Identification of Gene-specific DNA Methylation Signature Between the Crohn's Disease and Colorectal Cancer

Daopo Lin

Wenzhou medical university

Haoqi Zhu

Wenzhou Medical University

Guolong Ma

Wenzhou medical university

Xiaoxiao Shao

Wenzhou medical university

Yi Jiang (✉ wzjiangyi@yeah.net)

Department of Gastroenterology, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou, Zhejiang Province, China.

Research

Keywords: differentially methylated genes, colorectal cancer , methylation, Crohn's disease

Posted Date: October 15th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-91620/v1>

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

[Read Full License](#)

Abstract

Background: Crohn's disease (CD) is a dynamic, chronic inflammatory condition that is closely related to an increased incidence of colorectal cancer (CRC). However, the underlying pathophysiology remains still unclear. The present study was planned to explore the gene-specific DNA methylation signature between the CD and CRC and, therefore, to discover potential biomarkers during the initiation and progression of CRC.

Methods: DNA methylation data of CD were extracted from Gene Expression Omnibus (GEO) database, while the microarray data of CRC were collected from The Cancer Genome Atlas (TCGA). The differentially methylated genes (DMGs) were screened between diseases group and normal controls. Using the coexisted DMGs in CD and CRCs, we further performed a series of functional enrichment analyses. Then correlation analysis for DNA methylation and mRNA expression was conducted. Next, we analyzed another two independent data sets from GEO to confirm the expression levels of the hub genes between CD and CRC.

Results: Totally 4301 DMGs were screened from CD samples compared with healthy control samples, while 2076 DMGs were discovered between CRC tissues and the control specimens. 180 DMGs coexisted in CD and CRC, including 72 comethylated genes. Hypermethylation genes were mostly involved in cAMP signaling pathway, chemokine signaling pathway and calcium signaling pathway. Hypomethylation genes were related to pathways as tryptophan metabolism. Correlation analysis for DNA methylation and mRNA expression data showed that 31 gene methylation status was negatively correlated with their mRNA expression. Moreover, two genes were found to associate with the initiation and progression of CD and CRC.

Conclusions: Our results suggested that ISX and SULF1 gene were discovered to be related to the occurrence and development of CD and CRC, and SULF1 may function as a signature for prognosis for CD-related CRC.

Background

As one form of inflammatory bowel disease (IBD), Crohn's disease (CD) is a chronic, incurable inflammatory disorder that is featured by relapsing and remitting inflammation of the entire gastrointestinal tract. Because of excessive and long-term inflammation of digestive tract, CD is closely associated with numerous malignant tumor of the digestive tract, especially CD-related colorectal cancer (CRC) [1, 2]. Although initially more extensively evaluated for ulcerative colitis (UC), a body of evidence indicates that CD carries a similar magnitude of risk for CRC over the past decade [3]. Compared with sporadic CRC, patients with onset of CD-associated CRC are younger and usually accompany by stenosis and fistulae, which enhance the difficulty of performing endoscopy [4]. What's worse, CD-related CRC is often diagnosed at more advanced stages and therefore has a poorer prognosis. Thus, it is an urgent task to provide a novel target for CD, which may also in favor of early diagnosis and treatment for CRC.

The pathogenesis of IBD-related CRC involves environmental and genetic factors, including epigenetic modifications, genetic mutations, intestinal microbiota and immune abnormalities [5]. Under the influence of environmental factors such as illness and diet, epigenetic changes have been confirmed in development, aging, and pathogenesis [6]. As a major conserved epigenetic mark, DNA methylation alters the structure of the DNA molecule by adding a methyl group to the 5' carbon of cytosine, which may result in possible modifications of gene expression patterns and thereby affect the development and progression of complex diseases, such as CD and CRC [7, 8].

Two recent studies concerning genome-wide DNA methylation evaluated changes in colonic mucosa in untreated pediatric IBD and childhood-onset CD, and found plenty of disease-related methylation alterations in many genetic loci, such as the HLA region and miR21 [9, 10]. Furthermore, the risk of IBD associated CRC is also related to DNA methylation. Compared with UC controls, cancer-related genes (RUNX3 and MINT1), and age-related genes such as MYOD were more highly methylated in UC-related carcinomas than in UC controls. Besides, there are distinctive methylation patterns in both neoplastic and non-neoplastic mucosa in patients with UC-related carcinomas, indicating that the methylation of cancer- and age-related genes may lead to these two disease processes to differing degrees [11]. Azuara et al. found methylation of SLIT2 and TMEFF2 was more frequently detected in the mucosa of UC and CD patients at high risk of dysplasia or cancer than patients at low risk, indicating that evaluation of the methylation status could provide biomarkers for the early detection of IBD-related CRC [7]. However, the explicit DNA methylation pattern in the course of occurrence and development of CD-related CRC, and the role of methylation in targeted treatment remains largely unknown.

In this bioinformatics analysis, DNA methylation data of CD were downloaded from Gene Expression Omnibus (GEO) database, while the microarray data of CRC were collected from The Cancer Genome Atlas (TCGA). We tried to determine the differentially methylated genes (DMGs) in CD and CRC. Firstly, intersections between CD DMGs and CRC DMGs were identified. Then functional enrichment analysis, construction of protein-protein interaction (PPI) network, DNA methylation-mRNA regulatory net analysis and validation of the hub genes between CD and CRC were performed. Finally, gene-specific DNA methylation signatures for diagnosis and prognosis during the progression of CD-related CRC were determined.

Materials And Methods

Data source and pre-processing

The DNA methylation data set (GSE105798) contained 8 CD samples and 3 healthy control samples, which was based on the GPL13534 (Illumina HumanMethylation450 BeadChip). Publicly available methylation data and the corresponding clinical information about colon adenocarcinoma cases (COAD) were obtained from TCGA (<http://cancergenome.nih.gov/>) based on Human Methylation 450 k Illumina Infinium methylation arrays, which included 474 CRC samples and 74 normal tissue samples.

Identification of differentially methylated gene

We used 'Champ' package in R software to analyze GSE105798, in an effort to identify differentially methylated probes (DMPs). As for the DMPs, we set false discovery rate (FDR) < 0.05 and cut-off $\beta > 0.3$. DMPs located in the gene region were assigned to the corresponding genes, which were defined as DMGs. The 'Limma' package was used to analyze the data collected from TCGA database. The differences for gene methylation were characterized with $|\log FC \text{ (fold change)}| > 0.5$, FDR < 0.05. Then an online Venn diagram tool (<http://bioinfoggp.cnb.csic.es/tools/venny/>) was employed to recognize overlapping DMGs between CD and CRC.

Functional enrichment analysis

The overlapping DMGs between CD and CRC were distributed to four quadrants of rectangular coordinate system on the basis of methylation patterns. 'ClusterProfiler' package in R software was used to perform Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis for the overlapping DMGs between CD and CRC. The threshold *P* value was set as 0.05, and the results were exhibited as bubble charts.

Protein-Protein interaction (PPI) network construction

The PPI networks of overlapping DMGs were constructed based on the STRING database [12] (<https://string-db.org/cgi/input.pl>) and visualized by The Cytoscape 3.3 [13] (www.cytoscape.org/). Degrees served as a valuable way to determine the role of protein nodes in the network.

Validation of the DNA methylation-mRNA regulatory net

Adding methyl groups at cytosine-guanine dinucleotides (CpGs) in regulatory/promoter regions in DNA, called DNA methylation, typically results in transcriptional repression and decreased expression of the gene. We performed a correlation analysis to detect whether DNA methylation causes CRC by negatively regulating gene. We downloaded the COAD expression data from TCGA database. The 'Limma' package was used to identify genes that were differentially expressed between CRC tumor samples and normal tissues. In addition, a variety of highly correlated differentially expressed genes and DNA methylated locis were identified between tumor samples and normal samples using Pearson correlation analysis. The value of $P < 0.05$ was regarded as the threshold value.

Validation of the hub genes between CD and CRC

To confirm the expression levels of the hub genes, we analyzed another two independent data sets from GEO (GSE52746 and GSE75214). GSE52746 was retrieved from GPL17996 (Affymetrix Human Genome U133 Plus 2.0 Array), involving 15 active CD samples and 17 healthy controls. Similarly, GSE75214 data set was based on GPL6244 (Affymetrix Human Gene 1.0 ST Array), which contained 59 CD samples and 22 healthy control samples. The expression levels of hub genes between CD and control groups were compared using Student t-test, with a significance threshold of $P < 0.05$. Besides, so as to confirm our results in CRC, the hub genes were validated by TCGA expression data. The expressions of the hub genes in CRC specimens were compared with normal tissues. Next, the expression levels of hub genes in CRC

tissues were compared with matched normal tissues using publicly available gene expression data in the Oncomine Cancer Microarray database (<http://www.oncomine.org>). Moreover, the Human Protein Atlas database [14] (<https://www.proteinatlas.org/>) was applied to validate the translational levels of the hub genes. Finally, GEPIA (Gene Expression Profiling Interactive Analysis) [15], an online server derived from TCGA, was used to assess the association between the hub genes and the prognosis.

Results

DMGs of CD samples and CRC specimens

In total, 11065 DMPs were screened out by comparing the 8 CD tissues with the 3 healthy control samples. The heat map of the DMPs is presented in Fig. 1a. After the DMPs were assigned to the corresponding genes, 4301 DMGs were obtained, of which 937 were downregulated, and 3364 upregulated. Besides, totally 2076 DMGs were discovered in CRC tissues after being compared with its control specimens, and were visualized in Volcano Plot (Fig. 1b). By comparing CD DMGs with CRC DMGs, a total of 180 DMGs were found to coexist in both groups (Fig. 2). It is noteworthy that 72 genes were detected to have the same methylation direction, of which 54 genes were high-methylated (DMG1). In addition, there were 108 genes that had opposite methylation pattern. There were 100 genes that were high-methylated in CD but were low-methylated in CRC (DMG4). Besides, only 8 genes that had a high methylation level in CRC while a low-methylation level in CD (DMG2). The fold distribution of the four lists DMGs is presented in Fig. 2b.

KEGG enrichment analysis for the four lists DMGs

The role of 180 cointeraction genes was further explored through the KEGG enrichment analysis. DMG1 were mostly enriched in cAMP signaling pathway, chemokine signaling pathway and calcium signaling pathway, and so on (Fig. 3a). There were 18 high-methylated genes coexisting in both groups (DMGs 3), which were involved in metabolic pathways as tryptophan metabolism (Fig. 3a). Additionally, the KEGG enrichment analysis was also carried out on the 108 genes with opposite directions of methylation in the two groups (Fig. 3b). DMGs2 were mostly associated with folate biosynthesis, galactose metabolism, fructose and mannose metabolism, and so on. And DMGs4 were mostly related to immunological pathways as cytokine-cytokine receptor interaction, viral protein interaction with cytokine and chemokine signaling pathway, and so on.

PPI network analysis on the cointeraction DMGs

Next, the software Cytoscape 3.3 was used to construct PPI network on the 180 intersection genes (Fig. 4). In the coexpression network, a total of 179 nodes and 128 edges of coexpression interactions were identified. According to degree, the top 10 hub genes in the networks were screened out, including IKZF1, CD69, CD53, MMP1, CCL20, CXCL1, SMAD2, DCN, ADAMTS5 and PI3. Notably, CD53 was the hypomethylated gene in both CD and CRC, which deserves further investigation.

Validation of the DNA methylation–mRNA regulatory net

We carried out correlation analysis to detect whether DNA methylation caused CRC by regulating gene expression. We selected all the methylation-mRNA pairs to evaluate the relationship between DNA methylation and gene expression in CRC sample. After a comprehensive analysis, we found that the DNA methylation levels of 31 genes were significantly negatively correlated with their expressions. Fig. 5 showed the top 9 most relevant genes.

Validation of the hub genes between CD and CRC

To verify these 31 genes, the expression levels of these target genes were further tested in two independent GEO data set for CD and TCGA expression data set concerning CRC. In GSE52746 and GSE75214 data set for CD, the expression level of Intestine Specific Homeobox (ISX) was significantly lower in CD than in control groups ($\log FC = -1.19$, $P < 0.0001$; $\log FC = -0.73$, $P < 0.001$; Fig. 6). When compared with healthy controls, patients with CD showed an increased expression level of Sulfatase1 (SULF1) ($\log FC = 0.41$, $P < 0.01$; $\log FC = -0.81$, $P < 0.001$; Fig. 6). With regard to ISX and SULF1 gene expression levels in TCGA, the trend was consistent with that in CD. Using the Oncomine microarray database, the expression levels of ISX mRNA in CRC tissues were decreased compared with normal tissues from five independent studies [16-20], while the expression levels of SULF1 mRNA in CRC tissues were increased [16, 18-21] (Figs. 7-8). What's more, the immunohistochemistry staining acquired from the Human Protein Atlas database showed the increased expression of SULF1 genes in CRC (Fig. 9a). Finally, we found that the up-regulation of the expression of oncogene SULF1 was closely associated with overall survival (OS) (Fig. 9b).

Discussion

In this study, DMGs between CD group and the control group, as well as those between the CRC tissues and its corresponding control samples were identified. The potential pathological mechanisms of IBD-related CRC were discovered through exploring the differentially methylated pattern. More specifically, there were 54 high-methylated genes and 18 low-methylated genes that were coexisting in both CD DMGs and CRC DMGs, respectively. Overlapping genes possessing the same methylation direction between CD and CRC were mostly enriched in pathways as cAMP signaling pathway, chemokine signaling pathway, calcium signaling pathway and tryptophan metabolism. We assumed that CD may influence the initiation and progression of CRC by affecting the immunity and metabolism.

In view of the micro-environment within which colitis related dysplasia and CRC are arising, it is not surprising that host immune and inflammatory responses exert an enormous function on the pathogenesis. It is believed that the mechanisms of chronic inflammation contributing to CRC are by the induction of cytokines and chemokines, followed by ensuing alterations in epithelial cell proliferation, survival, and migration [22]. However, the exact pathogenesis of colitis-related CRC remains unclear. Extension of additional evaluations such as DNA methylation has tremendous potential to broaden our knowledge of colitis related CRC pathogenesis and chances for monitoring and prevention. In our study,

the KEGG pathway enrichment analysis discovered that genes with a high methylation level in both CD and CRC were mainly involved in cAMP signaling pathway, chemokine signaling pathway and calcium signaling pathway while genes with low-methylation level were mainly related to tryptophan metabolism. Chemokines and their interaction with chemokine receptors constitute a highly complex communication system, which allows targeted function of circulating immune cells [22, 23]. Chemokines have emerged as the most significant regulators of leukocyte trafficking during infection or inflammation and, therefore, have been involved in inflammatory changes and progression to CRC in IBD patients. A number of chemokines are shown to be influenced by inflammatory mediators including NF- κ B, which has a great effect on the development of IBD-related CRC [24]. Besides, chemokines can recruit immune cells and vascular cells into the tissues with intestinal inflammation. For example, CXCL8 is up-regulated in colitis and CRC by signaling through CXCR1, and has been associated with epithelial transition in colonic carcinoma [25]. Further researches are needed to detect how chemokines interact with immune cells, cellular receptors, and inflammatory mediators to result in neoplastic changes in IBD patients. Susanna and colleagues found that serum levels of tryptophan in patients with IBD were significantly lower than control, and the serum levels of tryptophan were negatively correlated with disease activity or level of c-reactive protein [26]. Animal studies have discovered that the administration of tryptophan or tryptophan metabolites might be a therapeutic method to IBD [27], suggesting that tryptophan and its metabolites might not only function as useful biomarkers, but also be promising therapeutic candidates in IBD. Several studies have shown that decreased tryptophan levels and increased kynurenine pathway metabolites in CRC patients [28]. We supposed that CD may affect the initiation and progression of CRC by affecting the immunity and metabolism.

What's more, we found that SULF1 and ISX may be closely related to the occurrence and development of CD and CRC. ISX, expressed in intestinal epithelial cells, plays a vital role in maintaining immunity and tolerance in response to diet-derived signals in the intestinal barrier, and is an integral regulator of vitamin A production [29]. A genome-wide association study by Irina et al. found strong correlation between ISX gene polymorphisms and CD [30]. As a proto-oncogene, ISX is also found to be an important regulator of hepatocellular carcinoma (HCC) progression and gastric carcinogenesis with significant potential as a prognostic and therapeutic target. Although the relationship between CRC and ISX is still unclear, multiple studies have demonstrated that decreased levels of vitamin A are associated with an increased risk for CRC [31, 32], suggested that ISX may also have an important role in CRC. SULF1 has recently been identified and shown to desulfurize cellular heparan sulfate proteoglycans (HSPGs). Since sulfated HSPGs serve as co-receptors for many cytokines and growth factors, it is expected that SULF1 can modulate growth factor and cytokine signaling. Studies concerning mouse model of colitis have clearly shown that HSPGs has important roles in regulation of chronic inflammation and gut barrier function [33]. Additionally, SULF1 gene was determined to serve as a potential biological marker to predict prognosis and a therapeutic target in CRC. Consistent with our research, Wei et al. discovered that SULF1 gene is closely related to the survival of patients with CRC [34]. Thus, the novel biomarker of ISX and SULF1 may contribute to understanding the molecular mechanism of the relationship between CD and CRC, the underlying mechanism needs further investigation.

Conclusion

To sum up, bioinformatics analysis in our study discovered that CD may affect the initiation and progression of CRC by affecting the immunity and metabolism. More importantly, two genes involved in the development of CRC were found, of which SULF1 gene was considered as a potential prognostic marker for CRC. However, the exact underlying pathophysiology between CD and CRC needed much further study.

Declarations

Acknowledgments

None

Authors' contributions

Prof Yi Jiang conceived and designed the experiments; Dr Daopo Lin and Dr Haoqi Zhu wrote the paper. Dr Guolong Ma and Dr Xiaoxiao Shao analyzed the data; All co-authors have carefully examined and verified the manuscript and approved to submit it to your journal.

Funding

This work was supported by grants of Natural Science Foundation of Zhejiang Province (Grant number: LY16H160055, LY17H030011, LY18H030009), and the Wenzhou Municipal Science and Technology Bureau (Grant number: Y20160102, Y20170062, Y20170637).

Availability of data and materials

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

Ethics approval and consent to participate

Ethics approval not applicable. The data do not compromise anonymity or confidentiality or breach local data protection law.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflict of interest

References

1. Larsen M, Mose H, Gislum M, et al. Survival after colorectal cancer in patients with Crohn's disease: A nationwide population-based Danish follow-up study. *Am J Gastroenterol*. 2007;102:163-7.
2. Munkholm P. Review article: the incidence and prevalence of colorectal cancer in inflammatory bowel disease. *Aliment Pharmacol Ther*. 2003;18:1-5.
3. Keller DS, Windsor A, Cohen R, Chand M. Colorectal cancer in inflammatory bowel disease: review of the evidence. *Tech Coloproctol*. 2019;23:3-13.
4. Averboukh F, Ziv Y, Kariv Y, et al. Colorectal carcinoma in inflammatory bowel disease: a comparison between Crohn's and ulcerative colitis. *Colorectal Dis*. 2011;13:1230-5.
5. Dong J, Liang W, Wang T, et al. Saponins regulate intestinal inflammation in colon cancer and IBD. *Pharmacol Res*. 2019;144:66-72.
6. Kellermayer R. Challenges for epigenetic research in inflammatory bowel diseases. *Epigenomics*. 2017;9:527-38.
7. Azuara D, Rodriguez-Moranta F, de Oca J, et al. Novel methylation panel for the early detection of neoplasia in high-risk ulcerative colitis and Crohn's colitis patients. *Inflamm Bowel Dis*. 2013;19:165-73.
8. Weisenberger DJ, Liang G, Lenz HJ. DNA methylation aberrancies delineate clinically distinct subsets of colorectal cancer and provide novel targets for epigenetic therapies. *Oncogene*. 2018;37:566-77.
9. Adams AT, Kennedy NA, Hansen R, et al. Two-stage genome-wide methylation profiling in childhood-onset Crohn's Disease implicates epigenetic alterations at the VMP1/MIR21 and HLA loci. *Inflamm Bowel Dis*. 2014;20:1784-93.
10. Harris RA, Nagy-Szakal D, Mir SA, et al. DNA methylation-associated colonic mucosal immune and defense responses in treatment-naïve pediatric ulcerative colitis. *Epigenetics*. 2014;9:1131-7.
11. Garrity-Park MM, Loftus EV Jr, Sandborn WJ, Bryant SC, Smyrk TC. Methylation status of genes in non-neoplastic mucosa from patients with ulcerative colitis-associated colorectal cancer. *Am J Gastroenterol*. 2010;105:1610-9.
12. Stelzl U, Worm U, Lalowski M, et al. A human protein-protein interaction network: a resource for annotating the proteome. *Cell*. 2005;122:957-68.
13. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498-504.
14. Pontén F, Jirström K, Uhlen M. The Human Protein Atlas—a tool for pathology. *J Pathol*. 2008;216:387-93.
15. Tang Z, Li C, Kang B, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45:W98-W102.
16. Gaedcke J, Grade M, Jung K, et al. Mutated KRAS results in overexpression of DUSP4, a MAP-kinase phosphatase, and SMYD3, a histone methyltransferase, in rectal carcinomas. *Genes Chromosomes Cancer*. 2010;49:1024-34.

17. Hong Y, Downey T, Eu KW, Koh PK, Cheah PY. A 'metastasis-prone' signature for early-stage mismatch-repair proficient sporadic colorectal cancer patients and its implications for possible therapeutics. *Clin Exp Metastasis*. 2010;27:83-90.
18. Kaiser S, Park YK, Franklin JL, et al. Transcriptional recapitulation and subversion of embryonic colon development by mouse colon tumor models and human colon cancer. *Genome Biol*. 2007;8:R131.
19. Ki DH, Jeung HC, Park CH, et al. Whole genome analysis for liver metastasis gene signatures in colorectal cancer. *Int J Cancer*. 2007;121:2005-12.
20. Skrzypczak M, Goryca K, Rubel T, et al. Modeling oncogenic signaling in colon tumors by multidirectional analyses of microarray data directed for maximization of analytical reliability. *PLoS One*. 2010;5:e13091.
21. Sabates-Bellver J, Van der Flier LG, de Palo M, et al. Transcriptome profile of human colorectal adenomas. *Mol Cancer Res*. 2007;5:1263-75.
22. Dulai PS, Sandborn WJ, Gupta S. Colorectal Cancer and Dysplasia in Inflammatory Bowel Disease: A Review of Disease Epidemiology, Pathophysiology, and Management. *Cancer Prev Res (Phila)*. 2016;9:887-94.
23. Cardona AE, Sasse ME, Liu L, et al. Scavenging roles of chemokine receptors: chemokine receptor deficiency is associated with increased levels of ligand in circulation and tissues. *Blood*. 2008;112:256-63.
24. Azer SA. Overview of molecular pathways in inflammatory bowel disease associated with colorectal cancer development. *Eur J Gastroenterol Hepatol*. 2013;25:271-81.
25. Ottaiano A, Franco R, Aiello Talamanca A, et al. Overexpression of both CXC chemokine receptor 4 and vascular endothelial growth factor proteins predicts early distant relapse in stage II-III colorectal cancer patients. *Clin Cancer Res*. 2006;12:2795-803.
26. Nikolaus S, Schulte B, Al-Massad N, et al. Increased Tryptophan Metabolism Is Associated With Activity of Inflammatory Bowel Diseases. *Gastroenterology*. 2017;153:1504-1516.e2.
27. Bettenworth D, Nowacki TM, Ross M, et al. Nicotinamide treatment ameliorates the course of experimental colitis mediated by enhanced neutrophil-specific antibacterial clearance. *Mol Nutr Food Res*. 2014;58:1474-90.
28. Santhanam S, Alvarado DM, Ciorba MA. Therapeutic targeting of inflammation and tryptophan metabolism in colon and gastrointestinal cancer. *Transl Res*. 2016;167:67-79.
29. Widjaja-Adhi MAK, Palczewski G, Dale K, et al. Transcription factor ISX mediates the cross talk between diet and immunity. *Proc Natl Acad Sci U S A*. 2017;114:11530-5.
30. Dinu I, Mahasirimongkol S, Liu Q, et al. SNP-SNP interactions discovered by logic regression explain Crohn's disease genetics. *PLoS One*. 2012;7:e43035.
31. Leenders M, Leufkens AM, Siersema PD, et al. Plasma and dietary carotenoids and vitamins A, C and E and risk of colon and rectal cancer in the European Prospective Investigation into Cancer and Nutrition. *Int J Cancer*. 2014;135:2930-9.

32. Luo H, Fang YJ, Lu MS, et al. Dietary and serum vitamins A and E and colorectal cancer risk in Chinese population: a case-control study. *Eur J Cancer Prev.* 2019;28:268-77.

33. Patterson AM, Delday MI, van Kuppevelt TH, et al. Expression of heparan sulfate proteoglycans in murine models of experimental colitis. *Inflamm Bowel Dis.* 2012;18:1112-26.

34. Wei S, Chen J, Huang Y, et al. Identification of hub genes and construction of transcriptional regulatory network for the progression of colon adenocarcinoma hub genes and TF regulatory network of colon adenocarcinoma. *J Cell Physiol.* 2020;235:2037-48.

Figures

Figure 1

Results of differentially methylated probes (DMPs) in GSE105798 and differentially methylated genes (DMGs) in TCGA. a Heat map of the DMPs. Hierarchical clustering showed separate groupings of DMPs for CD tissue and normal tissue. Green and blue colors indicate higher expression and lower expression, respectively. b Volcano Plot for differentially methylated genes for CRC samples and healthy control samples. Red and Blue colors indicate higher expression and lower expression, respectively. DMPs: differentially methylated probes; DMGs: differentially methylated genes; TCGA: The Cancer Genome Atlas; CD: Crohn’s disease; CRC: colorectal cancer

Figure 2

Differentially methylated genes between the CD DMGs and CRC DMGs. a Intersection between the CD DMGs and CRC DMGs. DMGs1 represents the genes with high-methylation levels in both CD and CRC. DMGs2 is on behalf of the genes that are hypermethylated in CRC but hypomethylated in CD. DMGs3 represents the genes with low-methylation levels in both CD and CRC. DMGs4 is on behalf of the genes that hypermethylated in CD but hypomethylated in CRC. b Distribution of the four lists DMGs. DMGs: differentially methylated genes; CD: Crohn's disease; CRC: colorectal cancer

Figure 3

Results of the functional enrichment analysis for the four list genes. a Bubble chart of the functional enrichment analysis for DMGs1 and DMG3. b Bubble chart of KEGG results of DMGs2 and DMG4. DMGs: differentially methylated genes; KEGG: Kyoto Encyclopedia of Genes and Genomes

Figure 5

Correlation between mRNA and DNA methylation in the TCGA COAD study. TCGA: The Cancer Genome Atlas; COAD: colon adenocarcinoma cases

Figure 6

Validation of the hub genes in CD and CRC. Analysis of SULF1 and ISX gene expression in CD vs normal tissues across each independent GEO data set, and the DEGs in unpaired CRC (N=462) vs normal tissues (N=41) in TCGA CD: Crohn's disease; CRC: colorectal cancer; GEO: Gene Expression Omnibus; DEGs: differentially expressed genes; TCGA: The Cancer Genome Atlas

Figure 7

SULF1 mRNA is overexpressed in colorectal cancer. Data sets in a single panel were from the same study. a Kaiser et al (Kaiser et al., 2007): 1, 5 colon samples; 2, 41 colon adenocarcinoma samples. b Skrzypczak et al (Skrzypczak et al., 2010): 1, 10 colon samples; 2, 5 colon carcinoma samples. c Skrzypczak et al (Skrzypczak et al., 2010): 1, 24 colorectal samples; 2, 36 colorectal carcinoma samples. d Hong et al (Hong, Downey, Eu, Koh, & Cheah, 2010): 1, 12 colon samples; 2, 70 colorectal carcinoma samples. e Ki et al (Ki et al., 2007): 1, 28 colon samples; 2, 13 liver samples; 3 50 colon adenocarcinoma samples. f Gaedcke et al (Gaedcke et al., 2010): 1, 65 rectum tissues samples; 2, 65 rectal adenocarcinoma samples. *P<0.05.

Figure 8

The mRNA expression level of ISX is decreased in colorectal cancer. Data sets in a single panel were from the same study. (a) Kaiser et al (Kaiser et al., 2007): 1, 5 colon samples; 2, 41 colon adenocarcinoma samples. (b) Skrzypczak et al (Skrzypczak et al., 2010): 1, 24 colorectal samples; 2, 36 colorectal carcinoma samples. (c) Skrzypczak et al (Skrzypczak et al., 2010): 1, 10 colon samples; 2, 5 colon adenoma samples. (d) Hong et al (Hong, Downey, Eu, Koh, & Cheah, 2010): 1, 12 colon samples; 2, 70 colorectal carcinoma samples. (e) Sabates-Bellver et al (Sabates-Bellver et al., 2007): 1, 4 ascending colon samples; 2, 5 descending colon samples; 3, 7 rectum colon samples; 4, 15 sigmoid colon samples; 5, 1 transverse colon sample; 6, 25 colon adenoma samples. (f) Gaedcke et al (Gaedcke et al., 2010): 1, 65 rectum samples; 2, 65 rectal adenocarcinoma samples. * $P < 0.05$.

Figure 9

Validate the translational levels of the hub genes and assess the association between the hub genes and the prognosis in CRC. (a) Validation of SULF1 gene on a translational level through the Human Protein Atlas database. (b) SULF1 was significantly associated with overall survival in CRC patients, using a Kaplan-Meier curve and a log-rank test. The patients were stratified into a high-level group and a low-level group according to median.