indicate a significant PPI interaction. The processed PPI network was exported in "txt" format. Then, it was transferred into "csv" file and imported into Cytoscape v3.8.2 software for visualization.

The graphical PPI networks of the up- and downregulated DEGs are shown in Figs. 2A.2B. The backbone network of the up- and downregulated genes consist of 186 nodes and 138 nodes with clustering coefficients of 0.437 and 0.573, respectively. The Cytoscape plug-in Network Analyzer was used to analyze the networks for both the up- and downregulated DEGs. Table 2 shows the topological parameters of the up- and downregulated PPI networks.

We evaluating the degree of connection of up-regulated DEGs by Cytohubba in Cytoscape, and identified 10 hub genes (ADCY9, CCL19, CXCL12, CXCL13, GNG2, LPAR1, CCL5, PYY, P2RY14, SST). The hub genes shown with red nodes (ADCY9, GNG2, and LPAR1) had the strongest connections with others, while those shown with orange (PYY, SST) and yellow nodes (CCL19, CXCL12, CXCL13, P2RY14) had the weakest connections, respectively. (Fig. 2C). The Table 3 demonstrates the functional roles of 10 hub genes.

Analysis of GO and Enrichment

To explore the potential GO classifications and KEGG pathway-enriched genes from the dataset, we imported all target DEGs into the online analysis tool WebGestalt (<http://www.webgestalt.org/option.php>) to execute the annotation process, the significant level is top 10 and FDR < 0.05. According to the website, the input 472 users IDs in which 459 users IDs are unambiguously mapped to 459 unique entrezgene IDs and 13 user IDs can't be mapped to any entrezgene ID. The annotated results for the GO was divided according to the BP (biological process), CC (cell component), and MF (molecular function), ($p < 0.05$, FDR < 0.05). The results of the GO biological process (BP) analysis revealed that the upregulated DEGs were mainly enriched in the adjustment of the homeostasis of various ions, maintaining cellular homeostasis, and regulation of cellular component movement; the downregulated DEGs were mainly elevated in antimicrobial humoral immune response mediated by antimicrobial peptide, and are effective in regulating cell proliferation, cell cycle and cell cycle transition (Fig. 3A). For the GO molecular function (MF) analysis, the upregulated DEGs were significantly enriched in xenobiotic transmembrane transporting ATPase activity, carbonate dehydratase activity, phosphoric diester hydrolase activity, and extracellular matrix structural constituent, whereas the downregulated DEGs were largely enriched in cytokine activity, chemokine activity, cyclin – dependent protein serine/threonine kinase activity, cyclin – dependent protein kinase activity, and chemokine receptor binding (Fig. 3B). Concerning the GO cell component (CC) analysis, the upregulated DEGs were mostly enriched in the brush border and brush border membrane, cluster of actin – based cell projections, and microvillus, while the downregulated DEGs were enriched in preribosome, small subunit precursor, cyclin – dependent protein kinase holoenzyme complex, mitotic spindle, serine/threonine protein kinase complex (Fig. 3C). Moreover, we used the WebGestalt to categorize the DEGs involved in various signaling pathways according to the KEGG reference pathways ($p < 0.05$, FDR < 0.05). By examining the KEGG pathways, we noticed that the upregulated DEGs were enriched in nitrogen metabolism, proximal tubule bicarbonate reclamation, aldosterone – regulated sodium reabsorption, mineral absorption, and bile secretion (Fig. 3D). In total, the GO and KEGG pathway enrichment analysis of both up- and downregulated DEGs were sorted out (Table 4).

Survival Analysis and Alterative Frequency of Hub Genes

GEPIA survival analysis was used to evaluate the overall association with survival of 10 hub genes (ADCY9, CCL19, CXCL12, CXCL13, GNG2, LPAR1, CCL5, PYY, P2RY14, SST) were significant related to poor prognosis and survival times in CRC patients. In conclusion, we discovered that the high expression of CCL19 (HR = 0.88) (Fig. 4B), CXCL13 (HR = 0.88) (Fig. 4C), GNG2 (HR = 0.98) (Fig. 4E), LPAR1 (HR = 0.73) (Fig. 4F), CCL5 (HR = 0.77) (Fig. 4G), P2RY14 (HR = 0.83) (Fig. 4I), and PYY (HR = 0.80) (Fig. 4H) were associated with improved overall survival in the CRC cell line. However, the high expression of ADCY9 (HR = 1) (Fig. 4A), CXCL12 (HR = 1.1) (Fig. 4D) and SST (HR = 1.1) (Fig. 4J) were linked with worse overall survival in the CRC cell line. The frequencies of genetic alterations of the 10 hub genes in CRC were analyzed using the cBioPortal database. Approximately 6.4% of CRC clinical cases exhibited significant alterations in the 10 hub genes (Fig. 5A). Moreover, the interactions between those ten hub genes were calculated in the GeneMANIA (Fig. 5B). Hierarchical clustering showed that the hub genes could basically differentiate the colorectal cancer samples from the noncancerous samples (Fig. 5C). Furthermore, the box plot analysis in GEPIA calculated the level of expression of the core genes in 275 CRC tissue samples and 349 normal tissue samples. The boxplot in (Fig. 6B, 6C, 6E) shows a considerable increase in the level of core gene expression (CCL5, CCL19 and CXCL13) in the CRC cell line. Moreover, the expression levels of ten hub genes between tumor and normal tissues were also compared in the Oncomine to depict an overall graph (Fig. 6K), which was in consistent with the findings in GEPIA. Combined together, the results show that CCL5, CCL19 and CXCL13 work as core genes that have a close relationship with CRC.

KEGG Analysis of the Hub Genes

After the verification of 10 hub genes applying Kaplan–Meier Overall Survival Analysis and GEPIA, KEGG pathways of these hub genes were re-analyzed and transformed into another form via Uniprot to better understand their functions; we identified six pathways associated with the 10 hub genes (Table 5). CXCL13, ADCY9, GNG2, CCL5, CCL19 and CXCL12 were enriched in the chemokine signaling pathway (Fig. 7A), ADCY9, LPAR1, GNG2 and CXCL12 were enriched in the pathway in cancer (Fig. 7D), and CXCL13, CCL5, CCL19 and CXCL12 were enriched in viral protein interaction with cytokine and cytokine receptor (Fig. 7C), ADCY9, GNG2, CCL5 and CXCL12 were enriched in the human cytomegalovirus infection (Fig. 7F), CXCL13, CCL5, CCL19 and CXCL12 were enriched in the cytokine-cytokine receptor interaction (Fig. 7B), LPAR1, PYY, SST, P2RY14 were enriched in the neuroactive ligand-receptor interaction ($P > 0.05$) (Fig. 7E). Research has proved that chemokine signaling pathway and cytokine-cytokine receptor interaction play important roles in the progression of CRC.^{15,16} Therefore, CXCL13, CCL5, CCL19 may play important roles in the occurrence and development of CRC, which is in accordance with the results proposed by survival analysis.

Discussion

In spite of continuously progress in diagnostic and therapeutic methods, colorectal cancer still remains one of the deadliest causes in malignant tumor of human digestive system. Due to relapse, drug resistance and other reasons, the long-term survival rate of CRC patients is still unsatisfactory. Early diagnosis plays a vital role in the prevention and prognosis of colorectal cancer. Currently, biomarker performs as a significant part in the detection and treatment of CRC patients. CEA has been most widely accepted as tumor marker and used in monitoring the treatment effect of CRC patients.¹⁷ However, CEA has specific expression in not only CRC, but also in other neoplasms too, e.g. gastric and pancreatic cancers, and in inflammatory conditions. Furthermore, CEA is unable to differentiate benign and malignant polyps.^{18,19} Therefore, it is of of great urgency to study the expression profiles and discover the effective target genes of CRC.

In the present study, the microarray dataset of GSE32323 was selected to identify DEGs between CRC and normal cells with a total of seventeen CRC samples and seventeen normal samples, and GSE4183 was selected to identify DEGs between CA and normal cells with a total of fifteen CRC samples and eight normal samples. A total of 749 DEGs were screened, including 471 upregulated and 278 downregulated genes. Network has been confirmed as reliable method, which allows researchers to gain insight into the organization and structure of interacting proteins. A comprehensive set of topological parameters are computed and displayed through this application, which includes the number of nodes, edges, and connected components, the network centralization, heterogeneity, and clustering coefficient, average clustering coefficients, and shortest path lengths.²⁰ These parameters were applied to analyze the nodes in PPI networks to imply the significance of their functions in networks with different characteristics. Furthermore, GO and KEGG were carried out to determine the biological process (BP), cell component (CC), molecular function (MF) and pathways of these DEGs. For BP, the DEGs were significantly associated with ion homeostasis, inorganic ion homeostasis, cation homeostasis, cellular response to endogenous stimulus and chemical homeostasis. In terms of CC, the upregulated DEGs were particularly enriched in brush border membrane, cluster of actin-based cell projections, apical plasma of membrane and collagen-containing extracellular matrix. Additionally, KEGG pathway enrichment showed remarkable involvement of DEGs in pathways of nitrogen metabolism, proximal tubule bicarbonate reclamation, aldosterone-regulated sodium reabsorption, mineral absorption and bile secretion and thyroid hormone signaling pathway. According to the MF, DEGs mainly related to phosphoric diester hydrolase activity, extracellular matrix structural constituent and carbonate dehydratase activity. A phosphodiester hydrolase is a group of isoenzyme, which hydrolyze cyclic nucleotides, thereby reducing the levels of cAMP and cGMP in cells, leading to tumorigenic effects.²¹ cAMP plays an important role in cell proliferation, apoptosis and cell cycle regulation in many tumor cells. Moreover, some in vitro findings have demonstrated the role of cAMP in the treatment of CRC.²² cGMP commonly presents as intracellular second messenger and involved in various cellular processes as like cell growth, ion channel conductance, apoptosis, cell mobility and contractility.²³ The effects of anticancer in cGMP are mostly associated with the activation of PKG and downstream effector pathways.²⁴ Increasingly evidence implies that, the effective inhibition of phosphoric diester hydrolase greatly contributes to the therapeutic opportunities of patients with CRC.^{25,26} Therefore, the results we obtained are consistent with the

function of upregulated genes in pathways that cause CRC. Simultaneously, the KEGG assessment revealed that the downregulated genes were mainly enriched in cell cycle, IL-17 signaling pathway and p53 signaling pathway. Interleukin 17 is a proinflammatory cytokine, which is associated with cancer progression.²⁷ Many researches have proven that IL-17 plays a vital role in metastasis and prognosis of CRC,²⁸ specifically, its impacts on the tumorigenesis, angiogenesis, and metastasis of CRC.^{29,30} Thus, the regulation of IL-17 affects the occurrence of CRC, which is consistent with the results of our analysis of downregulated DEGs in KEGG pathway. Furthermore, p53 is the most commonly mutated tumor suppressor gene in various kinds of malignant tumors,³¹ and its mutation are generally considered to occur in the transition from adenoma to cancer.³² Meanwhile, discoveries found that those CRC patients with p53 mutations were likely to receive a worse outcome and shorter survival time.³³ Thus, p53 utilized clinically as a therapeutic target for CRC to improve the prognosis of patients. Generally speaking, all the theories provided above were consistent with our bioinformatic analysis result.

After more in-depth analysis of DEGs in PPI network, ten hub genes including ADCY9, CCL19, CXCL12, CXCL13, GNG2, LPAR1, CCL5, PYY, P2RY14 and SST were screened, which expressed significantly different in colon cancer tissues compared with normal colon tissues. In addition, the expression levels of the hub genes between CRC tissues and normal tissues were examined in the GEPIA2. Using the data from Kaplan Meier plots on GEPIA2 to generate overall survival curves, we discovered that high expression of CCL19, CXCL13, GNG2, LPAR1, CCL5, PYY and P2RY14 were associated with better survival, and high expression of ADCY9, CXCL12 and SST were associated with decreased survival in the CRC cell line. On the whole, CCL5, CCL19 and CXCL13 were confirmed as core genes that were strongly associated with overall survival in CRC. Thus, these three genes could greatly contribute to CRC metastasis.

C-C motif chemokine ligand 5 (CCL5), also known as RANTES (Regulated upon Activation, Normal T-cells Expressed, and Secreted), belongs to CC chemokine subfamily,³⁴ which induces tumor proliferation and infiltration, promotes angiogenesis, and stimulates tumor cell metastasis. The interactions between CCL5 and its receptor CCR5 may result in tumor development and further may be associated with cancer progression and metastasis,³⁵ in accordance with the study that those two genes are overexpressed within the primary as well as liver and pulmonary metastases compared to the healthy tissues.³⁶ CCL5 favors the in vitro growth and the migratory responses of CRC cells from both human and mouse origins.³⁵ More recently, a new mechanism of immune escape mediated by CCL5 was proposed by Chang et al.³⁷ The knockout of CCL5 from CT26 mouse CRC cells suppresses apoptosis of tumor-infiltrating CD8 + T cells and reduces tumor growth in mice, which proved that CCL5 is correlated with higher levels of apoptosis of CD8 + T cells and infiltration of T-regulatory cells (T-reg). Thus, this research demonstrated the potential value of CCL5 as a therapeutic target. There are two possible ways to apply CCL5 and CCR5 as therapeutic targets in clinical use. Firstly, the inhibitors of CCL5/CCR5 interactions. Quantities of specific small molecule CCR5 antagonists that are being used as antiviral therapies, which are also effective in blocking CCR5 signal transduction, such as maraviroc and vicriviroc.³⁸⁻⁴⁰ Secondly, the inhibitors of CCL5 secretion. Zoledronic acid significantly influences the secretion of CCL5 and

interleukin 6 in bone marrow-derived mesenchymal stem cells (MSCs)⁴¹ meaning that the drug could have effect on antitumor activity by affecting the ability of MSCs to interact with cancer cells. Generally, our current research leads us to speculate the CCL5/CCR5 axis as a potential therapeutic target in CRC. However, applying this axis into practical application requires further research to clarify the effects of CCL5 on cancer progression and the formation of an immunosuppressive microenvironment to insure that such treatments are supported by the appropriate reasons.

C-C motif chemokine ligand 19 (CCL19), also named as macrophage inflammatory protein 3-beta (MIP-3b) or Epstein-Barr virus-induced molecule 1 ligand chemokine (ELC), which receptor is CCR7.^{42,43} This type of cytokine may mainly play a role in normal lymphocyte recirculation and homing. Moreover, it greatly contributes to the transport of T cells in thymus and the migration of T cells and B cells to secondary lymphoid organs. Some research demonstrated that the CCL19/CCR7 axis reduces tumor cells metastasis to regional lymph nodes to some extent,⁴⁴ but it also acts as antitumor responses in some sorts of solid tumors.⁴⁵ Therefore, the CCL19 may perform both good and bad effects in cancer prognosis. Jun Lu et al. proved that CCL19 have positive effect on suppressing the growth and invasion of CRC proliferation,⁴⁶ which may work through inhibiting wnt/ β -catenin signaling pathway.⁴⁷ Furthermore, the mechanism of CCL19 that inhibits the CRC was revealed. As dendritic cells (DCs) mediate antitumor immunity by T-cell responses, the recruitment of dendritic cells (DCs) is regarded as a useful method in antitumor immunotherapy. According to the report, CCL19 is able to recruit DCs, which was highly expressed its specific receptor (CCR7).^{48,49} For the reason that, CCL19 is expected to be designed as an antitumor targeted drug and applied to the clinic to improve the survival rate and prognosis of patients.

C-X-C motif chemokine ligand 13, originally named B cell attracting chemokine 1 (BCA-1), which functions in the homing of B lymphocytes to follicles.⁵⁰ CXCL13 is the only ligand for CXCR5, which is a member of the G protein-coupled receptors (GPCR) family. The CXCL13-CXCR5 axis is demonstrated to participate in mediating inflammatory diseases. Besides, it also take part in the development of tumor, as several researches have proved that CXCL13 and its receptor CXCR5 can be applied as new target for the detection and treatment of lymphoma.⁵¹ And moreover, the axis is overexpressed in tumor tissue and peripheral blood of breast cancer patients,⁵² which is closely related to the poor prognosis of breast cancer patients.⁵³ In recent years, scientists discovered that CXCL13 and CXCR5 are overexpressed in colon cancer tissues, and have high correlation with poor overall (OA) survival of CRC patients.⁵⁴ Zhengyu Zhu et al.⁵⁵ proved that the knockdown of CXCR5 weakened the CXCL13 mediated growth of SW260 cells in colon cancer tissues, and further indicated that CXCL13-CXCR5 axis contributes to the growth of colon cancer cells. In their study, CXCL13 strengthened the migration and invasion of colon cancer cells with the help of CXCR5 in mainly two ways. The first one is influencing the expression of MMPs, a family of zinc-binding endopeptidases, which are regarded as effective markers for tumor invasion and metastasis.⁵⁶ With the discovery of knocking down of CXCR5 decreased the production of MMP-13 induced by CXCL13, further confirming the conclusion that CXCL13 contributes to the

progression and secretion of MMP-13, and demonstrated that CXCL13-CXCR5 axis is involved in the regulation of colon cancer cell migration and invasion.⁵⁵ Another way is that CXCL13-CXCR5 axis induces the activation of AKT in colon cancer cells. PI3K/AKT pathway plays a vital role in the development and progression of cancer.⁵⁷ The blockage of PI3K/AKT pathway suppresses the CXCL13-mediated growth, migration, and invasion of colon cancer cells, and further provides a new treatment strategy for colon cancer.⁵⁵

Overall, meaningful results were found in this study, our bioinformatics assessment proved that DEGs may play a significant role in the incidence, prognosis, growth, and development of CRC. Furthermore, the core genes and pathways might become potential biomarkers that could be used for the detection and targeting of CRC cells for therapy. However, there are still some limitations. To start with, due to the lack of experimental verification, some connections and mechanisms can't yet to be confirmed. Secondly, the specific mechanisms of how these ten hub genes influence the tumorigenesis and progression of CRC are not yet fully understood. Therefore, further investigations are required to figure out the function and the possible mechanism of these hub genes.

Conclusion

In conclusion, with the integrated bioinformatics analysis for gene expression profiles in colorectal cancer and colon adenoma, we found out ten hub genes (ADCY9, CCL19, CXCL12, CXCL13, GNG2, LPAR1, CCL5, PYY, P2RY14 and SST) and three core genes (CCL5, CCL19 and CXCL13) closely associated with the pathogenesis and prognosis of CRC. These hub genes may potentially become novel diagnostic and prognostic biomarkers for early identification of CA with cancerous potential, prognosis, targeted therapy and gene therapy. However, further in-depth study (in vivo and in vitro experiment) is necessary to clarify the biological function of these genes in colorectal cancer. Moreover, whether there are differences in the expression of these genes in different stages of CRC also needs to be further studied and discussed.

Declarations

Ethics approval and consent to participate:

Not applicable.

Consent for publication:

Not applicable.

Availability of data and materials:

The data that support the findings of this study are available upon request from the corresponding author.

Competing interests:

The authors have no conflict of interest to report.

Funding: The research has no project funding.

Authors' contributions and Acknowledgements:

The first author contribute to data collection, data analysis and manuscript writing. The corresponding author for project designing and manuscript proofreading.

Acknowledgements

Yuxuan Huang contribute to data collection, data analysis and manuscript writing. Ge CUI for project designing and manuscript proofreading.

Conflicts of interest

The authors declare no conflicts of interest.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

References

1. Global cancer statistics 2018. GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2020;70:313–3.
2. Akagi T, Inomata M. Essential advances in surgical and adjuvant therapies for colorectal cancer 2018–2019. *Annals of Gastroenterological Surgery.* 2020;4:39–46.
3. Babaei M, et al. Minimally invasive colorectal cancer surgery in Europe. *Med. (United States)* 95, (2016).
4. Babaei M, et al. Administration of adjuvant chemotherapy for stage II-III colon cancer patients: An European population-based study. *Int J Cancer.* 2018;142:1480–9.
5. Babaei M, et al. Neoadjuvant Therapy in Rectal Cancer Patients With Clinical Stage II to III Across European Countries: Variations and Outcomes. *Clin Colorectal Cancer.* 2018;17:e129–42.
6. Ganesh K, et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nature Reviews Gastroenterology Hepatology.* 2019;16:361–75.
7. McQuade RM, Stojanovska V, Bornstein JC, Nurgali K. Colorectal Cancer Chemotherapy: The Evolution of Treatment and New Approaches. *Curr. Med. Chem.* 24, (2017).
8. Siegel RL, et al. Colorectal cancer statistics, 2020. *CA. Cancer J. Clin.* 70, 145–164 (2020).
9. Hill MJ, Morson BC, Bussey HJ. R. AETIOLOGY OF ADENOMA-CARCINOMA SEQUENCE IN LARGE BOWEL. *Lancet.* 1978;311:245–7.

10. Click B, Pinsky PF, Hickey T, Doroudi M, Schoen RE. Association of colonoscopy adenoma findings with long-term colorectal cancer incidence. *JAMA - J Am Med Assoc.* 2018;319:2021–31.
11. Aubert J, Bar-Hen A, Daudin JJ, Robin S. Determination of the differentially expressed genes in microarray experiments using local FDR. *BMC Bioinformatics* 5, (2004).
12. Szklarczyk D, et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43:D447–52.
13. Cline MS, et al. Integration of biological networks and gene expression data using cytoscape. *Nat Protoc.* 2007;2:2366–82.
14. Cerami E, et al. The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012;2:401–4.
15. Long X, et al. Peptostreptococcus anaerobius promotes colorectal carcinogenesis and modulates tumour immunity. *Nat Microbiol.* 2019;4:2319–30.
16. West NR, Mccuaig S, Franchini F, Powrie F. Emerging cytokine networks in colorectal cancer. *Nat Rev Immunol.* 2015;15:615–29.
17. GOLD P, FREEDMAN SO, DEMONSTRATION OF TUMOR-SPECIFIC, ANTIGENS IN HUMAN COLONIC CARCINOMATA BY IMMUNOLOGICAL TOLERANCE AND ABSORPTION TECHNIQUES. *J Exp Med.* 1965;121:439–62.
18. Duffy MJ, et al. Tumor markers in colorectal cancer, gastric cancer and gastrointestinal stromal cancers: European group on tumor markers 2014 guidelines update. *Int J Cancer.* 2014;134:2513–22.
19. Locker GY, et al. ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J Clin Oncol.* 2006;24:5313–27.
20. Assenov Y, Ramírez F, Schelhorn SESE, Lengauer T, Albrecht M. Computing topological parameters of biological networks. *Bioinformatics.* 2008;24:282–4.
21. SUTHERLAND EW, RALL TW. Fractionation and characterization of a cyclic adenine ribonucleotide formed by tissue particles. *J Biol Chem.* 1958;232:1077–91.
22. Niles RM, Ludwig KW, Makarski JS. Differential growth inhibition in two human carcinoma cell lines by cyclic adenosine 5'-monophosphate analogs. *J Natl Cancer Inst.* 1979;63:909–11.
23. Ashman DF, Lipton R, Melicow MM, Price TD. Isolation of adenosine 3',5'-monophosphate and guanosine 3',5'-monophosphate from rat urine. *Biochem Biophys Res Commun.* 1963;11:330–4.
24. Tao L, Zhu Y, Gu Y, Zheng J, Yang J. ARHGAP25: A negative regulator of colorectal cancer (CRC) metastasis via the Wnt/ β -catenin pathway. *Eur. J. Pharmacol.* 858, (2019).
25. Kim DU, Kwak B, Kim SW. Phosphodiesterase 4B is an effective therapeutic target in colorectal cancer. *Biochem Biophys Res Commun.* 2019;508:825–31.
26. Hirsh L, et al. Phosphodiesterase inhibitors as anti-cancer drugs. *Biochemical Pharmacology* Vol. 68 pp. 981–8 (Biochem Pharmacol, 2004).

27. Yang B, et al. The role of interleukin 17 in tumour proliferation, angiogenesis, and metastasis. *Mediators of Inflammation* vol. 2014 (2014).
28. Liu J, et al. IL-17 is associated with poor prognosis and promotes angiogenesis via stimulating VEGF production of cancer cells in colorectal carcinoma. *Biochem Biophys Res Commun*. 2011;407:348–54.
29. Chin CC, et al. Interleukin-17 induces CC chemokine receptor 6 expression and cell migration in colorectal cancer cells. *J Cell Physiol*. 2015;230:1430–7.
30. Wang Q, Feng M, Yu T, Liu X, Zhang P. Intratumoral regulatory T cells are associated with suppression of colorectal carcinoma metastasis after resection through overcoming IL-17 producing T cells. *Cell Immunol*. 2014;287:100–5.
31. Takayama T, Miyanishi K, Hayashi T, Sato Y, Niitsu Y. Colorectal cancer: genetics of development and metastasis. *Journal of gastroenterology*. 2006;41:185–92.
32. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1990;61:759–67.
33. Schulz-Heddergott R, et al. Therapeutic Ablation of Gain-of-Function Mutant p53 in Colorectal Cancer Inhibits Stat3-Mediated Tumor Growth and Invasion. *Cancer Cell*. 2018;34:298–314.e7.
34. Lv D, Zhang Y, Kim HJ, Zhang L, Ma X. CCL5 as a potential immunotherapeutic target in triple-negative breast cancer. *Cellular Molecular Immunology*. 2013;10:303–10.
35. Aldinucci D, Colombatti A. The inflammatory chemokine CCL5 and cancer progression. *Mediators of Inflammation* vol. 2014 (2014).
36. Cambien B, et al. CCL5 neutralization restricts cancer growth and potentiates the targeting of PDGFR β in colorectal carcinoma. *PLoS One* 6, (2011).
37. Chang LY, et al. Tumor-derived chemokine CCL5 enhances TGF- β -mediated killing of CD8 + T cells in colon cancer by T-regulatory cells. *Cancer Res*. 2012;72:1092–102.
38. Maeda K, Das D, Nakata H, Mitsuya H. CCR5 inhibitors: Emergence, success, and challenges. *Expert Opinion on Emerging Drugs*. 2012;17:135–45.
39. Lipson EJ Re-orienting the immune system: Durable tumor regression and successful re-induction therapy using anti-PD1 antibodies. *Oncoimmunology* 2, (2013).
40. Aldinucci D, Borghese C, Casagrande N. The CCL5/CCR5 Axis in Cancer Progression. *Cancers (Basel)*. 2020;12:1–30.
41. Gallo M, De Luca A, Lamura L, Normanno N. Zoledronic acid blocks the interaction between mesenchymal stem cells and breast cancer cells: Implications for adjuvant therapy of breast cancer. *Ann Oncol*. 2012;23:597–604.
42. Dieu MC, et al. Selective recruitment of immature and mature dendritic cells by distinct chemokines expressed in different anatomic sites. *J Exp Med*. 1998;188:373–86.
43. Ngo VN, Tang HL, Cyster JG. Epstein-Barr virus-induced molecule 1 ligand chemokine is expressed by dendritic cells in lymphoid tissues and strongly attracts naive T cells and activated B cells. *J Exp Med*. 1998;188:181–91.

44. Hou T, et al. Atypical chemokine receptors predict lymph node metastasis and prognosis in patients with cervical squamous cell cancer. *Gynecol Oncol.* 2013;130:181–7.
45. Zhou R, Sun J, He C, Huang C, Yu H. CCL19 suppresses gastric cancer cell proliferation, migration, and invasion through the CCL19/CCR7/AIM2 pathway. *Hum Cell.* 2020;33:1120–32.
46. Lu J, et al. Antitumor Efficacy of CC Motif Chemokine Ligand 19 in Colorectal Cancer. *Dig Dis Sci.* 2014;59:2153–62.
47. Lu J, et al. CC motif chemokine ligand 19 suppressed colorectal cancer in vivo accompanied by an increase in IL-12 and IFN- γ . *Biomed Pharmacother.* 2015;69:374–9.
48. Sun JC, et al. Dendritic cells-mediated CTLs targeting hepatocellular carcinoma stem cells. *Cancer Biol Ther.* 2010;10:368–75.
49. Ilett EJ, Prestwich RJD, Melcher AA. The evolving role of dendritic cells in cancer therapy. *Expert Opinion on Biological Therapy.* 2010;10:369–79.
50. Gunn MD, et al. A B-cell-homing chemokine made in lymphoid follicles activates Burkitt's lymphoma receptor-1. *Nature.* 1998;391:799–803.
51. Hussain SK, et al. Serum levels of the chemokine CXCL13, genetic variation in CXCL13 and its receptor CXCR5, and HIV-associated non-Hodgkin B-cell lymphoma risk. *Cancer Epidemiol Biomarkers Prev.* 2013;22:295–307.
52. Panse J, et al. Chemokine CXCL13 is overexpressed in the tumour tissue and in the peripheral blood of breast cancer patients. *Br J Cancer.* 2008;99:930–8.
53. Razis E, et al. Improved outcome of high-risk early HER2 positive breast cancer with high CXCL13-CXCR5 messenger RNA expression. *Clin Breast Cancer.* 2012;12:183–93.
54. Qi XW, et al. Expression features of CXCR5 and its ligand, CXCL13 associated with poor prognosis of advanced colorectal cancer. *Eur Rev Med Pharmacol Sci.* 2014;18:1916–24.
55. Z Z, et al. CXCL13-CXCR5 axis promotes the growth and invasion of colon cancer cells via PI3K/AKT pathway. *Mol. Cell. Biochem.* 400, (2015).
56. Brown PD. Matrix metalloproteinase inhibitors in the treatment of cancer. *Med Oncol.* 1997;14:1–10.
57. Agarwal E, Brattain MG, Chowdhury S. Cell survival and metastasis regulation by Akt signaling in colorectal cancer. *Cell Signal.* 2013;25:1711–9.

Tables

Due to technical limitations, table 1-5 is only available as a download in the Supplemental Files section.

Figures

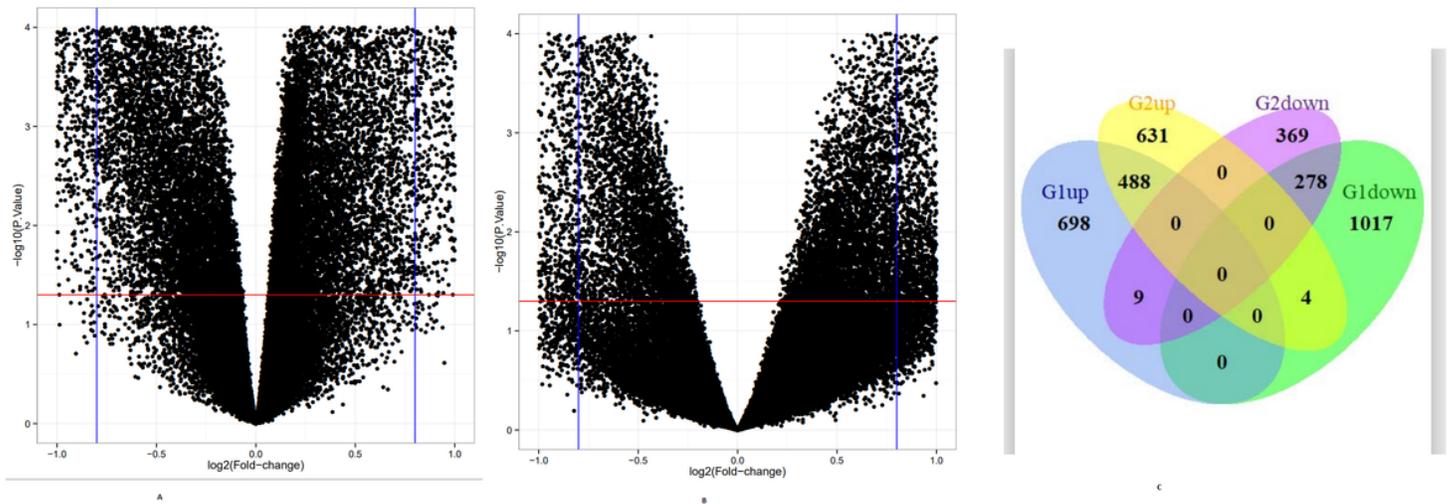


Figure 1

Volcano plot and Venn diagram of DEGs from dataset of colorectal tumor and normal cell lines. A. Volcano plot of DEGs between CRC tissues and normal cell lines from the GSE32323 dataset (defined as G1); B. Volcano plot of DEGs between CA tissues and normal cell lines from GSE4183 dataset (defined as G2). x-axis: \log_2FC , large-scale fold changes; y-axis: $-\log_{10}$ of the p-value showing the statistical significance. Each black point represents one gene. The black points above the red line and beside the blue line (left side and right side) represent $\log_2FC \geq 1$ and $p\text{-value} < 0.05$; all DEGs that were CRC-related are shown below the red line, which represents $\log_2FC < 1$ or $p\text{-value} > 0.05$. C. Venn diagram of up-regulated overlapping 488 DEGs and down-regulated overlapping 278 DEGs from GSE32323, GSE483 datasets.

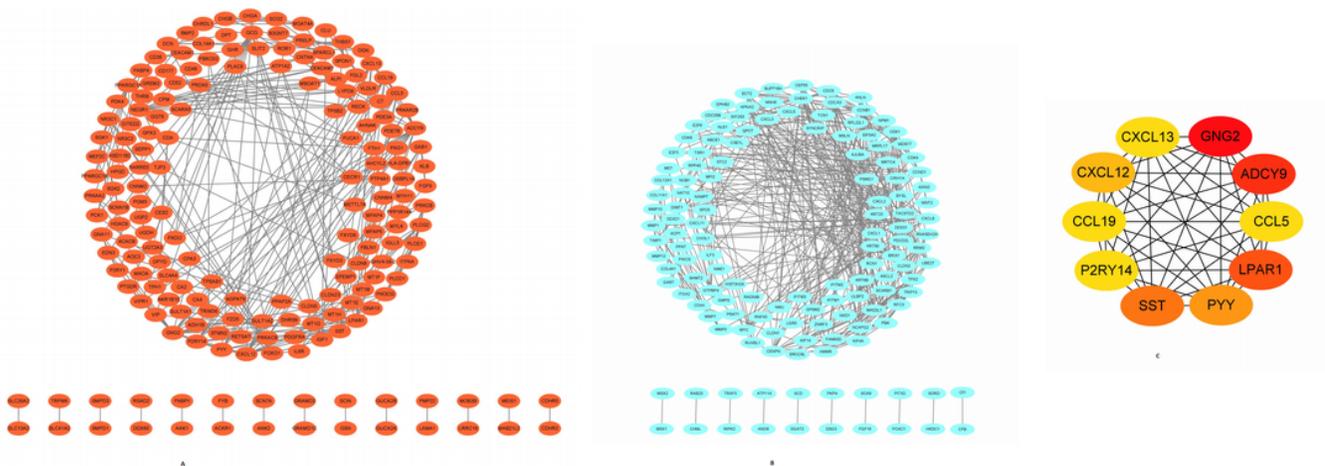


Figure 2

PPI networks of DEGs and hub genes. A. The protein networks of the upregulated DEGs visualized applying Cytoscape are shown; B. The protein networks of the downregulated DEGs visualized applying Cytoscape are shown. The representation is as follows: spheres represent the nodes, and the edges are

shown as lines; C. A total of 10 Hub genes were identified and selected from the PPI network. The connection degree of each molecule with others gradually decreases from the red nodes to yellow ones.

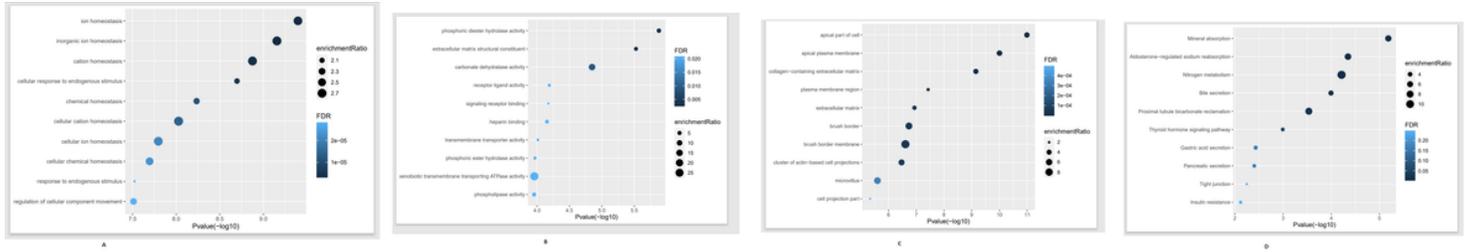


Figure 3

GO and KEGG pathway analysis of the upregulated DEGs in CA and CRC. A. Top 10 GO biological process (BP) analysis. B. Top 10 GO molecular function (MF) analysis. C. Top 10 GO cell component (CC) analysis. D. Top 10 KEGG pathways profiles. All of the enrichment pathways were produced utilizing the ggplot2 package in R language

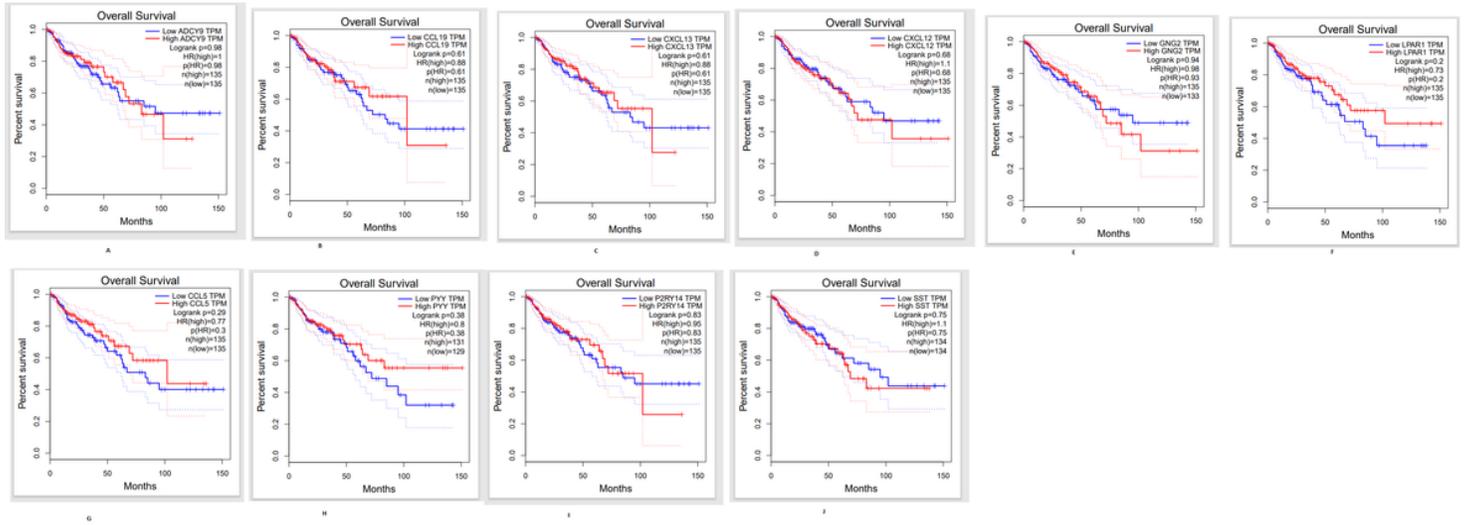


Figure 4

The survival curves were plotted applying the online tool Kaplan Meier plotter in GEPIA2 web server. The genes with high expression in the cohorts are shown in red, and the blue line demonstrates the low-expression cohort. The survival curves are represented as dotted lines, and the solid line represents the 95% confidence interval. HR stands for hazard ratio. The p-values were calculated using log-rank statistics.

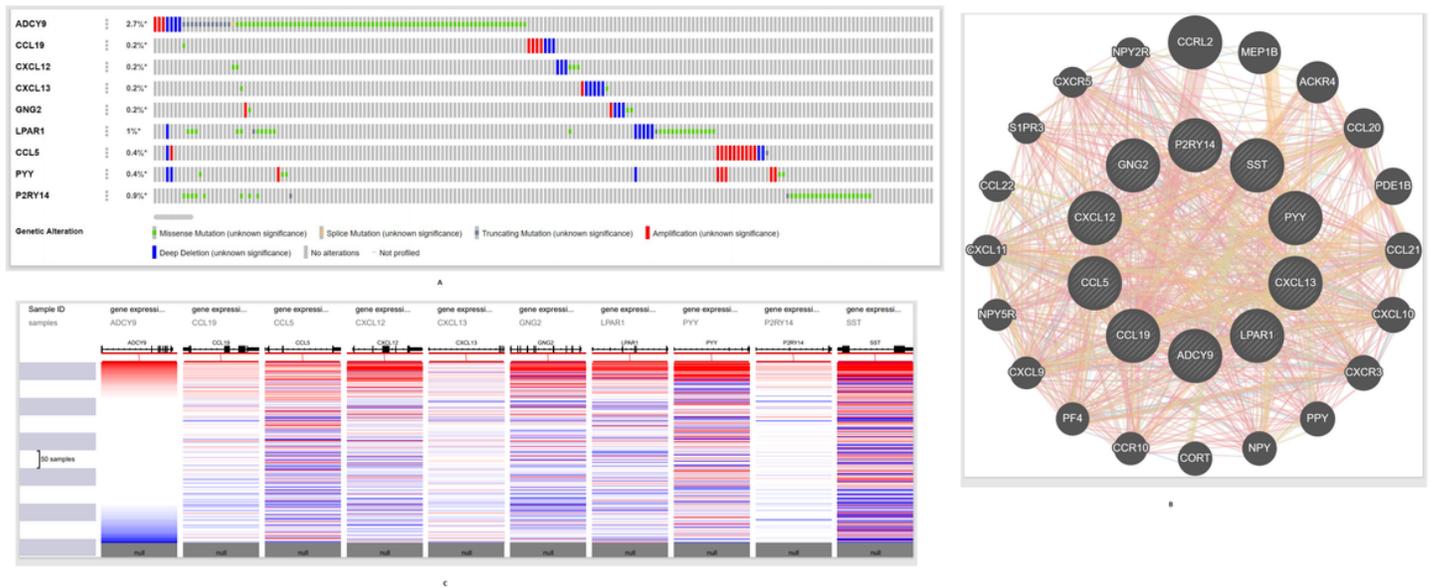


Figure 5

Genetic alterations associated with hub genes in TCGA-CRC. A visual summary across a query of ten hub genes showing genetic alteration of ten hub genes in TCGA-CRC patients. B. The network contains 30 genes (10 hub genes and 20 relative genes). Relationship between hub genes and their receptors is also illustrated. C. Hierarchical clustering of hub genes was constructed using UCSC. The samples under the pink bar are non-cancerous samples and the samples under the blue bar are CRC samples. Upregulation of genes is marked in red; downregulation of genes is marked in blue.

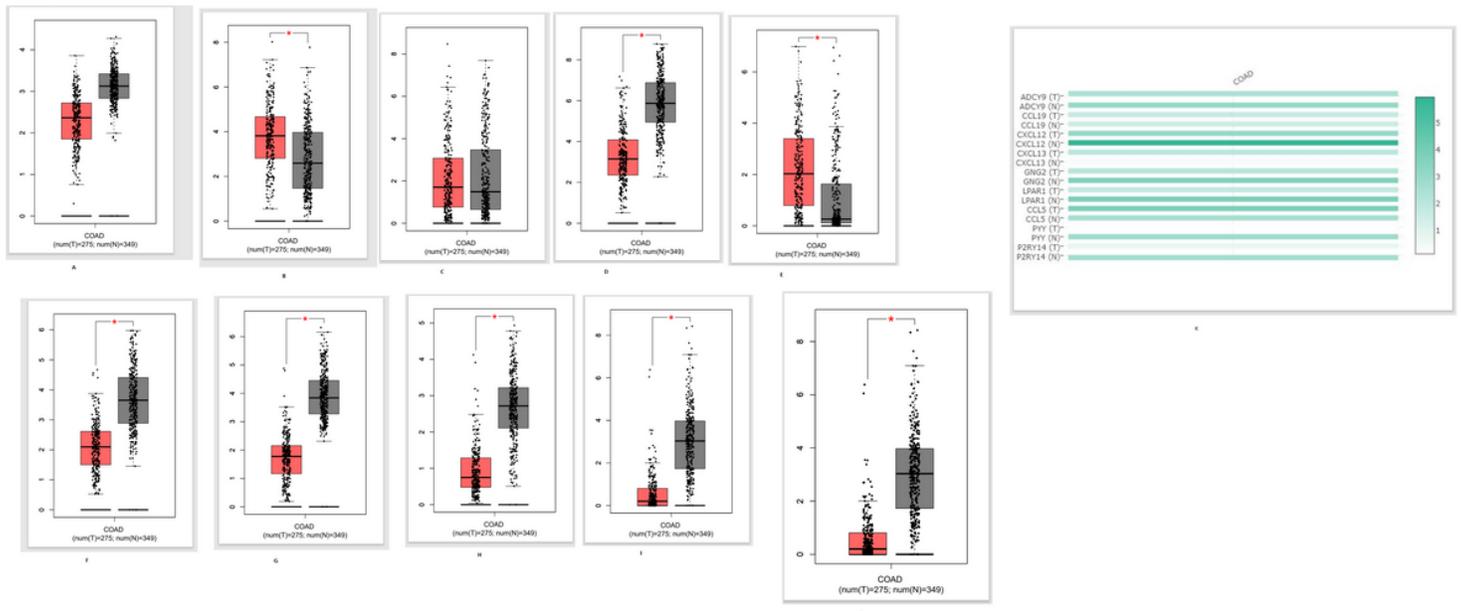


Figure 6

The expression level of hub genes between CRC and normal tissue according to GEPIA database. Three hub genes were significantly higher ($P < 0.01$) in CRC tissues than those in normal colon tissues (Figure

6B, 6C, 6E), which was in consistent with the findings in Oncomine (Figure 6K).

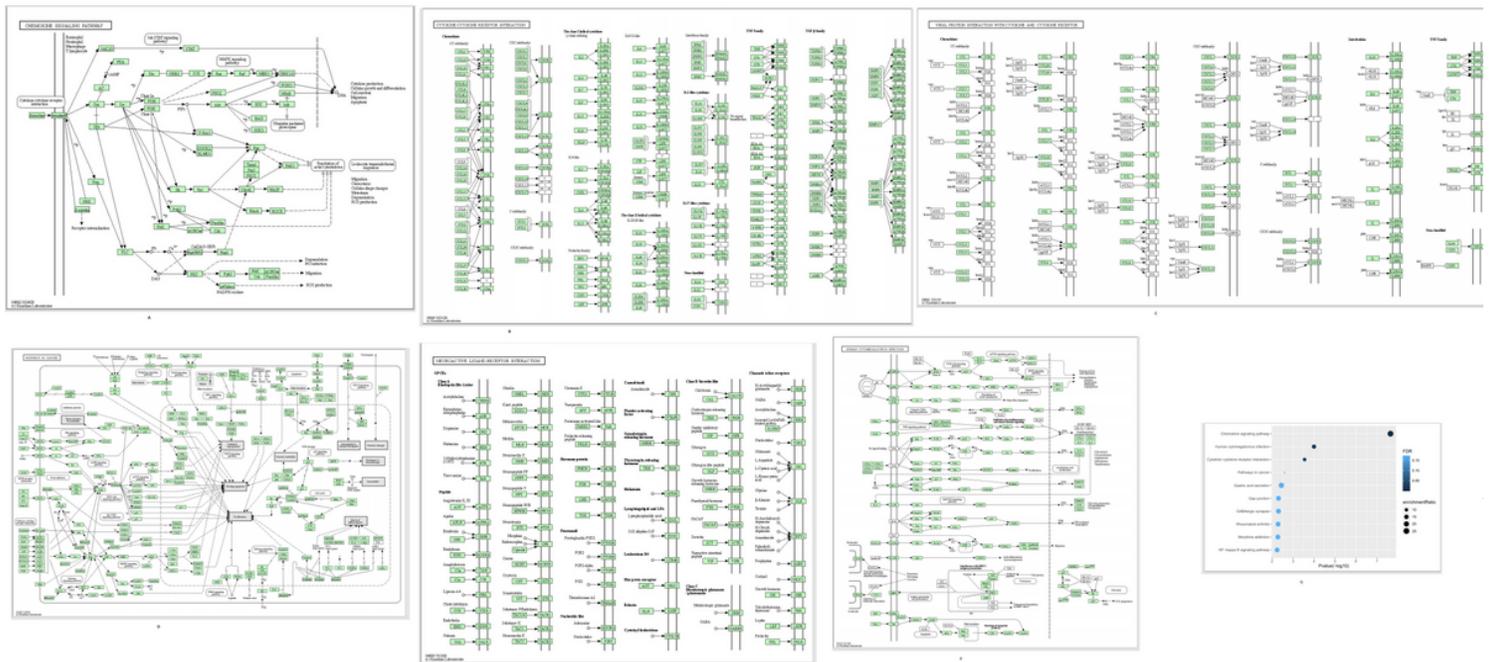


Figure 7

KEGG enrichment analysis of the hub genes. A. CXCL13, ADCY9, GNG2, CCL5, CCL19, and CXCL12 are enriched in the chemokine signaling pathway (Chemokine represents CXCL13, CXCL12, CCL5, and CCL19, AC represents ADCY9, and G β represents GNG2). B. CXCL13, CCL55, CCL19 and CXCL12 are enriched in the cytokine- cytokine receptor interaction. C. CXCL13, CCL5, CCL19, and CXCL12 are enriched in viral protein interaction with cytokine and cytokine receptor. D. ADCY9, LPAR1, GNG2, and CXCL12 are enriched in pathways in cancer (AC represents ADCY9, and β represents GNG2, and GPCR represents LPAR1, and SPF1 represents CXCL12). E. LPAR1, PYY, SST, and P2RY14 are enriched in the neuroactive ligand- receptor interaction (EDGL represents LPAR1, and NPY represents PYY). F. ADCY9, GNG2, CCL5, and CXCL12 are enriched in the human cytomegalovirus infection (AC represents ADCY9, and G β represents GNG2). G. The conclusion of KEGG enrichment analysis of the hub genes, which is graphed in the R Studio.

Supplementary Files

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

- [Table1.xlsx](#)
- [Table2.xlsx](#)
- [Table3.xlsx](#)
- [Table4.xlsx](#)

- [Table5.xlsx](#)