Combined Analysis of the Aberrant Epigenetic Alteration of Colorectal Cancer

Xiao-liang Xing

Hunan University of Medicine

Zhi-Yong Yao

Hunan University of Medicine

Ti Zhang

Hunan University of Medicine

Ning Zhu

Hunan University of Medicine

Xingyu Liu

Hunan University of Medicine

Liyun Xiao

Hunan University of Medicine

Kang Cheng

Hunan University of Medicine

Zhi Huang

Hunan University of Medicine

Yuanwu Liu (✉ yuanwu_liu@126.com)

China Agricultural University <https://orcid.org/0000-0001-5132-6252>

Jing Peng

Xiangya Hospital Central South University

Primary research

Keywords: Colon adenocarcinomas, Rectal adenocarcinomas, Methylation, Survival analysis, Pathologic TNM

Posted Date: June 30th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-37093/v1>

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

[Read Full License](#)

Abstract

Background

Colorectal cancer (CRC) is the third most common cancer which could be classified as colon adenocarcinoma (COAD) and rectum adenocarcinoma (READ) mainly. Accumulating evidence indicated methylations are involved in multiple tumors.

Methods

To know the effect of abnormal methylation in COAD, READ and CRC, we downloaded methylation and mRNA data of COAD and READ from The Cancer Genome Atlas (TCGA) database. And then, we used DESeq2, ChAMP, DAVID 6.8, Cytoscape_3.7.2 and Correlation analysis to identify the potential biomarkers.

Results

We obtain 12 potential biomarkers (APBB1, CDC42SE2, EIF4E3, FBXO17, FES, GNPAT1, HSPA1A, OSBPL3, RORC, SALL1, SPEG, and TCF7L1) directly associated with the pathologic TNM of COAD maybe regulated by 28 differential methylations; 2 potential biomarkers (AQP1 and HOXA3) directly associated with pathologic NM maybe regulated by 8 differential methylations; and 15 potential biomarkers (ADAMTSL3, ANXA9, APBB1, AQP1, C2CD4A, CLIP3, DNAJC15, EIF4E3, FAM160A1, GNG4, HLX, HSPA1A, LAYN, NR3C2, and SYT1) directly associated with the pathologic TNM of COAD maybe regulated by 29 differential methylations. Furthermore, we construct the network of differential methylation and differential expression genes. In addition, we also obtain 1 methylation (cg18149207), 2 methylations (cg02000808 and cg21134232) and 2 methylations (cg04408595 and cg19413725) associated with the overall survival.

Conclusion

Our results provide an analysis of theoretical knowledge and clinical outcomes, but more researches are needed to confirm our findings.

1. Background

Colorectal cancer (CRC) is the second most common cancer diagnosed in females and third in males, accounts for approximately 10% of all annually diagnosed cancers and *cancer*-related deaths[1]. There are approximately 900,000 deaths annually[1]. In the past decades, the mortality associated with CRC diagnosis has declined progressively which could be attributed to cancer screening programs, improved surgical techniques, and the availability of more-effective systemic therapies for early-stage and advanced-stage disease and so on[2, 3]. Biomarkers mostly refer to DNA, mRNA, microRNA (miRNA), epigenetic changes or antibodies are playing an increasingly important role in the screen and treatment of CRC[4].

Epigenetics is first illustrated by Conrad H. Waddington in 1942[5]. It is now clear that multiple processes of CRC stem cells are the result of the progressive accumulation of genetic and epigenetic alterations, which could inactivate tumor-suppressor genes and activate oncogenes and leads to tumorigenesis[6-8]. DNA methylation is a major epigenetic modification. Abnormal methylation could affect the functions of crucial genes by altering their expression in tumorigenesis. Several studies have demonstrated that DNA methylation exerted an early event, and new efforts are focused on finding biomarkers for early disease detection, prognostication, and treatment selection[9]. In this study, we used systemic analysis to identify a group of candidate prognosis gene, which may be regulated by DNA methylation. Our study might be the groundwork for further elucidation of the CRC mechanism and screening of the diagnostic biomarkers.

2. Material And Method

2.1 Data source and Data processing

In this study, mRNA expression, DNA methylation and the corresponding clinical information of COAD and READ were obtained from TCGA database (<https://portal.gdc.cancer.gov/>). In COAD, 497 (41 control vs 456 tumor) samples were included in the gene expression profiles while 334 (38 control vs 296 tumor) samples were included in the methylation analysis. For READ, 176 samples (10 control vs 166 tumor) were included in the gene expression profiles while 105 samples (7 control vs 98 tumor) were included in the methylation analysis. The mRNA expression profile data were analyzed by DESeq2 package in R software[10], $\text{padj} < 0.05$, $|\log\text{FC}| \geq 0.5$ and $\text{basemean} > 50$ as selected criteria for differential expression genes. The methylation levels were analyzed by ChAMP package in R software[11], $\text{padj} < 0.05$, $|\log\text{FC}| \geq 0.2$ as cutoff for differential methylation probes selection.

2.2 Correlation analysis

In this study, all of these overlap genes between differential expression genes and differential methylation genes were used for the correlation analysis. And the specific criteria for correlation analysis was set like that $p\text{ value} \leq 0.05$ and $r \geq -0.3$.

2.3 Survival Analysis

The genes which expression significantly changed in tumor samples were used for survival analysis by RegParallel and survival packages via Cox Proportional Hazards regression. According to the medium value, all differentially expressed mRNAs data were transformed to low expression and high expression group. The gene which significantly correlated with survival rate were selected as $p\text{value} < 0.05$ and used for the next analysis.

For pathologic TNM, the samples were divided into two groups according to the tumor pathology. And then the survival analyses were performed by IBM SPSS statistics 22, with log-rank $P < 0.05$ indicating a significant correlation with overall survival outcomes.

2.4Statistical analysis

A repeated-measure ANOVA followed by Bonferroni post hoc tests or unpaired two-tail Student’s t test was used as indicated. All statistical analyses were performed using the Prism 6.01 (Graph Pad Software, San Diego, CA).

3. Result

3.1Demography

After excluding those patients without methylation or clinical information, 294COAD and 97 READ patients were included in the study. The clinical and pathological information of these patients displayed in Table 1-2. In COAD cohort, 3.40% of patients were less than 30–39 years old, 11.22% were 40–49 years old, 18.71% were 50–59 years old, 26.19% were 60–69 years old, 25.85% were 70–79 years old, and 14.63% were above 80 years old. There were, respectively, 44 patients with pathologic TNM stage I, 114 patients with pathologic TNM stage II, 85 patients with pathologic TNM stage III, 41 patients with pathologic TNM stage IV, and 10 patients with an unknown TNM stage in our study (table 1).

In READ cohort, 3.09% of patients were less than 30–39 years old, 14.43% were 40–49 years old, 22.68% were 50–59 years old, 23.71% were 60–69 years old, 31.96% were 70–79 years old, and 4.12% were above 80 years old. There were, respectively, 11 patients with pathologic TNM stage I, 29 patients with pathologic TNM stage II, 35 patients with pathologic TNM stage III, 13 patients with pathologic TNM stage IV, and 9 patients with an unknown TNM stage in our study (table 2).

Table 1
COAD clinical characteristics

Colon Clinical variables	Clinical values (N=294)
Sex (Male/Female)	158/136
Age (Mean + SEM)	64.94 ± 0.7738
Race (Asian/Black/White/NA)	11/58/206/19
Pathologic stage (I/II/III/IV/V/NA)	44/114/85/41/10
Pathologist T stage (T1/T2/T3/T4/NA)	7/43/203/10/1
Pathologic N stage (N0/N1+2/NA)	171/74/49
Pathologic M stage (M0/M1/NA)	198/41/55

Table 2
READ clinical characteristics

Rectum Clinical variables	Clinical values (N=97)
Sex (Male/Female)	53/44
Age (Mean + SEM)	62.86 ± 1.251
Race (Asian/Black/White/NA)	1/5/76/15
Pathologic stage (I/II/III/IV/V/NA)	11/29/35/13/9
Pathologist T stage (T1/T2/T3/T4/NA)	4/12/59/11/1
Pathologic N stage (N0/N1+2/ NA)	42/30/25
Pathologic M stage (M0/M1/NA)	69/12/16

3.2 Differential methylation and expression analysis

We downloaded 334 COAD (38 control vs 296 tumor) and 105 READ (7 control vs 98 tumor) methylation data from the TCGA database and set thresholds for the relevant parameters ($p_{adj} < 0.05$ and $|\log FC| \geq 0.2$) to identify the differentially methylation probes (DMPs). There were 41847 DMPs (21079 hypermethylated, 20768 hypomethylated) and 42841 DMPs (27250 hypermethylated, 15591 hypomethylated) which had been identified in COAD and READ respectively (Fig 1a-b). Combined all methylation samples from COAD and READ, 40936 DMPs (21998 hypermethylated, 18938 hypomethylated) had been identified (Fig 1c). By considering the CpG content and the neighboring context, the hypomethylation rate of island was shown to be the highest (71.87%, 67.63%, 69.09%) while the hypermethylation rate of open sea was shown to be the highest (70.31%, 65.31%, 69.65%) in COAD, READ, and CRC respectively (Fig 1d-f). Examining the sites surrounding genes revealed that the hypomethylation rate of body was shown to be the highest (26.31%, 27.65%, 26.48%) while the hypermethylation rate of IGR was shown to be the highest (39.71%, 38.44%, 39.58%) in COAD, READ, and CRC respectively (Fig g-i).

Similarly, we also downloaded 497 COAD (41 control vs 456 tumor) and 176 READ (10 control and 166 tumors) mRNA sequencing data from TCGA database and set thresholds for the relevant parameters ($p_{adj} < 0.05$, $|\log FC| \geq 0.5$, and $base\ mean \geq 50$) to identify the differentially expression genes (DEGs) by DESeq2 analysis standard workflow. After analysis, 7550 DEGs (4089 up-regulated, 3461 down-regulated) had been identified in COAD, and 6954 DEGs (3791 up-regulated, 3163 down-regulated) had been identified in READ (Fig 2a-b). When we combined all collected samples in COAD and READ, 7616 DEGs (4111 up-regulated, 3505 down-regulated) had been identified in CRC (Fig 2c).

By cross analysis of DEGs and DMGs, we found that there were 2080, 2064 and 2118 overlap genes in COAD, READ and CRC respectively, and 9221, 7647 and 8821 corresponding DMPs (Fig 2d-f). And by cross analysis of DEGs and DMGs located in TSS1500, TSS200 and 5'UTR, we found that there were 1131,

1078 and 1130 overlap genes in COAD, READ and CRC respectively which corresponds with 3549, 2685, and 3392 DMPs (Fig 2g-i).

3.3 Correlation Analysis

Previous studies indicated that the relationship between methylation and genes expression is negative correlation. To further narrow-down target genes which potentially regulated by methylation in CRC, we introduced correlation analysis and found that there were 348, 252 and 316 overlap genes in COAD, READ and CRC respectively between DMGs and DEGs under specific criteria for DMGs (p -value ≤ 0.05 and $r \leq -0.3$). The methylation degree distributions of top three hypermethylated and hypomethylated genes for COAD, READ and CRC patients were shown in figure 3. And the top 3 negative correlations of those genes for COAD, READ and CRC patients were also shown in figure 4.

To further know the biological effects of those overlap genes of DMGs and DEGs with negative correlation, we performed GO analysis. The result of GO analysis for COAD, READ and CRC revealed that biological process (BP), cellular component (CC) and molecular functions (MF) were enriched and displayed in supplementary table 1. We set p value ≤ 0.05 and FDR cutoff ≤ 0.05 , and found 3 BP, 1 CC and 2 MF were enriched in COAD involved 257 DEGs; 1 MF was enriched in READ involved 33 DEGs; 3 BP and 1 MF were enriched in CRC involved 166 DEGs (Fig 5a-d).

3.4 Survival Analysis

In order to evaluate the effect of methylation regulated genes on COAD, READ and CRC patient's prognosis, we conducted the Kaplan–Meier survival analysis and univariate Cox regression analysis. We found that 24 out of 348 genes were associated with the patient's overall survival (OS). Patients with low expression of 11 genes (APBB1, FBXO17, FES, GSTM2, HSPA1A, NPTX2, OSBPL3, SLC43A3, SPEG, TCF7L1, and TME) and high expression of 13 genes (CBFA2T3, CDC42SE2, EDNRB, EIF4E3, GNPAT1, HHIP, NRG1, REP15, RORC, SALL1, TSPAN11, UNC5C, and ZNF132) exhibited better OS in COAD. And 16 out of 252 genes were associated with the READ patient's overall survival (OS). Patients with low expression of 11 high expression genes (ABCC2, AQP1, CLSTN3, EPHX1, HOXA3, HOXB3, HYAL2, PGC, S100A11, ST3GAL2, and ZNF415) and 5 high expression genes (C2CD4A, EPHX4, EREG, FCHO1, and NOL12) exhibited better OS in READ. And 28 out of 316 differential genes were associated with the all patient's overall survival (OS). Patients with low expression of these 15 genes (ADAMTSL3, ANXA9, APBB1, AQP1, CBFA2T3, CLIP3, DNAJC15, GNG4, HSPA1A, LAYN, NDRG4, NPTX2, SMAD9, SYT1, TMEM106A, ZNF132) and high expression of these 13 genes (C2CD4A, DPP10, EIF4E3, EPHX4, FAM160A1, HHIP, HLX, NR3C2, RELN, REP15, TSPAN11, UNC5C, and VSIG2) exhibited better OS in CRC.

Table 3
Survival analyses of COAD, READ and CRC

COAD					
Symbol	LogRank	HR	HRlower	HRupper	P
APBB1	0.027	1.58	1.05	2.39	0.029
CBFA2T3	0.013	0.59	0.39	0.90	0.014
CDC42SE2	0.009	0.58	0.38	0.88	0.010
EDNRB	0.045	0.65	0.43	0.99	0.047
EIF4E3	0.030	0.63	0.41	0.96	0.031
FBXO17	0.047	1.51	1.00	2.27	0.048
FES	0.015	1.66	1.10	2.51	0.017
GNPNAT1	0.018	0.61	0.41	0.92	0.019
GSTM2	0.017	1.64	1.09	2.47	0.018
HHIP	0.030	0.64	0.42	0.96	0.032
HSPA1A	0.000	2.18	1.42	3.35	0.000
NPTX2	0.029	1.57	1.04	2.37	0.030
NRG1	0.013	0.59	0.39	0.90	0.014
OSBPL3	0.017	1.65	1.09	2.48	0.018
REP15	0.024	0.63	0.41	0.94	0.026
RORC	0.035	0.65	0.43	0.97	0.037
SALL1	0.028	0.63	0.41	0.95	0.029
SLC43A3	0.021	1.62	1.07	2.43	0.022
SPEG	0.020	1.61	1.07	2.43	0.022
TCF7L1	0.035	1.55	1.03	2.33	0.037
TMEM106A	0.003	1.85	1.22	2.79	0.004
TSPAN11	0.018	0.60	0.39	0.92	0.019
UNC5C	0.011	0.58	0.38	0.89	0.012
ZNF132	0.034	0.64	0.43	0.97	0.036
READ					
Symbol	LogRank	HR	HRlower	HRupper	P

ABCC2	0.041	2.33	1.01	5.35	0.047
AQP1	0.043	2.30	1.00	5.26	0.049
C2CD4A	0.015	0.37	0.16	0.85	0.020
CLSTN3	0.006	3.19	1.34	7.56	0.009
EPHX1	0.027	2.40	1.08	5.35	0.032
EPHX4	0.004	0.31	0.14	0.73	0.007
EREG	0.014	0.37	0.16	0.84	0.017
FCH01	0.037	0.43	0.19	0.98	0.043
HOXA3	0.027	2.48	1.08	5.67	0.032
HOXB3	0.020	2.58	1.13	5.89	0.025
HYAL2	0.026	2.43	1.09	5.42	0.031
NOL12	0.016	0.37	0.16	0.86	0.020
PGC	0.020	2.53	1.13	5.68	0.024
S100A11	0.023	2.58	1.11	6.00	0.028
ST3GAL2	0.009	2.92	1.26	6.77	0.012
ZNF415	0.042	2.31	1.01	5.27	0.048
CRC					
Symbol	LogRank	HR	HRlower	HRupper	P
ADAMTSL3	0.031	1.48	1.03	2.13	0.032
ANXA9	0.010	1.61	1.12	2.31	0.011
APBB1	0.035	1.47	1.03	2.11	0.036
AQP1	0.028	1.49	1.04	2.14	0.029
C2CD4A	0.011	0.63	0.44	0.90	0.012
CBFA2T3	0.005	0.59	0.41	0.86	0.006
CLIP3	0.043	1.45	1.01	2.07	0.044
DNAJC15	0.035	0.68	0.48	0.98	0.036
DPP10	0.034	0.68	0.47	0.97	0.035
EIF4E3	0.027	0.66	0.46	0.96	0.028
EPHX4	0.017	0.65	0.45	0.93	0.018

FAM160A1	0.006	0.60	0.41	0.87	0.006
GNG4	0.043	1.45	1.01	2.07	0.045
HHIP	0.015	0.64	0.45	0.92	0.016
HLX	0.015	1.56	1.09	2.24	0.015
HSPA1A	0.000	2.01	1.39	2.92	0.000
LAYN	0.012	1.58	1.10	2.27	0.013
NDRG4	0.049	1.43	1.00	2.05	0.050
NPTX2	0.042	1.45	1.01	2.08	0.043
NR3C2	0.033	0.68	0.47	0.97	0.034
RELN	0.026	0.66	0.46	0.95	0.027
REP15	0.029	0.67	0.47	0.96	0.030
SMAD9	0.038	0.68	0.48	0.98	0.039
SYT1	0.009	1.62	1.12	2.33	0.010
TMEM106A	0.004	1.70	1.18	2.44	0.004
TSPAN11	0.003	0.57	0.39	0.83	0.003
UNC5C	0.027	0.66	0.46	0.96	0.028
VSIG2	0.022	0.66	0.46	0.94	0.023
ZNF132	0.017	0.65	0.45	0.93	0.018

As is well-known, the TNM staging were correlated with the OS. We also evaluated their relationship of TNM staging with the OS, and found that the TNM staging were actually associated with the overall survival in COAD, READ and CRC, but not for the T staging in READ (Fig 6a-i). Then, we analyzed the relationship of potential prognosisgenes with TNM staging. The results showed that 3genes (HSPA1A, SALL1, and TCF7L1) were associated with T staging, 9 genes (APBB1, CDC42SE2, FBXO17, FES, GNPAT1, OSBPL3, RORC, SPEG, and TCF7L1) were associated with N staging and 5 genes (APBB1, EIF4E3, GNPAT1, HSPA1A, and TCF7L1) were associated with M staging in COAD (Fig 6j). Similarly, AQP1 and HOXA3 were associated with N staging;AQP1 was associated M staging in READ (Fig 6k). Combined analysis, 3 genes (EIF4E3, FAM160A1, HSPA1A) were associated with T staging, 15 genes (ADAMTSL3, ANXA9, APBB1, AQP1, C2CD4A, CLIP3, DNAJC15, EIF4E3, FAM160A1, GNG4, HLX, HSPA1A, LAYN, NR3C2, and SYT1) were associated with N staging and 10 genes (ANXA9, APBB1, AQP1, C2CD4A, EIF4E3, FAM160A1, GNG4, HSPA1A, NR3C2, and SYT1) were associated the M staging in CRC(Fig 6l).

By retrospective examination, we found that those 12 prognosis genes in COAD were negative correlated with 27 DMPs, 2 genes in READ were negative correlated with 8 DMPs, and 15 genes in CRC were negative correlated with 29 DMPs as shown in Fig 7a. Survival analysis also indicated that the patients with low methylation of cg18149207 (hypomethylated gene: RORC) in COAD, cg04408595 (hypomethylated gene: EIF4E3) in CRC and cg19413725 (hypomethylated gene: FAM160A1) in CRC exhibited better OS (Fig 7b, e-f); the patients with high methylation of cg02000808 (hypomethylated gene: HOXA3) and cg21134232 (hypomethylated gene: HOXA3) in READ exhibited better OS (Fig 7c-d).

4. Discussion

Colorectal cancer is the third most common cancer and the second mortality[1]. Gender and aging have shown strong associations with disease incidence consistently by epidemiological studies[1]. Most patients of CRC arise from a polyp which begins with a neoplastic precursor lesion[12, 13]. It is now clear that CRC tumorigenesis is the consequence of the progressive accumulation of genetic and epigenetic alterations, which causes dysregulation of the homeostatic functions and leads to neoplastic transformation[6-8]. Until now, the mainly treatment for CRC is still the surgical which mainly based CT colonography and histology diagnosis[14]. So it is very important to find the tumor based markers for the screening strategies and development of more effective treatments for CRC.

Epigenetic alteration plays a vital role in carcinogenesis and tumor development progression. Abnormal methylation could affect the functions of crucial genes by altering their expression. Increasing studies have demonstrated that DNA methylation is referred to as an early phenomenon, and new efforts are focused on finding tumor biomarkers of early disease detection, prognostication, and treatment[15-17].

In this work, we integrated DNA methylation and gene expression data and screen DNA methylation driven tumor genes, and survival analysis was further to determinate these prognostic genes associated with TNM staging. Survival analysis indicated that there were 24, 16 and 29 genes maybe the prognostic genes for COAD, READ and CRC respectively. There were 11 overlap genes (HSPA1A, TMEM106A, UNC5C, CBFA2T3, TSPAN11, REP15, APBB1, NPTX2, EIF4E3, HHIP, and ZNF132) between COAD and CRC. Similarly, there were 3 overlap genes (EPHX4, C2CD4A, AQP1) between READ and CRC.

Actually, previous studies indicated that HSPA1A[18, 19], UNC5C[20, 21], CBFA2T3[22, 23], NPTX2[24, 25], EPHX4[26], and AQP1[27-29] have been reported to be associated with CRC. This result also suggested the feasibility of our present results. While other genes have not been reported to be associated with COAD/READ/CRC, their associations with other cancers have also been reported, such as APBB1, HHIP, and HOXA3 with lung cancer[30-33]; TCF7L1, HOXB3, and ABCC2 with pancreatic cancer[34-37]; SALL1 with neck cancer[38, 39]; C2CD4A with breast cancer[40]. These results suggest both their relevance and their role as prognostic genes for corresponding cancers. Based on the results of our study and the relationship between those genes and other kind of cancers found in previous studies, it is suggested that those genes mentioned above may be used as prognostic genes for colorectal cancer.

TNM staging were closely associated with the overall survival. We also evaluated the relationship between prognosis genes and TNM staging. And there were 3 overlap genes (HSPA1A, APBB1, and EIF4E3) between COAD and CRC, and 1 overlap genes (AQP1) between READ and CRC. This result suggested that all of these genes not only maybe serve as prognostic genes, but also maybe as therapeutic targets for preventing tumor metastasis.

5. Conclusions

Integrated analysis of the abnormal methylation alteration in CRC indicated that DMGs may be involved in the occurrence of CRC. Moreover, the present study could help clinicians to further known the function of DMGs in CRC. Our study might be the fundamental work for further mechanisms elucidation of CRC and identification of the prognostic genes of CRC. However, it was worth emphasizing that the regulatory network of methylation-mRNA is particularly complex, and the number of case and control data used in the study was not sufficient. We just provide an analysis direction depended on theoretical knowledge and clinical outcomes, more scientific research are needed to confirm our findings.

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 openly available in TCGA at <https://portal.gdc.cancer.gov/>

Competing interests

The authors declare no competing interests.

Funding

This project is financially supported by the Natural Science Foundation of Hunan Province (2019JJ40204), Education Department Foundation of Hunan Province (19A354).

Acknowledgements

Not applicable.

Author contribution

Jing Peng and Yuan-wu Liu are corresponding authors. J.P., Y.L., and X.X., conceived and designed the experiments; X.L., Y.X., Z.C., Z.H., Y.L., and X.X. performed the analysis; Y.Y., T.Z., and N.Z. helped to analyze the data; Y.L., and X.X., wrote the paper.

Reference

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.** *CA Cancer J Clin* 2018, **68**(6):394-424.
2. Jayne DG, Guillou PJ, Thorpe H, Quirke P, Copeland J, Smith AM, Heath RM, Brown JM, Group UMCT: **Randomized trial of laparoscopic-assisted resection of colorectal carcinoma: 3-year results of the UK MRC CLASICC Trial Group.** *J Clin Oncol* 2007, **25**(21):3061-3068.
3. Matsuda T, Yamashita K, Hasegawa H, Oshikiri T, Hosono M, Higashino N, Yamamoto M, Matsuda Y, Kanaji S, Nakamura T *et al*: **Recent updates in the surgical treatment of colorectal cancer.** *Ann Gastroenterol Surg* 2018, **2**(2):129-136.
4. Zamani M, Hosseini SV, Mokarram P: **Epigenetic biomarkers in colorectal cancer: premises and prospects.** *Biomarkers* 2018, **23**(2):105-114.
5. Waddington CH: **The epigenotype. 1942.** *Int J Epidemiol* 2012, **41**(1):10-13.
6. Okugawa Y, Grady WM, Goel A: **Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers.** *Gastroenterology* 2015, **149**(5):1204-1225 e1212.
7. Naghibalhossaini F, Zamani M, Mokarram P, Khalili I, Rasti M, Mostafavi-Pour Z: **Epigenetic and genetic analysis of WNT signaling pathway in sporadic colorectal cancer patients from Iran.** *Mol Biol Rep* 2012, **39**(5):6171-6178.
8. Amirghofran Z, Jalali SA, Hosseini SV, Vasei M, Sabayan B, Ghaderi A: **Evaluation of CD44 and CD44v6 in colorectal carcinoma patients: soluble forms in relation to tumor tissue expression and metastasis.** *J Gastrointest Cancer* 2008, **39**(1-4):73-78.
9. Schneckeburger M, Diederich M: **Epigenetics Offer New Horizons for Colorectal Cancer Prevention.** *Curr Colorectal Cancer Rep* 2012, **8**(1):66-81.
10. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.** *Genome Biol* 2014, **15**(12):550.
11. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, Beck S: **ChAMP: 450k Chip Analysis Methylation Pipeline.** *Bioinformatics* 2014, **30**(3):428-430.
12. Medema JP: **Cancer stem cells: the challenges ahead.** *Nat Cell Biol* 2013, **15**(4):338-344.
13. Nassar D, Blanpain C: **Cancer Stem Cells: Basic Concepts and Therapeutic Implications.** *Annu Rev Pathol* 2016, **11**:47-76.

14. Brandi G, De Lorenzo S, Nannini M, Curti S, Ottone M, Dall'Olio FG, Barbera MA, Pantaleo MA, Biasco G: **Adjuvant chemotherapy for resected colorectal cancer metastases: Literature review and meta-analysis.** *World J Gastroenterol* 2016, **22**(2):519-533.
15. Ishihara H, Yamashita S, Fujii S, Tanabe K, Mukai H, Ushijima T: **DNA methylation marker to estimate the breast cancer cell fraction in DNA samples.** *Med Oncol* 2018, **35**(11):147.
16. Shindo T, Shimizu T, Nojima M, Niinuma T, Maruyama R, Kitajima H, Kai M, Itoh N, Suzuki H, Masumori N: **Evaluation of Urinary DNA Methylation as a Marker for Recurrent Bladder Cancer: A 2-Center Prospective Study.** *Urology* 2018, **113**:71-78.
17. Joosten SC, Deckers IA, Aarts MJ, Hoeben A, van Roermund JG, Smits KM, Melotte V, van Engeland M, Tjan-Heijnen VC: **Prognostic DNA methylation markers for renal cell carcinoma: a systematic review.** *Epigenomics* 2017, **9**(9):1243-1257.
18. Ohkouchi C, Kumamoto K, Saito M, Ishigame T, Suzuki SI, Takenoshita S, Harris CC: **ING2, a tumor associated gene, enhances PAI1 and HSPA1A expression with HDAC1 and mSin3A through the PHD domain and Cterminal.** *Mol Med Rep* 2017, **16**(5):7367-7374.
19. Nadin SB, Cuello-Carrion FD, Sottile ML, Ciocca DR, Vargas-Roig LM: **Effects of hyperthermia on Hsp27 (HSPB1), Hsp72 (HSPA1A) and DNA repair proteins hMLH1 and hMSH2 in human colorectal cancer hMLH1-deficient and hMLH1-proficient cell lines.** *Int J Hyperthermia* 2012, **28**(3):191-201.
20. Wu J, Wang G, He B, Chen X, An Y: **Methylation of the UNC5C gene and its protein expression in colorectal cancer.** *Tumour Biol* 2017, **39**(4):1010428317697564.
21. Guroo SA, Malik AA, Afroze D, Ali S, Pandith AA, Yusuf A: **Significant Pattern of Promoter Hypermethylation of UNC5C Gene in Colorectal Cancer and Its Implication in Late Stage Disease.** *Asian Pac J Cancer Prev* 2018, **19**(5):1185-1188.
22. Parang B, Bradley AM, Mittal MK, Short SP, Thompson JJ, Barrett CW, Naik RD, Bilotta AJ, Washington MK, Revetta FL *et al*: **Myeloid translocation genes differentially regulate colorectal cancer programs.** *Oncogene* 2016, **35**(49):6341-6349.
23. Williams CS, Bradley AM, Chaturvedi R, Singh K, Piazuelo MB, Chen X, McDonough EM, Schwartz DA, Brown CT, Allaman MM *et al*: **MTG16 contributes to colonic epithelial integrity in experimental colitis.** *Gut* 2013, **62**(10):1446-1455.
24. Xu C, Tian G, Jiang C, Xue H, Kuerbanjiang M, Sun L, Gu L, Zhou H, Liu Y, Zhang Z *et al*: **NPTX2 promotes colorectal cancer growth and liver metastasis by the activation of the canonical Wnt/beta-catenin pathway via FZD6.** *Cell Death Dis* 2019, **10**(3):217.
25. Rasmussen SL, Krarup HB, Sunesen KG, Johansen MB, Stender MT, Pedersen IS, Madsen PH, Thorlacius-Ussing O: **Hypermethylated DNA, a circulating biomarker for colorectal cancer detection.** *PLoS One* 2017, **12**(7):e0180809.
26. Roberts DL, O'Dwyer ST, Stern PL, Renehan AG: **Global gene expression in pseudomyxoma peritonei, with parallel development of two immortalized cell lines.** *Oncotarget* 2015, **6**(13):10786-10800.
27. Imaizumi H, Ishibashi K, Takenoshita S, Ishida H: **Aquaporin 1 expression is associated with response to adjuvant chemotherapy in stage II and III colorectal cancer.** *Oncol Lett* 2018, **15**(5):6450-

28. Dorward HS, Du A, Bruhn MA, Wrin J, Pei JV, Evdokiou A, Price TJ, Yool AJ, Hardingham JE: **Pharmacological blockade of aquaporin-1 water channel by AqB013 restricts migration and invasiveness of colon cancer cells and prevents endothelial tube formation in vitro.** *J Exp Clin Cancer Res* 2016, **35**:36.
29. Kang BW, Kim JG, Lee SJ, Chae YS, Jeong JY, Yoon GS, Park SY, Kim HJ, Park JS, Choi GS *et al*: **Expression of aquaporin-1, aquaporin-3, and aquaporin-5 correlates with nodal metastasis in colon cancer.** *Oncology* 2015, **88**(6):369-376.
30. Lee JH, Kim JY, Kim SY, Choi SI, Kim KC, Cho EW, Kim IG: **APBB1 reinforces cancer stem cell and epithelial-to-mesenchymal transition by regulating the IGF1R signaling pathway in non-small-cell lung cancer cells.** *Biochem Biophys Res Commun* 2017, **482**(1):35-42.
31. Xu J, Shang Y, Cai F, Zhang S, Xiao Z, Wang H, Fan Y, Li T, Sheng S, Fu Y *et al*: **Correlation between lung cancer and the HHIP polymorphisms of chronic obstructive pulmonary disease (COPD) in the Chinese Han population.** *Genes Immun* 2019, **20**(4):273-280.
32. Zhao JG, Wang JF, Feng JF, Jin XY, Ye WL: **HHIP overexpression inhibits the proliferation, migration and invasion of non-small cell lung cancer.** *PLoS One* 2019, **14**(11):e0225755.
33. Gan BL, He RQ, Zhang Y, Wei DM, Hu XH, Chen G: **Downregulation of HOXA3 in lung adenocarcinoma and its relevant molecular mechanism analysed by RT-qPCR, TCGA and in silico analysis.** *Int J Oncol* 2018, **53**(4):1557-1579.
34. Yuan YH, Zhou J, Zhang Y, Xu MD, Wu J, Li W, Wu MY, Li DM: **Identification of key genes and pathways downstream of the beta-catenin-TCF7L1 complex in pancreatic cancer cells using bioinformatics analysis.** *Oncol Lett* 2019, **18**(2):1117-1132.
35. Yang D, Yan R, Zhang X, Zhu Z, Wang C, Liang C, Zhang X: **Deregulation of MicroRNA-375 inhibits cancer proliferation migration and chemosensitivity in pancreatic cancer through the association of HOXB3.** *Am J Transl Res* 2016, **8**(3):1551-1559.
36. Gentiluomo M, Puchalt Garcia P, Galeotti AA, Talar-Wojnarowska R, Tjaden C, Tavano F, Strobel O, Kupcinkas J, Neoptolemos J, Hegyi P *et al*: **Genetic variability of the ABCC2 gene and clinical outcomes in pancreatic cancer patients.** *Carcinogenesis* 2019, **40**(4):544-550.
37. Cervenkova L, Vycital O, Bruha J, Rosendorf J, Palek R, Liska V, Daum O, Mohelnikova-Duchonova B, Soucek P: **Protein expression of ABCC2 and SLC22A3 associates with prognosis of pancreatic adenocarcinoma.** *Sci Rep* 2019, **9**(1):19782.
38. Misawa K, Misawa Y, Imai A, Mochizuki D, Endo S, Mima M, Ishikawa R, Kawasaki H, Yamatodani T, Kanazawa T: **Epigenetic modification of SALL1 as a novel biomarker for the prognosis of early stage head and neck cancer.** *J Cancer* 2018, **9**(6):941-949.
39. Imai A, Mochizuki D, Misawa Y, Nakagawa T, Endo S, Mima M, Yamada S, Kawasaki H, Kanazawa T, Misawa K: **SALL2 Is a Novel Prognostic Methylation Marker in Patients with Oral Squamous Carcinomas: Associations with SALL1 and SALL3 Methylation Status.** *DNA Cell Biol* 2019, **38**(7):678-687.

40. Cui Q, Tang J, Zhang D, Kong D, Liao X, Ren J, Gong Y, Xie C, Wu G: **A prognostic eight-gene expression signature for patients with breast cancer receiving adjuvant chemotherapy.** *J Cell Biochem* 2019.

Figures

Figure 1

Volcano plot of differentially methylations and hypomethylatedand hypermethylatedCpGs distribution for COAD, READ, and CRC. a-b, Volcano plot of differentially methylations for COAD (a), READ (b), and CRC (c).d-f, Distribution of DMPs in different regions, including islands, shores, shelves, and opensea. g-i,

Distribution of DMPs across gene regions (TSS1500, TSS200, 5'UTRs, first exons, gene bodies, and 3'UTRs).

Figure 2

Integrative analyses of DNA methylation and gene expression in COAD, READ, and CRC. a-b, Volcano plot of differentially expression genes for COAD (a), READ (b), and CRC (c). d-f, Scatter plot of log2 methylation fold change versus log2 genes expression fold change in COAD (d), READ (e), and CRC (f). Each point represents a methylation-gene pair. g-i, overlap situation of DEGs and DMPs (TSS1500, TSS200 and 5'UTR) target genes.

Figure 3

Summary of top three hypermethylated and top three hypomethylated genes in COAD, READ, and CRC. The abscissa is the degree of methylation, the ordinate is the density of methylated samples, the histogram represents the methylation distribution of the tumor samples, and the curve demonstrates the simulated trend curve of the methylation distribution in the tumor samples. The orange line represents the

distribution of methylation in tumor samples. The green line represents the distribution of methylation in control samples.

Figure 4

Correlation analyses between gene expression andhypermethylated/hypomethylated in COAD, READ, and CRC.**** p ≤0.0001.

a

GO (COAD)	Count	PValue	FDR
transcription, DNA-templated	68	0.000	0.001
regulation of transcription, DNA-templated	56	0.000	0.002
positive regulation of neuron projection development	11	0.000	0.012
integral component of plasma membrane	50	0.000	0.017
transcription factor activity, sequence-specific DNA binding	46	0.000	0.000
sequence-specific DNA binding	26	0.000	0.031

GO (READ)	Count	PValue	FDR
transcription factor activity, sequence-specific DNA binding	33	0.000	0.004

GO (CRC)	Count	PValue	FDR
regulation of transcription, DNA-templated	54	0.000	0.001
transcription, DNA-templated	59	0.000	0.033
axon guidance	13	0.000	0.033
transcription factor activity, sequence-specific DNA binding	40	0.000	0.001

b

c

d

Figure 5

GO analyses of overlap gene between DMPs target genes and DEGs in COAD, READ, and CRC.

Figure 6

Integrative analysis of prognostic genes with TNM staging in COAD, READ, and CRC. a-i, overall survival analysis of TNM staging in COAD (a-c), READ (d-f), and CRC (g-i). Associated analysis of prognostic genes with TNM staging in COAD (j), READ (k), and CRC (l). * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$.

Figure 7

Interaction networks of prognostic genes and methylation and overall survival analysis for COAD(a, d), READ (b, e), and CRC (c, f).