# Genome-Wide Analysis of Cell-Free DNA Methylation Profiling with MeDIP-Seq Identified Potential Biomarkers for Colorectal Cancer

**Xin Zhang** ( zhangxinmd@fudan.edu.cn )
Qingpu Branch of Zhongshan Hospital Affiliated to Fudan University, Department of General Surgery
https://orcid.org/0000-0001-7574-6542

**Tao Li**
Ruijin Hospital, Shanghai Jiao Tong Univerity, School of Medicine, Department of General Surgery

**Qiang Niu**
Shidong Hospital Affiliated to University of Shanghai for Science and Technology, Department of General Surgery

**Chang Jiang Qin**
Huaihe Hospital of Henan University, Department of Gastrointestinal Surgery

**Ming Zhang**
The people's hospital of Wuhai, General surgery

**Guang Ming Wu**
The people's hospital of Wuhai, General surgery

**Hua Zhong Li**
The people's hospital of Wuhai,General Surgery

**Yan Li**
The people's hospital of Wuhai, Digestive internal

**Chen Wang**
The people's hospital of Wuhai, Digestive internal

**Wen Fei Du**
Shanghai Jiao Tong University, Shanghai Center for Systems Biomedicien

**Chen Yang Wang**
Shanghai Jiao Tong University, Shanghai Center for Systems Biomedicine

**Qiang Zhao**
Shanghai Jiao Tong University, Shanghai Center for Systems Biomedicine

**Xiao Dong Zhao**
Shanghai Jiao Tong University, Shanghai Center for Systems Biomedicine

**Xiao Liang Wang**
Qingpu Branch of Zhongshan Hospital Affiliated to Fudan University, Department of General Surgery

**Jian Bin Zhu**

The people's hospital of Wuhai, Digestive internal

**Research Article**

**Keywords:** biomarkers, cfDNA, colorectal cancer, MeDIP-seq

**Posted Date:** June 24th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-643331/v1

# Abstract

**Background:** To verify the feasibility of genome-wide plasma cell-free DNA(cfDNA) methylation profiling for early diagnosis of colorectal cancer.

**Methods:** We performed a genome-wide cfDNA methylation profiling study of colorectal cancer patients by methylated DNA immunoprecipitation coupled with high throughput sequencing (MeDIP-seq).

**Results:** Compared with the control group, 939 differentially methylated regions (DMRs) located in promoter regions were found in colorectal cancer patients, 16 of these DMRs were hypermethylated and the remaining 923 were hypomethylated. In addition, these hypermethylated genes, mainly including PRDM14, RALYL, ELMOD1, and TMEM132E, were validated and confirmed in colorectal cancer by using publicly available DNA methylation data.

**Conclusions:** Our study indicates that MeDIP-seq can be used as an optimal approach for analyzing cfDNA methylomes, and the differentially methylated genes obtained by MeDIP-seq can be used as potential biomarkers for clinical application in patients with colorectal cancer.


# Background

Colorectal cancer is the third leading cause of cancer death worldwide[1]. Early diagnosis and treatment for colorectal cancer are crucial and often offer a good prognosis[2]. At present, the diagnosis of colorectal cancer mainly depends on colonoscopy and detection of serum carcinoembryonic antigen(CEA)[3, 4]. Colonoscopy is invasive and may bring serious complications[5]. The sensitivity of serum CEA detection is often low[6]. In reality, there are still many obstacles to the early diagnosis of colorectal cancer.

Genetic and epigenetic aberrations of tumor cells occur at the initiation of tumorigenesis[7, 8]. Some of the apoptotic or necrotic tumor cells lyse and release their DNA fragments, which make up cfDNA, into the blood[9]. Detection of cfDNA is helpful for early diagnosis and follow-up monitoring of tumors and has the advantages of non-invasive and real-time[10–12].

DNA methylation is an important component of epigenetic modification[13]. The main experimental approaches for profiling genome-wide DNA methylation include whole-genome bisulfite sequencing(WGBS), reduced-representation bisulfite sequencing(RRBS) and MeDIP[14]. Both RRBS and WGBS show substantial DNA degradation after bisulfite treatment, and WGBS is less cost-effective[14]. Recently, some scholars have reported that compared with other detection approaches, cfDNA methylated immunoprecipitation and subsequent high-throughput sequencing(cfMeDIP-seq) are more sensitive, accurate and economical for the early diagnosis of tumors[15]. At present, MeDIP-seq is rarely used in cfDNA methylation profiling of colorectal cancer in China. In this study, we detected the methylation

profiling of cfDNA in colorectal cancer patients by MeDIP-seq and therefore we performed data analysis and validation.

## Methods

# Sample Collection and cfDNA Extraction

All colorectal cancer blood samples (n = 4) were obtained from patients with adenocarcinoma in Shanghai General Hospital and control blood samples (n = 3) were from healthy volunteers(the corresponding information not provided). Informed consent was obtained from individuals, and the study was approved by the Ethics Committee of Shanghai General Hospital.

Blood from colorectal cancer patients and controls (~ 5ml) was collected in tubes containing EDTA as the anticoagulant. Centrifuged blood samples for 10 min at 1900 x g and 4°C temperature setting. Carefully collected the plasma supernatant and centrifuged the plasma samples for 10 min at 16,000 x g in a fixed-angle rotor and 4°C temperature setting. The plasma supernatant was carefully collected and frozen at -80°C.

The plasma cfDNA was extracted using QIAamp Circulating Nucleic Acid Kit (Qiagen, 55114) according to the instructions. Qubit (Invitrogen) was used to analyze the concentration of cfDNA in plasma. The Agilent Bioanalyzer 2100 was used to estimate the distribution of cfDNA size.

# MeDIP-seq Library Construction and Sequencing

The cfDNA was used for the preparation of the MeDIP-seq library, with some modifications[16]. Briefly, we used Illumina's NEBNext Ultra II DNA Library Preparation Kit (NEB, E7645) and ligated ~ 50ng cfDNA to the Illumina adapter according to the manufacturer's instructions. The resulting library was denatured at 95 C for 10 min, immediately incubated on ice for 10 min, and then immunoprecipitated with 5-methylcytosine (5-mC) monoclonal antibody (Epigentek, A-1014). The MeDIP DNA was amplified with Q5 high-fidelity DNA polymerase (NEB, M0491), and the amplified products were purified with AMPure XP beads (Beckman). The amplified libraries were evaluated using Bioanalyzer 2100 (Agilent Technologies), and deep sequencing was performed using Illumina Hiseq 2000.

# Data Processing and Analysis

All qualified reads in colorectal cancer patients and healthy individuals cfDNA MeDIP-seq raw data were mapped to the reference genome (Human hg38) using Bowtie (version 1.0.1)[17]. The MEDIPS analysis package (version 1.24.0) was used for the analysis and comparison of DNA methylation datasets of patients and controls[18].

The 450K methylation array data (Illumina, San Diego, CA, USA) from normal colorectal tissue and colorectal cancer patient samples were obtained from TCGA-COAD Samples Report (https://gdac.broadinstitute.org/runs/stddata__latest/samples_report/COAD.html) and GEO database

(GSE42752, GSE52270, GSE77718). Independent-Sample t-test was performed between normal samples and patient samples using R statistical programming language(3.4.3, http://www.R-project.org) on the data processed with beta (β) values (proportion of the methylated signal over the total signal), and the hypermethylated target genes with *p*-value < 0.05 were selected.

## Results

# Whole-genome MeDIP-seq Analysis of cfDNA

Plasma collected from colorectal cancer patients (n = 4) and healthy controls (n = 3) were used in this study. The clinicopathological information of the patients was shown in Table 1. The cfDNA was extracted from plasma using the QIAamp Circulating Nucleic Acid Kit.

Table 1
Clinicopathological information of colorectal cancer patients

| Sample name | Gender | Age | Stage | Histology |
|---|---|---|---|---|
| J730 | M | 77 | TMN III | Adenocarcinoma |
| J056 | F | 80 | TMN III | Adenocarcinoma |
| J228 | M | 60 | TMN IV | Adenocarcinoma |
| J474 | M | 68 | TMN IV | Adenocarcinoma |
| Note: J730, J056, J228 and J474 represent patients with colorectal cancer. | | | | |

The cfDNA derived from colorectal cancer patients (n = 4) and healthy controls (n = 3) was used for the construction of the MeDIP-seq libraries and followed by next-generation sequencing.

Illumina Hiseq 2000 was used to sequence the MeDIP-seq libraries. On average, 27 million and 52 million raw sequencing reads were obtained from the colorectal cancer patient group and the control group, respectively. And the proportions of reads matched with the reference genome (Human hg38) were 66.2% and 52.9%, respectively. After filtering out the repetitive reads, the patient group had an average of 15 million unique reads and the control group had an average of 5 million unique reads (Table 2).

Table 2
Summary statistics of MeDIP-seq data

| Sample | Number of total reads | Number of mapped reads | Total mapped read rate | Number of unique reads | Unique reads rate |
|---|---|---|---|---|---|
| J730 | 42,362,427 | 32,119,271 | 75.8% | 26,454,305 | 82.4% |
| J056 | 29,652,736 | 21,442,064 | 72.3% | 17,378,878 | 81.1% |
| J228 | 18,250,231 | 11,425,158 | 62.6% | 9,003,634 | 78.8% |
| J474 | 19,015,432 | 10,281,561 | 54.1% | 8,097,962 | 78.8% |
| C1 | 46,505,740 | 16,686,432 | 35.9% | 2,089,072 | 12.5% |
| C2 | 18,918,360 | 11,095,194 | 58.7% | 5,085,484 | 45.8% |
| C3 | 91,305,808 | 58,482,718 | 64.1% | 8,453,424 | 14.5% |

Note: J730, J056, J228 and J474 represent patients with colorectal cancer; C1, C2 and C3 represent healthy controls.

# Distinctive cfDNA Methylation Patterns between Colorectal Cancer Patients and Healthy Controls

To determine the overall cfDNA methylation in patients and healthy controls, we performed heuristic cluster analysis and unsupervised cluster analysis on cfDNA MeDIP data from colorectal cancer samples and normal samples, respectively. Through heuristic cluster analysis, we found out that the methylation patterns were distinctive between the patient group and the control group(Fig. 1A). And then genome-wide unsupervised cluster analysis also confirmed distinct methylation patterns between the two groups(Fig. 1B).

# Differentially Methylated Regions(DMRs) in Colorectal Cancer Patients

With the help of the MeDIPS analysis package, a total of 8398 DMRs were obtained from the genome-wide distribution of patients ($p$-value < 0.05). Among these DMRs, 1875(22.3%) were hypermethylated and 6523(77.7%) were hypomethylated(Supplementary Table 1). We examined the genomic distributions of hypomethylated and hypermethylated DMRs and found that the proportion of hypermethylated DMRs was higher in the intergenic and intron regions(Fig. 2A). The distribution of DMRs mapped to the whole genome on different chromosomes was shown in Fig. 2B. These 8,398 DMRS exhibiting distinct patterns between colorectal cancer patients and normal controls were shown in Fig. 2C.

It is well known that hypermethylation in the promoter region of tumor suppressor genes is positively correlated with the occurrence and development of tumors[19, 20]. Therefore, we further analyzed DMRs located in the promoter region. We found that 939 DMRs were located in promoter regions(Fig. 2D and

Supplementary Table 2), including 16 hypermethylated regions and 923 hypomethylated regions. Furthermore, these 939 DMRs in the promoter regions also exhibited distinct patterns between the patients and the controls.

# Validation of Differentially Methylated Genes by using Publicly Available DNA Methylation Data

As mentioned above, we found that 16 of the DMRs located in the promoter region were hypermethylated, and then we wanted to find out whether the methylation levels of these corresponding genes were helpful in distinguishing colorectal cancer patients from healthy individuals.

After annotating 16 DMRs with hypermethylated promoter regions, 13 genes were obtained and their corresponding promoter region microarray probes were screened. During the screening process, probes located in the sex chromosome and the 3'UTR regions and the gene body regions were excluded, as were the SNP-related probes. Only the probes located in the UCSC_CpG_Islands regions were retained, so finally, a total of 12 probes were obtained(Supplementary Table 3). And the corresponding genes of the 12 probes mentioned above are PRDM14, RALYL, ELMOD1 and TMEM132E.

The 450K methylation array data were obtained from TCGA and GEO datasets respectively, including both colorectal cancer patient samples (n = 295) and normal colorectal tissue samples (n = 193). Based on the aforesaid 12 probes, the predictive model of logistic regression algorithm was established, and the 488 original data were divided into the training dataset and validation dataset respectively according to the ratio of 4:1. The predictive ability of the model in the two datasets was shown in Fig. 3. According to the receiver operating characteristic (ROC) curves shown in the figure, the area under the curve (AUC) of the training dataset and the validation dataset were 0.928 and 0.915, respectively. Figure 3A and 3B showed the confusion matrix of the training dataset and the validation dataset respectively. This suggested high validity in the diagnosis of colorectal cancer based on the methylation levels of the 12 probes described above.

We then extracted the 12 probes for unsupervised cluster analysis based on the 488 cases of 450K methylation array dataset, and the results showed that the methylation data of the aforesaid 12 probes were distinct between the tumor and normal tissue in general (Fig. 4A). We also compared the methylation levels of the aforesaid 12 probes between the normal colorectal tissue and the colorectal cancer patient samples in the dataset, and we found the methylation levels of the aforesaid 12 probes were significantly different ($p$-value < 0.05). Compared with normal colorectal tissue, the methylation level of the 12 probes in the tumor tissue was hypermethylated (Fig. 4B). These results suggest that detecting the methylation levels of these 12 probes and their corresponding genes is helpful for the diagnosis of colorectal cancer.

# Discussion

Many reports have pointed out that liquid biopsy studies, including cfDNA tests, and their clinical application may be helpful for tumor diagnosis, drug screening, efficacy evaluation, prognosis prediction, and tumor surveillance[9, 21–23]. cfDNA also includes DNA fragments released into the blood after lysis by apoptotic or necrotic tumor cells, which is commonly referred to as circulating tumor DNA (ctDNA)[9, 20]. The ctDNA has methylation patterns similar to those found in tumor cells[24]. In recent years, there have been a few reports on the genome-wide detection of cfDNA methylation profiling by MeDIP-seq to screen potential tumor biomarkers. Xu et al.[25] identified hypermethylated DMRs in the promoter region, which could be used as the early diagnostic markers for lung cancer. Li et al. [26] identified hypermethylated DMRs which were located in promoter regions and completely overlapped with CGIs could be used for the non-invasive diagnosis of pancreatic cancer. So far, there have been few reports on the detection of cfDNA methylation profiling by MeDIP-seq in colorectal cancer in China.

Here in this study, we found 8398 DMRs in cfDNA collected from patients with colorectal cancer at the genome-wide level, and among these DMRs, 1875(22.3%) were hypermethylated and 6523(77.7%) were hypomethylated. And when we focused on DMRS located in the promoter region, 16(1.7%) were hypermethylated and 923(98.3%) were hypomethylated. This suggests that demethylation is widespread in cancer genomes at the genome-wide level[27], with a higher proportion of hypomethylation in promoter regions. Studies have shown that DNA demethylation plays an important role in activating specific gene expression and the initiation of reprogramming[28].

Subsequently, to evaluate the diagnostic value of hypermethylated genes in colorectal cancer, their methylation data were obtained from the publicly available DNA methylation datasets due to the lack of cfDNA methylation data in the public datasets. A predictive model was constructed to confirm its high validity. Methylation levels of the 12 probes in the public database were significantly different between colorectal cancer tissue and normal tissue. The corresponding genes of the 12 probes mentioned above are PRDM14, RALYL, ELMOD1 and TMEM132E. PRDM14 has been reported to be hypermethylated in lung cancer and has high accuracy in the diagnosis of lung cancer[29, 30]. Studies also have shown that the PRDM14 has several hypermethylated CpG sites in African-American colorectal cancer patients by using RRBS, which is consistent with our findings, although different from our experimental approach and the race we've studied[31]. RALYL has been reported to be down-regulated in clear cell renal cell carcinoma, and its reduced expression is associated with poor prognosis[32], which means it could serve as a tumor suppressor gene. These findings indicate that the methylated genes which were identified from cfDNA derived from colorectal cancer patients' plasma may have potential clinical application value.

# Conclusions

In summary, our study indicates that MeDIP-seq can be used as an optimal approach for analyzing cfDNA methylomes, and the differentially methylated genes obtained by MeDIP-seq can be used as potential biomarkers for clinical application in patients with colorectal cancer.

# Declarations

### Ethics approval and consent to participate

Informed consent was obtained from individuals, and the study was approved by the Ethics Committee of Shanghai General Hospital.

### Consent for publication

Not applicable.

### Availability of data and materials

The datasets used during the present study are available from the corresponding author upon reasonable request.

### Competing interests

The authors declare no conflict of interest.


### Funding

This work was supported by National Natural Science Foundation of China (81670514 and 81702337), and Inner Mongolia Science & Technology Plan.


### Authors' contributions

XD.Z and XL.W conceived the study and were in charge of the overall direction and planning. X.Z, T.L, and Q.N wrote and proofread the manuscript with input from all authors. X.Z, WF.D, CY.W, and HZ.L collected the samples. M.Z and GM.W analyzed the data. Q.Z and CJ.Q performed the computational framework. Y.L and C.W provided technical support. XL.W and JB.Z provided the funding support.


### Acknowledgments

Not applicable.

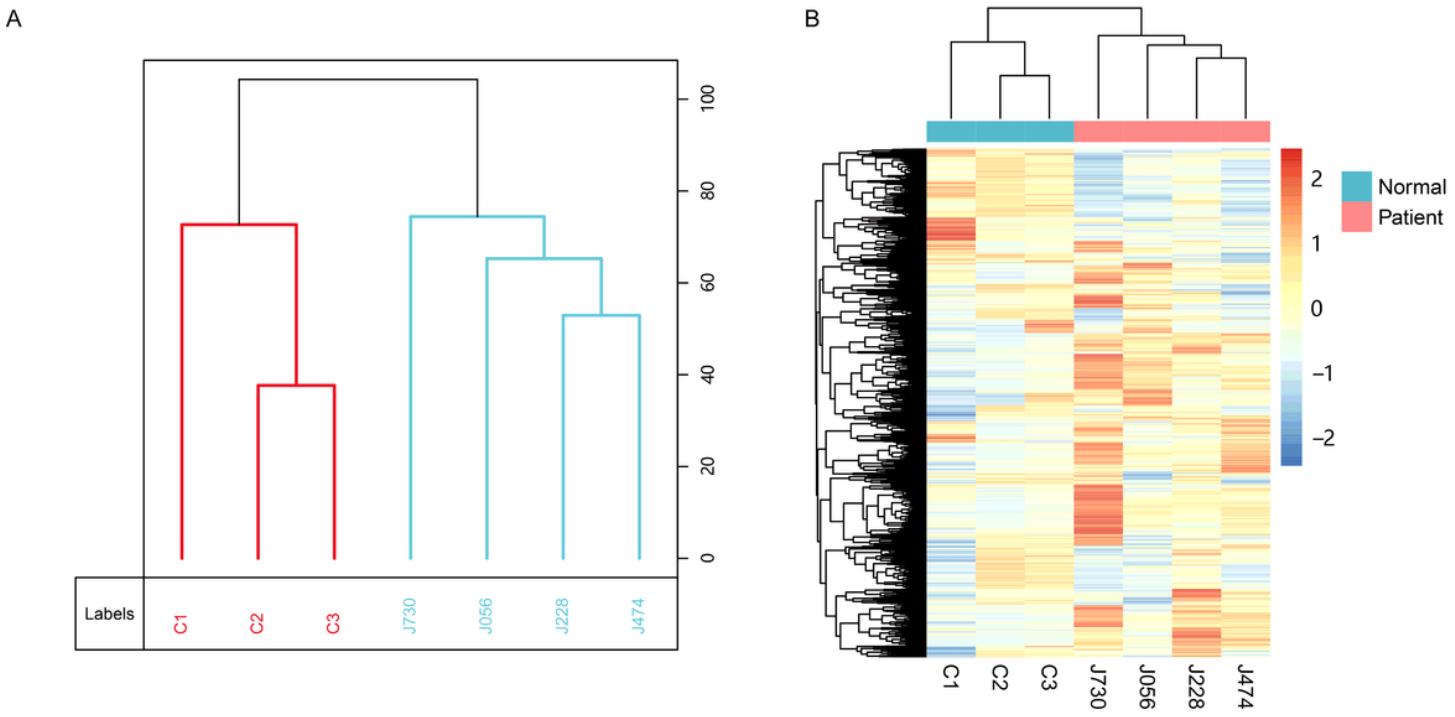

# References


1. González-González M, Gutiérrez ML, Sayagués JM, et al. Genomic profiling of sporadic liver metastatic colorectal cancer [J]. Sem Cancer Biol. 2021;71:98–108.
2. Insua YV, Cámara J, Vázquez EB, et al. Predicting Outcome and Therapy Response in mCRC Patients Using an Indirect Method for CTCs Detection by a Multigene Expression Panel: A Multicentric Prospective Validation Study [J]. Int J Mol Sci. 2017;18(6):1265.

3. Hosoe N, Limpias Kamiya KJ, Hayashi Y, et al. Current Status of Colon Capsule Endoscopy [J]. Digestive endoscopy: official journal of the Japan Gastroenterological Endoscopy Society, 2020 Jun 16.
4. Xiang W, Lv Q, Shi H, et al. Aptamer-based biosensor for detecting carcinoembryonic antigen [J]. Talanta. 2020;214:120716.
5. Dixon P, Kowdley GC, Cunningham SC. The role of surgery in the treatment of endoscopic complications [J]. Best practice & research Clinical gastroenterology, 2016, 30(5): pp. 841–51.
6. Baassiri A, Nassar F, Mukherji D, et al. Exosomal Non Coding RNA in LIQUID Biopsies as a Promising Biomarker for Colorectal Cancer [J]. Int J Mol Sci. 2020;21(4):1398.
7. Miranda Furtado CL, Dos Santos Luciano MC, Silva Santos RD, et al. Epidrugs: targeting epigenetic marks in cancer treatment [J] Epigenetics. 2019;14(12):1164–76.
8. Liu J, Adhav R, Xu X. Current Progresses of Single Cell DNA Sequencing in Breast Cancer Research [J]. Int J Biol Sci. 2017;13(8):949–60.
9. Gorgannezhad L, Umer M, Islam MN, et al. Circulating tumor DNA and liquid biopsy: opportunities, challenges, and recent advances in detection technologies [J]. Lab Chip. 2018;18(8):1174–96.
10. Kujala J, Hartikainen JM, Tengström M, et al. High mutation burden of circulating cell-free DNA in early-stage breast cancer patients is associated with a poor relapse-free survival [J]. Cancer medicine. 2020;9(16):5922–31.
11. Georgiadis A, Durham JN, Keefer LA, et al. Noninvasive Detection of Microsatellite Instability and High Tumor Mutation Burden in Cancer Patients Treated with PD-1 Blockade [J]. Clinical cancer research: an official journal of the American Association for Cancer Research. 2019;25(23):7024–34.
12. Rostami A, Lambie M, Yu CW, et al. Senescence, Necrosis, and Apoptosis Govern Circulating Cell-free DNA Release Kinetics [J]. Cell reports. 2020;31(13):107830.
13. Hou YC, Lu CL, Yuan TH, et al. The Epigenetic Landscape of Vascular Calcification: An Integrative Perspective [J]. Int J Mol Sci. 2020;21(3):980.
14. Yong WS, Hsu FM, Chen PY. Profiling genome-wide DNA methylation [J]. 9: Epigenetics & chromatin; 2016. p. 26.
15. Shen SY, Singhania R, Fehringer G, et al. Sensitive tumour detection and classification using plasma cell-free DNA methylomes [J]. Nature. 2018;563(7732):579–83.
16. Zhang XL, Wu J, Wang J, et al. Integrative epigenomic analysis reveals unique epigenetic signatures involved in unipotency of mouse female germline stem cells [J]. Genome biology. 2016;17(1):162.
17. Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome [J]. Genome biology. 2009;10(3):R25.
18. Lienhard M, Grimm C, Morkel M, et al. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments [J]. Bioinformatics, 2014, 30(2): 284–6.
19. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps [J]. Nat Rev Genet. 2007;8(4):286–98.

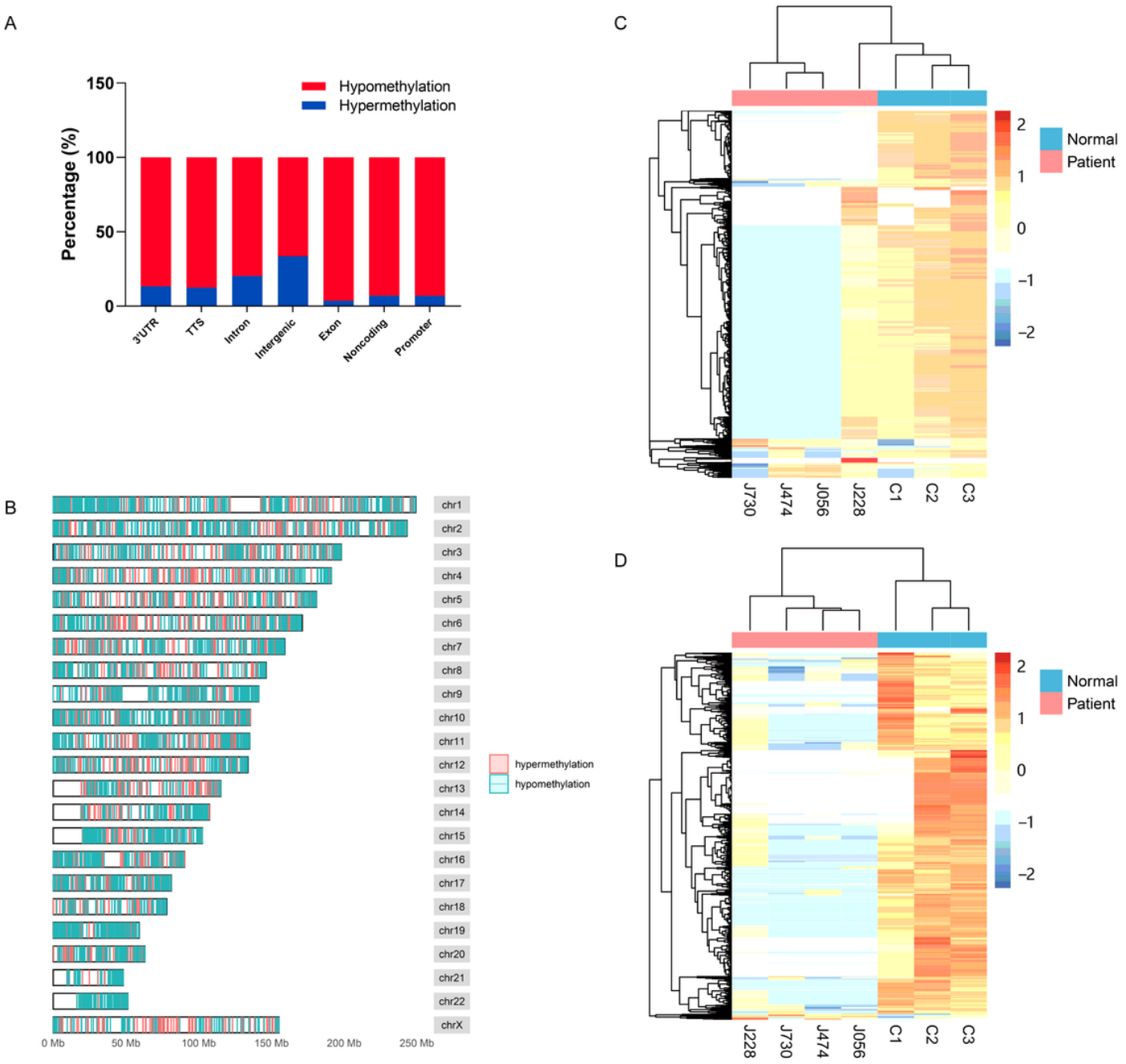
20. Xu RH, Wei W, Krawczyk M, et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma [J]. Nat Mater. 2017;16(11):1155–61.
21. Parikh AR, Leshchiner I, Elagina L, et al. Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers [J]. Nature medicine. 2019;25(9):1415–21.
22. Crowley E, Di Nicolantonio F, Loupakis F, et al. Liquid biopsy: monitoring cancer-genetics in the blood [J]. Nature reviews Clinical oncology. 2013;10(8):472–84.
23. Guo Q, Wang J, Xiao J, et al. Heterogeneous mutation pattern in tumor tissue and circulating tumor DNA warrants parallel NGS panel testing [J]. Mol Cancer. 2018;17(1):131.
24. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients [J]. Nat Rev Cancer. 2011;11(6):426–37.
25. Xu W, Lu J, Zhao Q, et al. Genome-Wide Plasma Cell-Free DNA Methylation Profiling Identifies Potential Biomarkers for Lung Cancer [J]. Disease markers, 2019, 2019:4108474.
26. Li S, Wang L, Zhao Q, et al. Genome-Wide Analysis of Cell-Free DNA Methylation Profiling for the Early Diagnosis of Pancreatic Cancer [J]. Frontiers in genetics. 2020;11:596078.
27. López-Moyado IF, Tsagaratou A, Yuita H, et al. Paradoxical association of TET loss of function with genome-wide DNA hypomethylation [J]. Proc Natl Acad Sci USA. 2019;116(34):16933–42.
28. Simonsson S, Gurdon J. DNA demethylation is necessary for the epigenetic reprogramming of somatic cell nuclei [J]. Nat Cell Biol. 2004;6(10):984–90.
29. Hubers AJ, Brinkman P, Boksem RJ, et al. Combined sputum hypermethylation and eNose analysis for lung cancer diagnosis [J]. J Clin Pathol. 2014;67(8):707–11.
30. Hubers AJ, Heideman DA, Burgers SA, et al. DNA hypermethylation analysis in sputum for the diagnosis of lung cancer: training validation set approach [J]. British journal of cancer. 2015;112(6):1105–13.
31. Ashktorab H, Shakoori A, Zarnogi S, et al. Reduced Representation Bisulfite Sequencing Determination of Distinctive DNA Hypermethylated Genes in the Progression to Colon Cancer in African Americans [J]. Gastroenterology research and practice, 2016; 2016:2102674.
32. Cui ZW, Xia Y, Ye YW, et al. RALY RNA binding protein-like reduced expression is associated with poor prognosis in clear cell renal cell carcinoma [J]. Asian Pacific journal of cancer prevention: APJCP. 2012;13(7):3403–8.

# Figures




## Figure 1

The cfDNA methylation patterns derived from MeDIP-seq datasets between colorectal cancer patients and controls. (A) Heuristic cluster analysis of methylation profiling between patients and controls. (B) Unsupervised cluster analysis of the genome-wide methylation profiling in patients and controls.




## Figure 2

Differentially methylated regions in patients and controls. (A) The genomic distributions of hypomethylated and hypermethylated DMRs in introns, intergenomic, exons, noncoding, promoters and other regions. (B) The distribution of DMRs mapped to the whole genome on different chromosomes in patients. (C) Heat map of total 8398 DMRs, including 1875 hypermethylated and 6523 hypomethylated. (D) Heat map of DMRs located in promoter regions in patients and controls, including 16 hypermethylated and 923 hypomethylated.

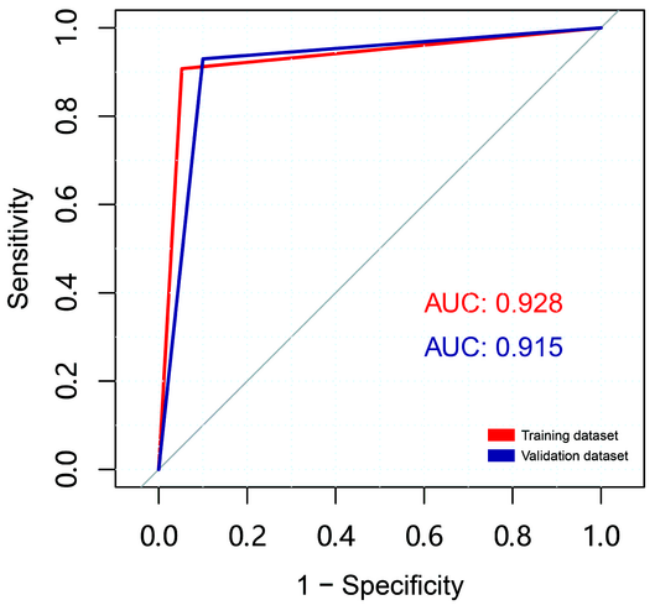
A

| Training dataset | Actual Normal | Actual COAD |
|---|---|---|
| Predict Normal | 145 | 22 |
| Predict COAD | 8 | 216 |
| Totals | 391 | |
| Sensitivity | | 90.8% |
| Specificity | 94.8% | |

B

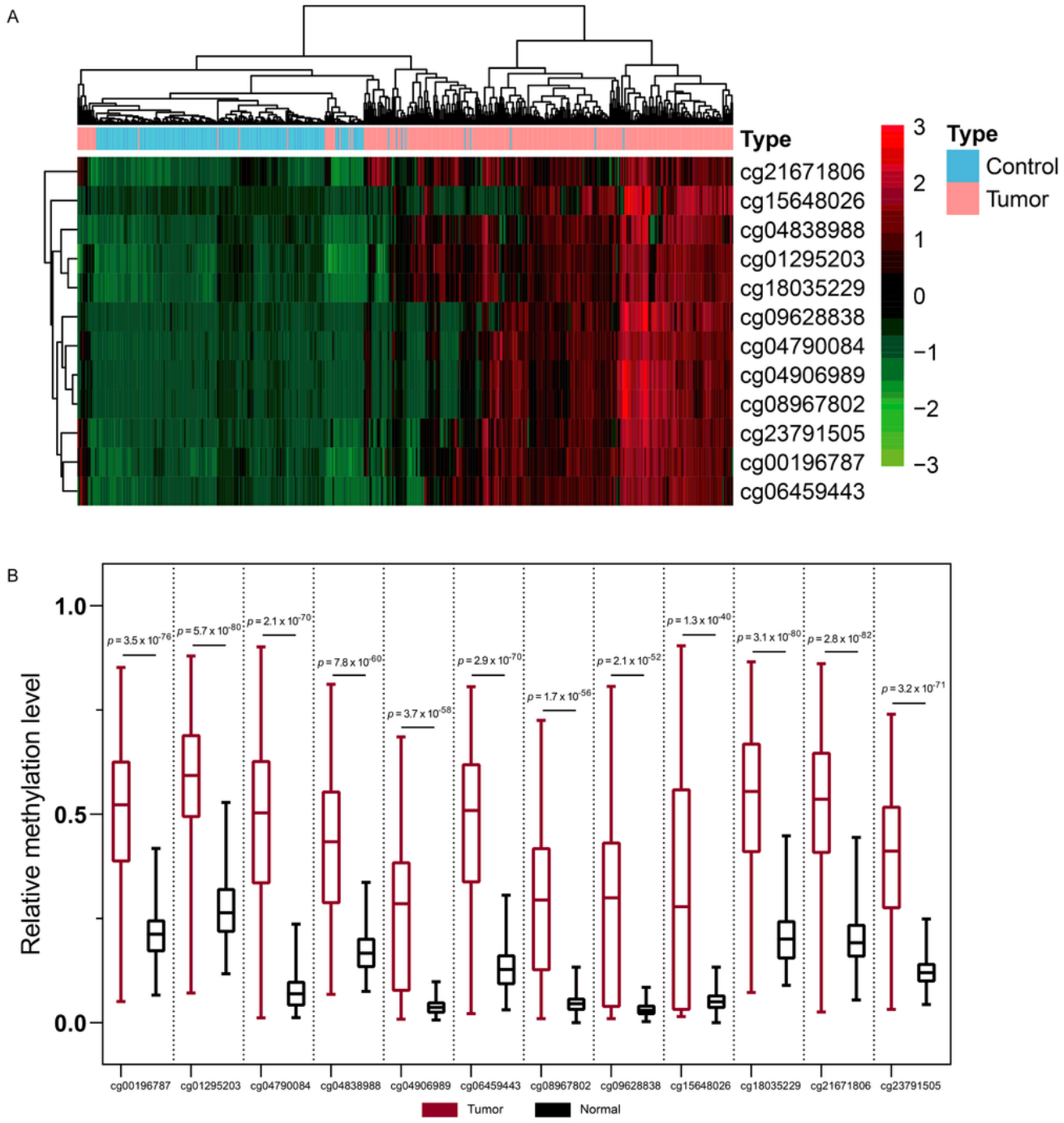
| Validation dataset | Actual Normal | Actual COAD |
|---|---|---|
| Predict Normal | 36 | 4 |
| Predict COAD | 4 | 53 |
| Totals | 97 | |
| Sensitivity | | 93.0% |
| Specificity | 90.0% | |

C




## Figure 3

Diagnostic predictive models and receiver operating characteristic (ROC) curves for colorectal cancer. (A and B) Confusion matrix built from the diagnostic predictive models in training (A) and validation (B) dataset. COAD, colon adenocarcinoma. (C) ROC curves and the associated area under the curve (AUC) of the training and validation dataset.




## Figure 4

Validation of hypermethylated genes by using publicly available DNA methylation data. (A) Unsupervised cluster analysis of these 12 probes extracted from the 488 cases of 450K methylation array dataset. (B) The comparison of methylation level between tumor and normal tissue of the 12 selected probes. All p-values<0.05.

# Supplementary Files

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

- SupplementaryTable1.xlsx
- SupplementaryTable2.xlsx
- SupplementaryTable3.xlsx