Comprehensive analysis of DNA methylation and gene expression profiles in rectal cancer

Cheng Zhang

The First Hospital of China Medical University <https://orcid.org/0000-0002-8121-2405>

Di Meng

China Medical University

Songjie Chao

China Medical University

Chunlin Ge (✉ 13998303666@139.com)

Department of Pancreatic and Biliary Surgery, The First Hospital of China Medical University

<https://orcid.org/0000-0001-8772-5554>

Primary research

Keywords: rectal cancer, methylation, bioinformatics

Posted Date: May 20th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-29003/v1>

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

[Read Full License](#)

Abstract

Background

Abnormal hypomethylation of oncogenes and hypermethylation of tumor suppressor genes play important roles in human tumorigenesis and cancer progression, including those of rectal cancer (RC). However, conjoint analysis of RC involving both gene expression and methylation profiling datasets remains rare. This study aimed to identify methylation-regulated differentially expressed genes (MeDEGs) and to evaluate their prognostic value in RC through bioinformatics analysis.

Methods

Gene expression (GSE20842 and GSE68204) and gene methylation (GSE75546) profiling datasets were obtained from the Gene Expression Omnibus database. GEO2R was adopted to identify differentially expressed genes (DEGs) and differentially methylated genes (DMGs). MeDEGs were obtained by overlapping the DEGs and DMGs and then subjected to protein–protein interaction (PPI) network analysis using STRING. Modules and hub genes within the network were identified using MCODE and CytoHubba, respectively. Prognostic MeDEGs were selected by univariate Cox regression. Finally, our findings were validated based on The Cancer Genome Atlas (TCGA) database.

Results

In total, 243 upregulated-hypomethylated and 51 downregulated-hypermethylated genes were identified as MeDEGs. A PPI network of MeDEGs was constructed with 290 nodes and 578 edges. Three modules and three hub genes—COL3A1, FPR1, and PLK1—within the network were identified. Three MeDEGs—NFE2, COMP, and LAMA1—were found to be survival-related. Furthermore, the expression and methylation status of two hub genes (excluding FPR1) and the three prognostic MeDEGs were also significantly altered in TCGA and were consistent with our findings.

Conclusions

We identified novel MeDEGs and explored their relationship with survival in RC. Our methodology may provide an effective bioinformatics basis for further understanding of the methylation-mediated regulatory mechanisms in RC.

Background

Colorectal cancer (CRC), which encompasses both colon cancer and rectal cancer (RC), is the third most common malignancy worldwide and the second most common cause of cancer-related death, only exceeded by lung cancer^[1, 2]. Its incidence and mortality have continually increased in recent years^[3]. The 5-year survival rate of CRC is still dismal (only 8%–9%) once metastasis has occurred^[3]. RC accounts for approximately 30% of all CRC cases and is associated with worse clinical outcomes^[1, 4]. Although it is frequently grouped with colon cancer, RC is thought to be a biologically separate disease. Crucially, the

rectum is enclosed by the pelvis, which elevates the likelihood of the local invasion of nearby tissue compared with colon cancer and increases the chance of local recurrence due to the challenging surgical resection^[5, 6]. In addition, the spread patterns and treatment strategies differ between colon cancer and RC^[3]. Therefore, in contrast to most CRC bioinformatics research, we specifically focus on RC.

DNA methylation, a type of central epigenetic modification, is involved in diverse biological activities, including organism development, genome regulation, aging, and disease^[7, 8]. Methylation generally occurs on cytosine residues in cytosine-phosphate-guanine (CpG) dinucleotides^[9, 10]. Typically, these CpG sites are located and enriched in large clusters forming CpG islands, found mostly at the promoter regions of genes^[10]. The methylation of CpG islands can recruit repressive methyl-binding proteins, impair transcription factor binding, and condense chromatin structure^[11, 12]. Therefore, CpG islands are generally methylated in promoters of silencing genes and not methylated in transcriptionally active genes^[13]. Furthermore, abnormal hypomethylation of oncogenes and hypermethylation of tumor suppressor genes have been found to play important roles in human tumorigenesis and cancer progression, including those of RC^[14-16]. For example, Ye et al.^[17] performed whole-exome sequencing in 50 RC patients and found pervasive decreased expression of PCDHB3 due to CpG island hypermethylation. In addition, the mRNA expression of MGMT and ERCC1 were shown to be downregulated in RC as a result of the hypermethylation of their promoter regions^[18]. Despite the identification of individual genes with aberrant methylation in RC, comprehensive network research based on methylation profiles and further enrichment analysis of these genes remain largely inadequate.

Over the past decade, gene profiling and next-generation sequencing technology have developed rapidly, permitting advances in the detection of genomic, transcriptional, proteomic, and epigenetic changes^[19, 20]. Bioinformatics analysis based on these multifarious data could provide a new avenue to the understanding of cancer biology through intelligent systems^[21]. However, previous studies based on profiling arrays mostly analyzed either gene expression data for differentially expressed genes (DEGs) or DNA methylation data for differentially methylated genes (DMGs), not both^[22-24]. Conjoint analysis of RC involving both gene expression and methylation profiling microarray datasets remains rare. Thus far, only one study, based on six RC patients, has explored the diagnostic values of methylation-regulated differentially expressed genes (MeDEGs)^[25].

In the present study, an integrated bioinformatics analysis was performed based on gene expression (GSE20842 and GSE68204) and gene methylation (GSE75546) profiling microarrays. We identified MeDEGs by overlapping three datasets and then performing protein–protein interaction (PPI) network analysis. Moreover, MeDEGs related to survival were selected and validated based on The Cancer Genome Atlas (TCGA) database. Through this method, we aimed to identify aberrantly methylated and expressed genes that contribute to patient survival in RC, which may potentially identify novel biomarkers and therapeutic targets.

Methods

Microarray data

In the present study, we obtained gene expression (GSE20842 and GSE68204) and gene methylation (GSE75546) profiling datasets from Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>). For the gene expression profiling datasets, 65 pairs of RC and matched normal samples were obtained from the GSE20842 dataset (platform: GPL4133 Agilent-014850 Whole Human Genome Microarray), whereas 59 RC and 21 matched normal samples were obtained from the GSE68204 dataset (platform: GPL17586 Affymetrix Human Transcriptome Array 2.0). The DNA methylation dataset (GSE75546), based on the GPL13534 platform (Illumina HumanMethylation450 BeadChip), included 12 paired RC and normal samples from six patients.

Identification of MeDEGs

GEO2R (<http://www.ncbi.nlm.nih.gov/geo/geo2r/>) is an online analysis tool that can be used to compare two or more groups of samples in a GEO series to identify deregulated genes. In our study, GEO2R was adopted to screen for DEGs and DMGs between normal and tumor samples separately in three datasets. Because two gene expression datasets (GSE20842 and GSE68204) were to be superimposed to identify overlapping DEGs, we adopted a slightly lower than usual threshold for DEGs, which was $|\log_2\text{FoldChange}| > 1$ and $\text{FDR} < 0.05$. For DMGs, $|t| > 5$ and $\text{FDR} < 0.05$ were considered statistically significant. After the two gene expression datasets and one methylation dataset were overlapped, the upregulated-hypomethylated and downregulated-hypermethylated genes were identified as MeDEGs.

PPI network construction, module analysis, and identification of hub genes

The PPI network of MeDEGs was constructed based on The Search Tool for the Retrieval of Interacting Genes (STRING, <https://string-db.org/>) database. An interaction score > 0.4 (medium confidence) was considered significant. Cytoscape software (<http://www.cytoscape.org/>) was adopted to visualize the network. Then, we used the Molecular Complex Detection (MCODE) clustering tool to screen modules within the PPI network with node and combined score thresholds of ≥ 6 and > 5 , respectively. Hub genes were screened using CytoHubba with the Degree-ranked method. Genes with the highest degree score in each module were considered hub genes.

Functional enrichment analysis

To analyze the MeDEGs for the functional enrichment, Gene Ontology (GO) and pathway enrichment analysis were performed using STRING (<https://string-db.org/>), Kyoto Encyclopedia of Genes and Genomes (KEGG; <https://www.kegg.jp/>), and Reactome (<https://reactome.org/>). $P < 0.05$ was considered statistically significant.

Survival analysis

The clinical and mRNA expression data of 165 RC patients were obtained from TCGA database (<https://portal.gdc.cancer.gov/>). We then extracted the expression matrix of the MeDEGs derived from the

three GEO datasets. Univariate Cox regression was applied to evaluate the association between overall survival and MeDEG expression using the survival package of R. $P < 0.05$ was considered statistically significant. These survival-related MeDEGs were considered prognostic MeDEGs.

Validation of prognostic MeDEGs and hub genes

MEXPRESS (<http://mexpress.be/>) is an intuitive web tool for the fast and straightforward querying and visualization of gene expression and methylation data from TCGA database on a single-gene level^[26, 27]. In our study, MEXPRESS was used to validate the expression and methylation levels of prognostic MeDEGs and hub genes. cBioPortal for Cancer Genomics (<http://cbioportal.org/>) is an open-access resource for exploring, visualizing, and analyzing multidimensional cancer genomics data^[28, 29]. We used cBioPortal to further explore the genetic alterations in these genes in RC.

Results

Identification of MeDEGs in RC

GEO2R was adopted to screen for DEGs and DMGs between normal and tumor samples separately in three datasets. For the expression microarray datasets, we obtained 2187 DEGs (1068 upregulated and 1119 downregulated) and 4556 DEGs (3418 upregulated and 1138 downregulated) from the GSE20842 and GSE68204 datasets, respectively. For the GSE75546 methylation microarray, 6313 DMGs were identified, comprising 1366 hypermethylated genes and 4947 hypomethylated genes. After the differential genes in the three datasets were overlapped, 243 upregulated-hypomethylated and 51 downregulated-hypermethylated genes were identified as MeDEGs (Figure 1).

Functional enrichment analysis

The gene ontology enrichment analysis of MeDEGs is summarized in Table 1. The enriched biological processes included biological adhesion, cell adhesion, extracellular matrix organization, and regulation of cell population proliferation. We also found that these MeDEGs were mainly enriched in protein binding, extracellular matrix structural constituent, and growth factor binding in the molecular function group and in extracellular region in the cell component group. As shown in Table 2, pathway enrichment analysis implied that these genes participated in many cancer-related pathways, including the PI3K-Akt signaling pathway, ECM-receptor interaction, focal adhesion, and degradation of the extracellular matrix.

Table 1. Gene ontology enrichment analysis of 294 MeDEGs

Term ID	Term description	Count	False discovery rate
Biological processes (BP)			
GO:0022610	Biological adhesion	48	2.13E-11
GO:0007155	Cell adhesion	46	1.45E-10
GO:0030198	Extracellular matrix organization	27	4.28E-10
GO:0048522	Positive regulation of cellular process	121	1.52E-07
GO:0002009	Morphogenesis of an epithelium	26	4.78E-07
GO:0042127	Regulation of cell population proliferation	56	5.16E-07
GO:0030154	Cell differentiation	92	1.45E-06
GO:0048869	Cellular developmental process	93	1.88E-06
GO:0016477	Cell migration	35	5.81E-06
GO:0001525	Angiogenesis	20	8.32E-06
Molecular function (MF)			
GO:0005515	Protein binding	149	5.26E-07
GO:0005201	Extracellular matrix structural constituent	11	1.47E-05
GO:0019838	Growth factor binding	13	3.37E-05
GO:0005488	Binding	215	2.30E-04
GO:0005102	Signaling receptor binding	46	5.50E-04
GO:0030246	Carbohydrate binding	15	2.70E-03
GO:0048407	Platelet-derived growth factor binding	4	5.80E-03
GO:0050786	RAGE receptor binding	4	5.80E-03
GO:0005539	Glycosaminoglycan binding	12	1.34E-02
GO:0005509	Calcium ion binding	23	3.19E-02
GO:0005518	Collagen binding	6	3.19E-02
GO:0005198	Structural molecule activity	22	4.24E-02
Cell component (CC)			
GO:0005576	Extracellular region	95	1.34E-15
GO:0031012	Extracellular matrix	28	4.07E-12
GO:0044421	Extracellular region part	61	4.07E-12

GO:0005581	Collagen trimer	15	3.28E-09
GO:0062023	Collagen-containing extracellular matrix	18	3.28E-09
GO:0005615	Extracellular space	44	7.50E-07
GO:0005604	Basement membrane	12	2.50E-06
GO:0005788	Endoplasmic reticulum lumen	20	3.35E-06
GO:0044420	Extracellular matrix component	10	3.35E-06
GO:0098644	Complex of collagen trimers	7	3.41E-06

Table 2. Pathway enrichment analysis of 294 MeDEGs

Term ID	Term description	Count	False discovery rate
KEGG:hsa04974	Protein digestion and absorption	12	6.58E-06
KEGG:hsa04151	PI3K-Akt signaling pathway	18	8.40E-04
KEGG:hsa04510	Focal adhesion	12	3.30E-03
KEGG:hsa04512	ECM-receptor interaction	8	3.30E-03
KEGG:hsa04060	Cytokine-cytokine receptor interaction	13	9.40E-03
KEGG:hsa05150	Staphylococcus aureus infection	5	4.65E-02
Reactome:HSA-1474244	Extracellular matrix organization	28	3.45E-11
Reactome:HSA-2022090	Assembly of collagen fibrils and other multimeric structures	14	9.89E-10
Reactome:HSA-1442490	Collagen degradation	14	1.40E-09
Reactome:HSA-8948216	Collagen chain trimerization	12	3.82E-09
Reactome:HSA-3000178	ECM proteoglycans	13	5.12E-08
Reactome:HSA-216083	Integrin cell surface interactions	13	1.33E-07
Reactome:HSA-1474228	Degradation of the extracellular matrix	15	5.86E-07
Reactome:HSA-186797	Signaling by PDGF	9	2.33E-05
Reactome:HSA-8875878	MET promotes cell motility	8	2.33E-05
Reactome:HSA-419037	NCAM1 interactions	8	3.20E-05
Reactome:HSA-8874081	MET activates PTK2 signaling	7	3.81E-05
Reactome:HSA-3000171	Non-integrin membrane-ECM interactions	8	2.40E-04
Reactome:HSA-9006934	Signaling by Receptor Tyrosine Kinases	19	1.70E-03
Reactome:HSA-6798695	Neutrophil degranulation	19	4.20E-03
Reactome:HSA-	MET interacts with TNS proteins	3	5.60E-03

Reactome:HSA-2243919	Crosslinking of collagen fibrils	4	8.60E-03
Reactome:HSA-422475	Axon guidance	19	1.90E-02
Reactome:HSA-3000170	Syndecan interactions	4	2.66E-02
Reactome:HSA-3000157	Laminin interactions	4	4.12E-02

PPI network construction, module analysis, and identification of hub gene

By uploading the MeDEGs to STRING, we constructed a PPI network with 290 nodes and 578 edges (Figure 2). The enrichment P value of the network was less than $1.0\text{e-}16$, which indicated that these genes were at least partially biologically connected as a group. Using MCODE in Cytoscape, three modules were identified within the PPI network with node and combined score thresholds ≥ 6 and > 5 , respectively (Table 3). These modules were then visualized using STRING (Figure 3). To further investigate their function in RC, pathway enrichment analyses of these modules were performed (Tables 4–6). In addition, three hub genes were identified from three modules using CytoHubba with the Degree-ranked method, namely, COL3A1, FPR1, and PLK1. All of these genes were upregulated-hypomethylated genes.

Table 3. Module analysis within the PPI network

Cluster	Combined Score	Nodes	Edges	Node IDs
1	12	12	66	COL3A1, COL5A2, COL4A2, COL15A1, COL4A1, COL18A1, COL9A3, COL6A3, COL11A1, COL12A1, COL5A1, COL8A1
2	9.857	15	69	CCL20, CCL4, FCN1, FCGR3B, NPY, TREM1, CSF3R, CXCR1, MND A, S100A8, FPR2, S100A9, FPR1, S100A12, GNG4
3	9	9	36	NEK2, MELK, CENPF, CEP55, TPX2, PLK1, CDC25C, RRM2, CDC6

Table 4. Pathway enrichment analysis of MeDEGs in Module 1

Term ID	Term description	Count	False discovery rate
KEGG:hsa04974	Protein digestion and absorption	11	5.16E-24
KEGG:hsa04512	ECM-receptor interaction	4	8.77E-07
KEGG:hsa04510	Focal adhesion	4	1.83E-05
KEGG:hsa04933	AGE-RAGE signaling pathway in diabetic complications	3	6.91E-05
KEGG:hsa04151	PI3K-Akt signaling pathway	4	7.13E-05
KEGG:hsa04926	Relaxin signaling pathway	3	8.87E-05
KEGG:hsa05222	Small cell lung cancer	2	1.80E-03
KEGG:hsa05200	Pathways in cancer	2	4.24E-02
Reactome:HSA-8948216	Collagen chain trimerization	12	2.62E-30
Reactome:HSA-2022090	Assembly of collagen fibrils and other multimeric structures	12	3.61E-29
Reactome:HSA-1442490	Collagen degradation	12	4.86E-29
Reactome:HSA-216083	Integrin cell surface interactions	9	6.30E-19
Reactome:HSA-419037	NCAM1 interactions	7	1.06E-15
Reactome:HSA-186797	Signaling by PDGF	7	5.46E-15
Reactome:HSA-3000178	ECM proteoglycans	7	3.50E-14
Reactome:HSA-3000171	Non-integrin membrane-ECM interactions	6	2.08E-12
Reactome:HSA-9006934	Signaling by Receptor Tyrosine Kinases	8	6.77E-11
Reactome:HSA-8874081	MET activates PTK2 signaling	4	6.78E-09

Table 5. Pathway enrichment analysis of MeDEGs in Module 2

Term ID	Term description	Count	False discovery rate
KEGG:hsa04062	Chemokine signaling pathway	4	3.60E-04
KEGG:hsa05150	Staphylococcus aureus infection	3	3.60E-04
KEGG:hsa04060	Cytokine-cytokine receptor interaction	4	5.60E-04
KEGG:hsa04657	IL-17 signaling pathway	3	5.60E-04
KEGG:hsa05034	Alcoholism	2	4.34E-02
Reactome:HSA-168256	Immune System	13	7.38E-10
Reactome:HSA-6798695	Neutrophil degranulation	9	7.38E-10
Reactome:HSA-168249	Innate Immune System	10	1.14E-08
Reactome:HSA-418594	G alpha (i) signaling events	7	1.79E-07
Reactome:HSA-375276	Peptide ligand-binding receptors	5	4.34E-06
Reactome:HSA-500792	GPCR ligand binding	6	9.94E-06
Reactome:HSA-6783783	Interleukin-10 signaling	3	7.78E-05
Reactome:HSA-449147	Signaling by Interleukins	5	1.30E-04
Reactome:HSA-6799990	Metal sequestration by antimicrobial proteins	2	1.30E-04
Reactome:HSA-444473	Formyl peptide receptors bind formyl peptides and many other ligands	2	1.50E-04
Reactome:HSA-5686938	Regulation of TLR by endogenous ligand	2	7.70E-04
Reactome:HSA-168898	Toll-like Receptor Cascades	3	1.30E-03
Reactome:HSA-380108	Chemokine receptors bind chemokines	2	3.90E-03

Table 6. Pathway enrichment analysis of MeDEGs in Module 3

Term ID	Term description	Count	False discovery rate
KEGG:hsa04110	Cell cycle	3	2.30E-04
KEGG:hsa04114	Oocyte meiosis	2	4.60E-03
KEGG:hsa04914	Progesterone-mediated oocyte maturation	2	4.60E-03
Reactome:HSA-69278	Cell Cycle, Mitotic	7	1.75E-08
Reactome:HSA-69275	G2/M Transition	5	3.18E-07
Reactome:HSA-156711	Polo-like kinase mediated events	3	1.12E-06
Reactome:HSA-69620	Cell Cycle Checkpoints	4	5.97E-05
Reactome:HSA-8854518	AURKA Activation by TPX2	3	5.97E-05
Reactome:HSA-69273	Cyclin A/B1/B2 associated events during G2/M transition	2	7.00E-04
Reactome:HSA-539107	Activation of E2F1 target genes at G1/S	2	7.30E-04
Reactome:HSA-68877	Mitotic Prometaphase	3	7.30E-04
Reactome:HSA-176187	Activation of ATR in response to replication stress	2	9.90E-04
Reactome:HSA-195258	RHO GTPase Effectors	3	1.40E-03

Survival analysis of MeDEGs

We extracted the expression matrix of MeDEGs from TCGA and evaluated their association with survival by univariate Cox regression analysis. As shown in Table 7, three survival-related MeDEGs—NFE2, COMP, and LAMA1—were found to be prognostic MeDEGs. NFE2 and COMP were hypomethylated and upregulated in RC, whereas LAMA1 was hypermethylated and downregulated. Hazard ratios (HRs) are the ratios of the hazard rates corresponding to the conditions described by two levels of an explanatory variable. In survival analysis, a HR value greater than 1 represents a harmful effect, whereas a HR value less than 1 represents a protective effect. Our results suggested that NFE2 and COMP may act as tumor promoters and that LAMA1 may act as a tumor suppressor.

Table 7. Identification of prognostic MeDEGs

Terms	HR	coef	95% CI lower	95% CI upper	Z	P value
NFE2	1.248086	0.221612	0.030114	0.413109	2.268176	0.023318
COMP	1.181946	0.167162	0.009313	0.325011	2.075601	0.037931
LAMA1	0.717404	-0.332117	-0.629092	-0.035141	-2.191887	0.028388

Validation of prognostic MeDEGs and hub genes

To further validate the results, we examined the expression and methylation of the three prognostic MeDEGs and the three hub genes using MEXPRESS. We found that the expression and methylation status of these genes were consistent with our findings, suggesting the reliability and stability of our results. As shown in Table 8, LAMA1 was downregulated, whereas NFE2, COMP, COL3A1, FPR1, and PLK1 were upregulated. With the exception of FPR1, the expression levels of these genes were significantly altered. As for the methylation levels, these genes all exhibited significant differences between RC and normal tissues (Figure 4a–f: NFE2, COMP, COL3A1, FPR1, PLK1, and LAMA1). In addition, the presence of genetic alterations in the six MeDEGs was explored using cBioPortal. We found that more than 30% of the RC patients had at least one gene alteration and that LAMA1 (14%) was the most frequently altered gene of the six queried genes (Figure 5).

Table 8. Validation of the expression levels of the three prognostic MeDEGs and the three hub genes based on TCGA database

Category	Gene ID	logFC	FDR
Downregulated	LAMA1	-3.00398	1.89E-11
Upregulated	NFE2	2.279814	1.87E-04
Upregulated	COMP	5.966875	1.09E-16
Upregulated	COL3A1	1.322626	3.17E-03
Upregulated	FPR1	0.453118	4.31E-01
Upregulated	PLK1	1.522735	3.92E-08

Discussion

DNA methylation has been identified and verified as a critical modification in the field of epigenetics. The maintenance of appropriate DNA methylation plays an important role in the regulation of a variety of biological processes, including chromosomal structure stability and control of gene expression^[30]. DNA hypermethylation can result in the transcriptional silencing of specific tumor suppressor genes, whereas hypomethylation may lead to the activation of oncogenes^[31]. More and more evidence suggests that

aberrant methylation is actively involved in multiple human malignancies, including RC^[32, 33]. For instance, Benard et al.^[34] studied the methylation status of promoter regions of key apoptosis genes in RC patients. They found that methylation of these genes in RC would lead to apoptotic pathway deregulation and would therefore correlate with poor survival and tumor recurrence^[34]. Moreover, methylation of CDKN2A has been proposed to be a negative prognostic factor for the overall survival of RC patients^[35]. Despite the identification of individual genes with aberrant methylation, high-throughput methylation profile research in RC remains largely inadequate. Additionally, conjoint analysis of RC involving both gene expression and methylation profiling datasets is even rarer. Thus far, only one such study, based on six RC patients, has explored the diagnostic values of MeDEGs^[25]. In contrast, we identified and validated MeDEGs based on multicenter clinical samples and also carried out prognostic analysis. Therefore, we believe that there is some merit to our research.

In this study, by integrating DEGs and DMGs, we identified 243 upregulated and 51 downregulated MeDEGs. Some of the 294 MeDEGs have previously been shown to be regulated by methylation. For instance, Sepulveda et al.^[36] found that CDH11 had significantly increased CpG methylation and decreased expression in gastric cancer. CDH11 was also downregulated in colorectal cancer due to its promoter methylation, which was linked to cell cycle arrest in the G0/G1 phase and apoptosis^[37]. Furthermore, the methylation-induced regulation of CDH11 has been found in many other cancers, such as melanoma and bladder cancer^[38-40]. However, the methylation regulation mechanisms of CDH11 and many of the other MeDEGs that we identified has not been revealed in RC and warrant further study.

We further performed gene ontology and pathway enrichment analysis to clarify the roles of these MeDEGs in RC. The enriched biological processes included biological adhesion, cell adhesion, extracellular matrix organization, and regulation of cell population proliferation. Our findings are reasonable because it is well-recognized that these significantly enriched terms are closely related to cancer. For example, many MeDEGs under the cell adhesion term are involved in tumorigenesis and progression, including TGFBI, KIFC3, CDH11, and MMP10. TGFBI (transforming growth factor beta-induced) is an extracellular secreted matrix protein that plays pivotal roles in cell adhesion by binding to integrin at the C-terminus^[41]. Steitz et al.^[42] found that TGFBI could promote ovarian cancer cell migration and was associated with short progression-free survival. TGFBI was also identified to be a hub gene strongly associated with oral squamous cell carcinoma and to be able to promote tumor metastasis^[43]. In addition, the epigenetic silencing of TGFBI by DNA methylation can confer resistance to trastuzumab in breast cancer^[44]. Furthermore, Wang et al.^[45] found the aberrant expression of TGFBI in RC. However, research into the methylation regulation of TGFBI in RC has not been reported. Pathway analysis also revealed that these MeDEGs were enriched in the PI3K-Akt signaling pathway, which was involved in multiple types of malignancies^[46, 47], including RC^[48]. In the current study, we identified many aberrantly methylated and expressed genes participating in the PI3K-Akt signaling pathway, which could provide a foundation for experimental research in RC.

After construction of a PPI network for MeDEGs, we observed that some modules in the network clearly had more interactions in which genes showed closer connections. Therefore, we performed module analysis using MCODE and identified three modules. All genes in Module 1 belong to the collagen family, such as COL3A1, COL5A2, COL4A2, and COL15A1. Collagens play structural roles in the mechanical properties of tissues and can interact with cells via several receptor families to regulate cell proliferation, migration, and differentiation^[49]. Enrichment analysis revealed that these genes in Module 1 were enriched in many cancer-related pathways, including focal adhesion^[50], the PI3K-Akt signaling pathway^[47], collagen degradation^[51], and pathways in cancer. We also found that MeDEGs in Module 2 were related to the immune system, chemokine signaling pathway, and signaling by interleukins. In recent decades, there has been a revolution in the development of therapeutics aimed at harnessing the immune system to mount effective antitumor responses^[52]. Tremendous progress has been made in the immunotherapy field, such as engineered T-cell transfer, PD-1 blockade, and dendritic cell vaccination^[53, 54]. Our findings may provide some clues for future immunotherapy research. In addition, we found that the MeDEGs in Module 3 were mainly enriched in cell cycle-related pathways, including G2/M transition, cell cycle checkpoints, mitotic prometaphase, and activation of E2F1 target genes at G1/S. These findings are reasonable because it is well-known that dysregulation of the cell cycle progression is closely involved in tumorigenesis and tumor development.

In the present study, three prognostic MeDEGs—NFE2, COMP, and LAMA1—were identified based on univariate Cox regression analysis. The hematopoietic transcription factor NFE2 is an important regulator of erythroid and megakaryocytic gene expression^[55]. NFE2 is significantly overexpressed in most patients with myeloproliferative neoplasms^[56], although it has not been linked to other types of cancer. COMP, an extracellular matrix protein, can suppress apoptosis by blocking the activation of caspase-3^[57]. Sun et al.^[58] found that hepatic stellate cell-derived COMP contributes to hepatocellular carcinoma epithelial-mesenchymal transition. COMP is also hypomethylated and upregulated in gastric cancer^[59]. However, no research into COMP has been reported for RC. Of the six queried genes (NFE2, COMP, COL3A1, FPR1, PLK1, and LAMA1), we found LAMA1 to be the most frequently altered gene in RC. However, its effect on tumorigenesis is controversial. Promoter methylation and downregulation of LAMA1 has been observed in small intestinal neuroendocrine tumors^[60]. However, Meng et al.^[61] found that LAMA1 was upregulated and could promote tumor cell proliferation and migration in esophageal squamous cell carcinoma. Therefore, the regulatory mechanism of LAMA1 requires further study.

Several limitations to this study should be acknowledged. First, the DNA methylation dataset used to identify DMGs only included six pairs of RC and normal samples. The small sample number may introduce false positives and reduce the reliability of our results. Second, our study only focused on hypermethylated downregulated genes and hypomethylated upregulated genes. Thus, we did not investigate contra-regulated genes. Finally, our study lacks experimental validation of our results.

Conclusions

In conclusion, we identified MeDEGs and explored their related pathways and functions in RC through comprehensive bioinformatics analysis. Moreover, three prognostic MeDEGs—NFE2, COMP, and LAMA1—and three hub genes—COL3A1, FPR1, and PLK1—were selected and validated based on several databases. Through this approach, we have provided new insights into the methylation-mediated regulatory mechanisms of RC progression.

Abbreviations

RC rectal cancer

CRC colorectal cancer

MeDEGs methylation-regulated differentially expressed genes

GEO Gene Expression Omnibus

DEGs differentially expressed genes

DMGs differentially methylated genes

PPI protein-protein interaction

MCODE Molecular Complex Detection

TCGA The Cancer Genome Atlas

GO Gene Ontology

STRING The Search Tool for the Retrieval of Interacting Genes

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The data used to support the findings of this study are included in the article.

Competing interests

Not applicable.

Funding

The study was supported by the Educational Commission of Liaoning Province, People's Republic of China (Grant number L2014294).

Authors' contributions

The authors contributed to this study and manuscript in the following manner: data collection, CZ and DM; statistical analysis, CZ and DM; writing and editing, CZ and SC; supervision, CG; funding acquisition, CG. All authors read and approved the final manuscript.

Acknowledgements

We thank BioMed Proofreading (<http://www.biomedproofreading.com/>) for its linguistic assistance during the preparation of this manuscript.

References

- [1] Giani A, Famularo S, Riva L, et al. Association between specific presurgical anthropometric indexes and morbidity in patients undergoing rectal cancer resection. *Nutrition*. 2020. 75-76: 110779.
- [2] Sipaviciute A, Sileika E, Burneckis A, Dulskas A. Late gastrointestinal toxicity after radiotherapy for rectal cancer: a systematic review. *Int J Colorectal Dis*. 2020 .
- [3] Wu J, Wu M, Wu Q. Identification of potential metabolite markers for colon cancer and rectal cancer using serum metabolomics. *J Clin Lab Anal*. 2020 : e23333.
- [4] Qiu H, Yu D, Ye S, Shan R, Ai J, Shi J. Long-term oncological outcomes in robotic versus laparoscopic approach for rectal cancer: a systematic review and meta-analysis. *Int J Surg*. 2020 .
- [5] Olson CH. Current Status of the Management of Stage I Rectal Cancer. *Curr Oncol Rep*. 2020. 22(4): 40.
- [6] Holm M, Joenväärä S, Saraswat M, et al. Preoperative Radiotherapy Leads to Significant Differences in the Plasma Protein Profile of Rectal Cancer Patients. *Oncology*. 2020 : 1-8.
- [7] Liu S, Li Q, Chen K, et al. The emerging molecular mechanism of m6A modulators in tumorigenesis and cancer progression. *Biomed Pharmacother*. 2020. 127: 110098.
- [8] Fröhlich J, Grundhoff A. Epigenetic control in Kaposi sarcoma-associated herpesvirus infection and associated disease. *Semin Immunopathol*. 2020 .

- [9] Ge R, Wang Z, Montironi R, et al. Epigenetic modulations and lineage plasticity in advanced prostate cancer. *Ann Oncol*. 2020. 31(4): 470-479.
- [10] Yu L, Xia K, Cen X, et al. DNA methylation of noncoding RNAs: new insights into osteogenesis and common bone diseases. *Stem Cell Res Ther*. 2020. 11(1): 109.
- [11] Klutstein M, Nejman D, Greenfield R, Cedar H. DNA Methylation in Cancer and Aging. *Cancer Res*. 2016. 76(12): 3446-50.
- [12] Ginder GD, Williams DC Jr. Readers of DNA methylation, the MBD family as potential therapeutic targets. *Pharmacol Ther*. 2018. 184: 98-111.
- [13] Ebrahimi V, Soleimanian A, Ebrahimi T, et al. Epigenetic modifications in gastric cancer: Focus on DNA methylation. *Gene*. 2020. 742: 144577.
- [14] Liang Z, Kidwell RL, Deng H, Xie Q. Epigenetic N6-methyladenosine modification of RNA and DNA regulates cancer. *Cancer Biol Med*. 2020. 17(1): 9-19.
- [15] Constâncio V, Nunes SP, Henrique R, Jerónimo C. DNA Methylation-Based Testing in Liquid Biopsies as Detection and Prognostic Biomarkers for the Four Major Cancer Types. *Cells*. 2020. 9(3).
- [16] Kokelaar RF, Jones H, Beynon J, Evans ME, Harris DA. Meta-analysis of the prognostic value of CpG island methylator phenotype in rectal cancer. *Int J Colorectal Dis*. 2018. 33(8): 995-1000.
- [17] Ye W, Ling S, Liu RY, et al. Exome sequencing reveals the genetic landscape and frequent inactivation of PCDHB3 in Chinese rectal cancers. *J Pathol*. 2018. 245(2): 222-234.
- [18] Shalaby SM, El-Shal AS, Abdelaziz LA, Abd-Elbary E, Khairy MM. Promoter methylation and expression of DNA repair genes MGMT and ERCC1 in tissue and blood of rectal cancer patients. *Gene*. 2018. 644: 66-73.
- [19] Kamps R, Brandão RD, Bosch BJ, et al. Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. *Int J Mol Sci*. 2017. 18(2).
- [20] Wu D, Wang X. Application of clinical bioinformatics in lung cancer-specific biomarkers. *Cancer Metastasis Rev*. 2015. 34(2): 209-16.
- [21] Banwait JK, Bastola DR. Contribution of bioinformatics prediction in microRNA-based cancer therapeutics. *Adv Drug Deliv Rev*. 2015. 81: 94-103.
- [22] Yin L, Xiao L, Gao Y, et al. Comparative bioinformatical analysis of pancreatic head cancer and pancreatic body/tail cancer. *Med Oncol*. 2020. 37(5): 46.
- [23] Li J, Ye M, Zhou C. Expression Profile and Prognostic Values of HOXA Family Members in Laryngeal Squamous Cell Cancer. *Front Oncol*. 2020. 10: 368.

- [24] Tong Y, Song Y, Deng S. Combined analysis and validation for DNA methylation and gene expression profiles associated with prostate cancer. *Cancer Cell Int.* 2019. 19: 50.
- [25] Wei J, Li G, Zhang J, et al. Integrated analysis of genome-wide DNA methylation and gene expression profiles identifies potential novel biomarkers of rectal cancer. *Oncotarget.* 2016. 7(38): 62547-62558.
- [26] Koch A, Jeschke J, Van Criekinge W, van Engeland M, De Meyer T. MEXPRESS update 2019. *Nucleic Acids Res.* 2019. 47(W1): W561-W565.
- [27] Koch A, De Meyer T, Jeschke J, Van Criekinge W. MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data. *BMC Genomics.* 2015. 16: 636.
- [28] Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012. 2(5): 401-4.
- [29] Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013. 6(269): pl1.
- [30] Ellis L, Atadja PW, Johnstone RW. Epigenetics in cancer: targeting chromatin modifications. *Mol Cancer Ther.* 2009. 8(6): 1409-20.
- [31] Morgan AE, Davies TJ, Mc Auley MT. The role of DNA methylation in ageing and cancer. *Proc Nutr Soc.* 2018. 77(4): 412-422.
- [32] Molinari C, Casadio V, Foca F, et al. Gene methylation in rectal cancer: predictive marker of response to chemoradiotherapy. *J Cell Physiol.* 2013. 228(12): 2343-9.
- [33] Saghafeinia S, Mina M, Riggi N, Hanahan D, Ciriello G. Pan-Cancer Landscape of Aberrant DNA Methylation across Human Tumors. *Cell Rep.* 2018. 25(4): 1066-1080.e8.
- [34] Benard A, Zeestraten EC, Goossens-Beumer IJ, et al. DNA methylation of apoptosis genes in rectal cancer predicts patient survival and tumor recurrence. *Apoptosis.* 2014. 19(11): 1581-93.
- [35] Exner R, Pulverer W, Diem M, et al. Potential of DNA methylation in rectal cancer as diagnostic and prognostic biomarkers. *Br J Cancer.* 2015. 113(7): 1035-45.
- [36] Sepulveda JL, Gutierrez-Pajares JL, Luna A, et al. High-definition CpG methylation of novel genes in gastric carcinogenesis identified by next-generation sequencing. *Mod Pathol.* 2016. 29(2): 182-93.
- [37] Yuan S, Li L, Xiang S, Jia H, Luo T. Cadherin-11 is inactivated due to promoter methylation and functions in colorectal cancer as a tumour suppressor. *Cancer Manag Res.* 2019. 11: 2517-2529.
- [38] Gao L, van den Hurk K, Moerkerk P, et al. Promoter CpG island hypermethylation in dysplastic nevus and melanoma: CLDN11 as an epigenetic biomarker for malignancy. *J Invest Dermatol.* 2014. 134(12): 2957-2966.

- [39] Lin YL, Gui SL, Ma JG. Aberrant methylation of CDH11 predicts a poor outcome for patients with bladder cancer. *Oncol Lett*. 2015. 10(2): 647-652.
- [40] Sandoval-Bórquez A, Polakovicova I, Carrasco-Véliz N, et al. MicroRNA-335-5p is a potential suppressor of metastasis and invasion in gastric cancer. *Clin Epigenetics*. 2017. 9: 114.
- [41] Han KE, Choi SI, Kim TI, et al. Pathogenesis and treatments of TGFBI corneal dystrophies. *Prog Retin Eye Res*. 2016. 50: 67-88.
- [42] Steitz AM, Steffes A, Finkernagel F, et al. Tumor-associated macrophages promote ovarian cancer cell migration by secreting transforming growth factor beta induced (TGFBI) and tenascin C. *Cell Death Dis*. 2020. 11(4): 249.
- [43] Wang BJ, Chi KP, Shen RL, et al. TGFBI Promotes Tumor Growth and is Associated with Poor Prognosis in Oral Squamous Cell Carcinoma. *J Cancer*. 2019. 10(20): 4902-4912.
- [44] Palomeras S, Diaz-Lagares Á, Viñas G, et al. Epigenetic silencing of TGFBI confers resistance to trastuzumab in human breast cancer. *Breast Cancer Res*. 2019. 21(1): 79.
- [45] Wang DZ, Chen GY, Li YF, Zhang NW. Comprehensive analysis of long non-coding RNA and mRNA expression profile in rectal cancer. *Chin Med J (Engl)*. 2020 .
- [46] War AR, Dang K, Jiang S, et al. Role of cancer stem cells in the development of giant cell tumor of bone. *Cancer Cell Int*. 2020. 20: 135.
- [47] Le Cras TD, Goines J, Lakes N, et al. Constitutively active PIK3CA mutations are expressed by lymphatic and vascular endothelial cells in capillary lymphatic venous malformation. *Angiogenesis*. 2020 .
- [48] Fernandes JM, Jandrey E, Koyama FC, et al. Metformin as an Alternative Radiosensitizing Agent to 5FU During Neoadjuvant Treatment for Rectal Cancer. *Dis Colon Rectum*. 2020 .
- [49] Ricard-Blum S. The collagen family. *Cold Spring Harb Perspect Biol*. 2011. 3(1): a004978.
- [50] Paluch EK, Aspalter IM, Sixt M. Focal Adhesion-Independent Cell Migration. *Annu Rev Cell Dev Biol*. 2016. 32: 469-490.
- [51] Brassart-Pasco S, Brézillon S, Brassart B, Ramont L, Oudart JB, Monboisse JC. Tumor Microenvironment: Extracellular Matrix Alterations Influence Tumor Progression. *Front Oncol*. 2020. 10: 397.
- [52] Desrichard A, Snyder A, Chan TA. Cancer Neoantigens and Applications for Immunotherapy. *Clin Cancer Res*. 2016. 22(4): 807-12.

- [53] Yang Y. Cancer immunotherapy: harnessing the immune system to battle cancer. *J Clin Invest*. 2015. 125(9): 3335-7.
- [54] van Willigen WW, Bloemendal M, Gerritsen WR, Schreibelt G, de Vries I, Bol KF. Dendritic Cell Cancer Therapy: Vaccinating the Right Patient at the Right Time. *Front Immunol*. 2018. 9: 2265.
- [55] Hung HL, Kim AY, Hong W, Rakowski C, Blobel GA. Stimulation of NF-E2 DNA binding by CREB-binding protein (CBP)-mediated acetylation. *J Biol Chem*. 2001. 276(14): 10715-21.
- [56] Peeken JC, Jutzi JS, Wehrle J, et al. Epigenetic regulation of NFE2 overexpression in myeloproliferative neoplasms. *Blood*. 2018. 131(18): 2065-2073.
- [57] Gagarina V, Carlberg AL, Pereira-Mouries L, Hall DJ. Cartilage oligomeric matrix protein protects cells against death by elevating members of the IAP family of survival proteins. *J Biol Chem*. 2008. 283(1): 648-59.
- [58] Sun L, Wang Y, Wang L, et al. Resolvin D1 prevents epithelial-mesenchymal transition and reduces the stemness features of hepatocellular carcinoma by inhibiting paracrine of cancer-associated fibroblast-derived COMP. *J Exp Clin Cancer Res*. 2019. 38(1): 170.
- [59] Liang Y, Zhang C, Dai DQ. Identification of DNA methylation-regulated differentially-expressed genes and related pathways using Illumina 450K BeadChip and bioinformatic analysis in gastric cancer. *Pathol Res Pract*. 2019. 215(10): 152570.
- [60] Fotouhi O, Adel Fahmideh M, Kjellman M, et al. Global hypomethylation and promoter methylation in small intestinal neuroendocrine tumors: an in vivo and in vitro study. *Epigenetics*. 2014. 9(7): 987-97.
- [61] Meng X, Chen X, Lu P, et al. MicroRNA-202 inhibits tumor progression by targeting LAMA1 in esophageal squamous cell carcinoma. *Biochem Biophys Res Commun*. 2016. 473(4): 821-827.

Figures

GSE20842
upregulated

GSE68204
upregulated

GSE75546
hypomethylated

GSE20842
downregulated

GSE68204
downregulated

GSE75546
hypermethylated

Figure 1

Identification of MeDEGs

Figure 3

Module analysis within the PPI network. a Module 1. b Module 2. c Module 3.

Figure 4

Page 24/25

Validation of the methylation levels of the three prognostic MeDEGs and the three hub genes based on TCGA database. a NFE2. b COMP. c COL3A1. d FPR1. e PLK1. f LAMA1.

Figure 5

Genetic alterations in the three prognostic MeDEGs and three hub genes in TCGA database. a Genetic alteration frequency of the six queried genes. Different colors represent different types of genetic alterations. b Summary of the alterations per sample. Each sample is presented in a column with each gene in a row.