**Comprehensive methylome analysis identifies a panel of zinc finger family genes to be diagnostic biomarkers for KRAS mutated colorectal cancer**

Weilin Pu 1,2,#, Fei Qian3,#, Jing Liu 2, Keke Shao4, Feng Xiao5, Qin Jin6, Qingmei Liu7, Shuai Jiang2, Rui Zhang8, Jun Zhang9, Shicheng Guo10, Jianfeng Zhang11, Jiucun Wang2, Yanyun Ma2, Shaoqing Ju1, Weifeng Ding1,\*

1Department of Laboratory Medicine, Affiliated Hospital of Nantong University, Nantong, Jiangsu Province, China

2State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China;

3 Department of Gastrointestinal Surgery, Affiliated Hospital of Nantong University, Nantong, Jiangsu Province, China

4 Department of Laboratory Medicine, the First people’s Hospital of Yancheng City, Yancheng, Jiangsu Province, China

5 Department of Pathology, the Third people’s Hospital of Nantong City, Nantong, Jiangsu Province, China

6 Department of Pathology, Affiliated Hospital of Nantong University, Nantong, Jiangsu Province, China

7 Department of dermatology, Huashan hospital of fudan university，Shanghai, China

8

9 Department of Gastroenterology, Huashan hospital of fudan university，Shanghai, China

10 Center for Precision Medicine Research, Marshfield Clinic Research Institute, Marshfield, WI, USA

11 Department of Gastroenterology, Affiliated Hospital of Nantong University, Nantong, China

#, Contributed equally to the manuscript

\*, Corresponding Author

**Abstract**

**Background**: Colorectal cancer (CRC) is one of the leading causes of cancer deaths worldwide, calling for the urgent need for early diagnosis. Previous studies have revealed the methylation status of zinc finger genes could be ideal for cancer diagnosis. However, no comprehensive evaluation of the methylation alterations of zinc finger genes in CRC has been conducted.

**Methods**: We proposed a two-stage design to build a zinc finger genes (ZEG) model for CRC diagnosis. In the discovery stage, we collected 1,426 genome-wide DNA methylation data including 1,104 CRC, 54 adenomas and 268 normal colon samples to discover the best panel among all zinc finger genes. The optimized panel was validated in two replication cohorts including 218 CRC patients with KRAS mutation status. We compared the prediction effects between ZEG model and SEPT9 model. Finally, we evaluated the prediction performance with both methylation and KRAS mutation.

**Results**: We identified a promising ZEG CRC prediction model including *ESR1*, *ZNF132*, *ZNF229*, *ZNF542* and *ZNF677*. We demonstrated ZEG model (AUC=0.87-0.93) have better performance than SEPT9 model (AUC=0.91). Furthermore, we demonstrated ZEG model have better prediction ability in the KRAS+ patients (AUC>0.98) than that in the KRAS- patients (AUC: 0.74-0.86).

**Conclusions**: Methylation profiles of the five genes from zinc finger family (ZEG model) could be ideal biomarkers for CRC early diagnosis, especially for the CRC patients with KRAS mutations.

**Keywords**: Colorectal cancer, DNA methylation, Zinc Finger Family, KRAS, Diagnosis

**Introduction**

Colorectal cancer (CRC) is the third most prevalent cancer and accounts for 10 % of the cancer deaths all over the world [ref1]. Previous studies have found that the accumulations of both genetic and epigenetic alterations lead to the carcinogenesis of CRC. The mutation of *APC*, *KRAS, TP53* and *PIK3CA* genes are the key events in CRC and the distribution of these mutations differed in different populations [ref2]. The five-year survival rate of CRC is above 70 % in its early stages (TNM I-III stages) while decreased significantly to about 14 % when detected at the later stage (IV) [www.cancer.org], suggesting the importance of early detection methods. Recently, owing to the widely application of screening modalities including colonoscopy and image-based detection, the mortality of CRC has been decreased significantly. However, the abdominal pain and discomfort as well as many contraindications such as hypertension, cardiopathy of these screening methods makes it not widely accepted across populations, calling for the need to develop the more practical methods for CRC early diagnosis.

[ref1] Siegel, R. L., et al. (2019). "Cancer statistics, 2019." CA Cancer J Clin **69**(1): 7-34.

[ref2] Cunningham, D., et al. (2010). "Colorectal cancer." Lancet **375**(9719): 1030-1047

DNA methylation is a crucial epigenetic modifications in the human genomes and plays key roles in embryonic development, transcription[1] regulation[1, 2] and genomic imprinting[3]. DNA methylation shown different patterns in different tissues and disease status which have been applied to develop for tissue-of-origin[4] and disease prediction[5], especially cancer[6] and immune diseases[7-9][6]. Until now, serials of candidate DNA methylation-based biomarkers have been found in CRC, especially *SEPT9*[10-15]. However, the performance of *SEPT9* was found not as good as that of the stool DNA test[16-18]. Therefore, the identification of better DNA methylation-based biomarkers with high accuracy will deeply benefit liquid biopsy of CRC.

Zinc finger proteins (ZFP) are prominent component of transcriptional factors in eukaryotes. It has been reported that ZFP could be divided into eight different classes, including Cys2His2 (C2H2) like, Gag knuckle, Treble clef, Zinc ribbon, Zn2/Cys6, TAZ2 domain like, Zinc binding loops and Metallothionein [19, 20]. Among them, the C2H2-type zinc finger motifs is the largest family of all zinc finger motif classes. Currently, a growing body of literature suggest that zinc finger proteins could contribute or suppress tumor progression via transcriptional regulation. Meanwhile, the DNA methylation alterations of multiple ZFPs have been recognized as promising biomarkers for cancer diagnosis, prognosis and drug response, indicating its vital role in cancers. However, few studies have assessed the DNA methylation profiles of these ZFGs in CRC comprehensively and systemically, which may be of importance for identifying key zinc finger proteins in the tumorigenesis and diagnosis of CRC.

In this study, we for the first time exhaustedly searched and combined public high-throughput DNA methylation microarray datasets, including 1,104 CRC samples, and 54 adenomas as well as 268 control/adjacent normal samples, to systemically explore the promising biomarkers belonging to the zinc family for CRC diagnosis. Through the detailed filtering procedures, we finally identified seven candidate genes and five of these were finally successfully validated in 104 CRC patients of Han Chinese using targeted bisulfite sequencing method. It is found that all of these five candidate genes were significantly hyper-methylated in CRC tumors (AUC ranges from 0.85 to 0.93), especially in the CRC tumors with KRAS mutation which the AUC of these five candidate genes could reach at least 0.98. To further confirm the findings, we recruited another independent cohort including 114 CRC patients of Han Chinese and yielded consistent results. Therefore, we proposed that the zinc family genes, including *ESR1*, *ZNF132*, *ZNF229*, *ZNF542* and *ZNF677* could be robust and reliable biomarkers for CRC diagnosis, especially for the KRAS mutated patients.

**Results**

**Comprehensive integration of public DNA methylation microarray datasets of CRC**

To identify the robust DNA methylation-based biomarkers, we searched the TCGA and GEO datasets concerning about the DNA methylation status of the colorectal cancer. After careful search, we identified 11 datasets, including 1,104 CRC tumors, 268 adjacent normal samples as well as 54 adenomas for further analysis (Table S1). Meanwhile, we have obtained the list of the genes belonging to the zinc finger family (Table S2). Based on the feature selection procedures described in “Methods” section (Figure 1), we finally identified five candidate genes including *ESR1*, *ZNF123*, *ZNF229*, *ZNF542* and *ZNF677* (Figure 1A-E). All of these candidate genes showed significantly hyper-methylated in both CRC and adenoma tissues compared to the adjacent normal tissues (Figure S1). Meanwhile, the expression levels of these genes were also significantly down-regulated in CRC tumors than that in adjacent normal tissues in TCGA dataset, which was in accordance with the hyper-methylated status in tumor tissues (Figure 1F). In addition, they were also showed hypo-methylated status in the PBMC (peripheral blood mononuclear cells), PBL (peripheral blood leucocytes) and WB (whole blood) samples of healthy controls (Table S3). Therefore, we proposed that these five candidate genes might be ideal biomarkers for CRC. To quantitatively characterize the abilities of these biomarkers in the combined discovery dataset, we constructed a logistic regression model using all of these five candidates and yielded a robust discrimination between CRC tumors and adjacent normal tissues (Sensitivity = 0.82 - 0.90, Specificity = 0.88 - 0.97, AUC = 0.93 - 0.97). To further validate the robustness and ability of these CRC diagnostic biomarkers, we conducted the validation study in two replication cohorts, consisting of 218 pairs of CRC and adjacent normal tissue samples from patients of Han Chinese population.

**The validation and evaluation of a panel of zinc finger family genes as the diagnostic biomarkers in replication cohort 1**

The characteristics of the CRC samples in replication cohort 1 were shown in Table 1. To characterize the methylation profiles of the candidates systemically, we conducted the targeted bisulfite sequencing approach and detected the methylation profiles of *SEPT9* as the positive control. It is found that the bisulfite conversion rate (C to T) was high (> 99%) in both CRC and control samples and no significant difference of the reads mapping rate was found between CRC and control samples (Figure S2). After quality control, 187 samples, including 98 CRC samples and 89 adjacent normal tissues were retained for further analysis. The PCA (principal component analysis) revealed a significant distinction between CRC and adjacent normal tissues (Figure S3). The differential methylation analysis was also conducted for all of these five candidates (Figure 3 and Figure S4). It is found that all of these candidates were significantly hyper-methylated in CRC samples of replication cohort 1 with the AUC ranged from 0.85 to 0.93 (Table 2). Specifically, we found that *ESR1* (AUC = 0.93) and *ZNF132* (AUC = 0.91) both achieved the diagnostic ability equaled or better than that of *SEPT9* (AUC = 0.91), indicating that these two zinc finger genes might be of great potential to be utilized for liquid biopsy.

In addition to the overall differential methylation analysis, we also evaluated the effects of age, gender, tumor stage, colon or rectum as well as the mutation status (*KRAS* gene) of the CRC samples. It is found that no significant differences of the diagnostic ability were found between CRC patients with young/old, male/female, early/late, colon/rectum subgroups (Table S4). However, we found that the diagnostic ability of CRC patients carrying KRAS mutation (KRAS+) was significantly superior to that of the KRAS wild-type (KRAS-) samples. In the KRAS+ group, we found that that sensitivity of each gene ranges from 0.86 to 0.98, while the specificity ranges from 0.89 to 1.00 and the AUC ranges from 0.97 to 1.00. However, in the KRAS- group, the sensitivity of each gene ranges from 0.52 to 0.82, while the specificity ranges from 0.80 to 0.98 and the AUC ranges from 0.74 to 0.86 (Table 3). Meanwhile, we conducted the hierarchical clustering analysis and found that the samples of the CRC tumors misclassified into the control group were all KRAS- type, indicating the significant differences between KRAS+ and KRAS- CRC patients (Figure 4).

Taken together, we found a promising ZEG CRC prediction model including *ESR1*, *ZNF132*, *ZNF229*, *ZNF542* and *ZNF677* in our corhort 1.

**ZEG model with more superior prediction for KRAS mutated colorectal cancer in replication cohort 2**

To further validate the efficacy and accuracy of the candidates in the diagnosis of CRC patients with/without KRAS mutation, we measured the methylation profiles in another independent cohort consisting of 114 pairs of CRC patients of Han Chinese population (Table 1). We observed excellent the bisulfite conversion rate (> 99%) and the reads mapping rate (>75%) (Figure S5). We validated the differentially methylated profiles of the candidates in replication cohort 2 (Figure S6). Moreover, the methylation profiles of the KRAS+ and KRAS- subgroups were also measured and significantly differences between these subgroups were found (Figure S7). In the KRAS+ subgroup, we found that the sensitivity of each gene ranges from 0.90 to 1.00, while the specificity ranges from 0.91 to 0.98 and the AUC ranges from 0.92 to 1.00. In contrast, the sensitivity of each gene ranges from 0.58 to 0.85, while the specificity ranges from 0.78 to 0.96 and the AUC ranges from 0.71 to 0.88 in the KRAS- subgroup. Furthermore, we also identified that the CRC samples misclassified to the adjacent normal group were belonging to the KRAS- subgroup, suggesting that the KRAS+ CRC samples were more epigenetically homogeneous than that of the KRAS- CRC samples.

**Overall diagnostic abilities of ZEG model superior than that of *SEPT9* in the combined cohorts**

To give a more reliable and robust estimation of the efficacy of the zinc finger genes in CRC diagnosis, we combined the datasets from replication 1 and replication 2 and tested the diagnostic performances of these biomarkers. As shown in Table 5, we found that *ZNF132* reached the highest diagnostic ability (Sensitivity = 0.83, Specificity = 0.97, AUC = 0.93) than other candidate biomarkers including *SEPT9* (SEN = 0.83, SPE = 0.87, AUC = 0.91) in the combined samples. Meanwhile, ESR1 also achieved comparable diagnostic ability (SEN = 0.78, SPE = 0.97, AUC = 0.91) when compared to the *SEPT9*. It is suggested that these two biomarkers could be further applied and tested in combination with *SEPT9* to achieve better diagnostic ability in the non-invasive liquid biopsy studies for CRC. Moreover, in the KRAS+ subgroups, all candidate biomarkers achieved superior performances (AUC >= 0.95) than that in the KRAS- and the total samples. Especially, we found that the *ZNF132* (SEN = 0.98, SPE = 0.99, AUC = 1.00) and *ZNF542* (SEN = 0.99, SPE = 0.95, AUC = 0.99) could nearly distinguish the CRC samples from adjacent normal samples without any misclassification (Figure 5). In addition, in the KRAS- subgroup, we found that none of the diagnostic abilities of the candidate biomarkers significantly decreased including *SEPT9*. It is worth mentioning that the diagnostic ability of *ZNF132* consistently superior than that of *SEPT9* in the KRAS+/KRAS- as well as the total samples, indicating that *ZNF132* could be of great potential for further utilization in CRC diagnosis.

In addition to the commonly used logistic regression method for evaluating the diagnostic performance, we also implemented 10 other machine-learning algorithms, including random forest (RF), supporting vector machine (SVM), neural network (NN), Naïve Bayes (NB), linear discriminant analysis (LDA), mixture discriminant analysis (MDA), flexible discriminant analysis (FDA), XGBoost and CatBoost for further assessment. The five-fold cross-validation method was applied to give an unbiased estimation of the model performance (Supplementary Table 13). It is found that in the combined dataset, the RF model achieved the best accuracy (Accuracy = 0.889). Meanwhile, in the KRAS+ subgroup, the NB algorithm performed best (Sensitivity = 0.99, Specificity = 0.97, Accuracy = 0.98) in the test data. In contrast, the NN model was found to be the best in the KRAS- subgroup (SEN = 0.75, SPE = 0.89, ACC = 0.82).

**Discussion**

Epigenetic alterations, including DNA methylation, histone methylation have been recognized as the potential biomarkers for multiple kinds of cancers. Meanwhile, the zinc finger gene family, which is the largest transcription factor family in human genome, has been widely reported to be differentially methylated in several kinds of tumor types. However, to our knowledge, few studies have concentrated on extensively explore the methylation alterations of zinc finger genes in colorectal cancer. Therefore, in our study, we for the first time integrated the high-throughput DNA methylation microarray datasets from TCGA and GEO, and yielded a total of 1,104 CRC samples, and 54 adenomas as well as 268 control/adjacent normal samples, ensuring the robustness and statistical power of the biomarkers identified. Based on the stringent quality control and preprocessing, we finally identified five hyper-methylated zinc finger genes as candidate biomarkers for CRC (nominated as ZEG model of CRC prediction). Furthermore, we validated these biomarkers in 104 pairs of CRC tumors and adjacent normal tissues of Han Chinese population. Among them, *ZNF132* reached the highest diagnostic ability, higher than that of *SEPT9*, which was recognized as the optimal CRC diagnostic biomarkers until now. To our surprise, we found that the diagnostic abilities of these five biomarkers were significantly higher in KRAS mutated CRC samples than that of the CRC samples without KRAS mutation. To further confirm the findings, we then recruited another cohort, consisting of 114 pairs of CRC tumors and adjacent normal tissues. Expectedly, we confirmed the differential diagnostic abilities of our biomarkers in KRAS+ and KRAS- subgroups. Therefore, our study revealed that the zinc finger gene-based biomarkers could distinguish the KRAS+ CRC samples with high confidence, suggesting the importance of combining genetic mutations and epigenetic alterations together for CRC diagnosis in the further study.

The interaction between genetic mutations and epigenetic alterations in the tumorigenesis of CRC has been reported previously. Claude Gazin et al performed a genome-wide RNA interference (RNAi) screen in K-ras-transformed NIH 3T3 cells, and identified 28 genes required for Ras-mediated epigenetic silencing of the pro-apoptotic Fas gene [21]. It is suggested that Ras-mediated epigenetic silencing could lead to the CRC tumorigenesis through the epigenetic inactivation of the key genes. Meanwhile, Takeshi Nagasaka et al also found that both KRAS and BRAF mutation could contribute to the global hyper-methylation phenotype of the CIMP genes in colon cancer. Furthermore, Ryan W Serra et al revealed that KRAS mutation could result in the hyper-methylation and transcriptional silencing of the CIMP genes through *ZNF304*, indicating the importance of zinc finger proteins in the tumorigenesis of CRC [22]. Herein, through the integration of public microarray datasets and our targeted bisulfite sequencing datasets, we identified five novel zinc finger genes (ZEG model), which were significantly associated with the KRAS mutation in CRC, suggesting that the KRAS mutation may alter the downstream pathway through the epigenetic regulation of these zinc finger genes.

Of the five zinc finger genes, *ZNF132* (zinc finger protein 132) was identified as the most promising biomarkers for CRC diagnosis in our combined analysis. *ZNF132* is located at the 19q13.4 and belongs to C2H2 zinc finger protein family [23]. Previous studies have identified the DNA methylation alterations of *ZNF132* in breast cancer, esophageal squamous cell carcinoma (ESCC), oropharyngeal squamous cell carcinoma and prostate cancer [1, 24-26]. Especially, *ZNF132* hyper-methylation could reduce the Sp1 transcript factor activity and decreased the abilities of cell in growth, migration and invasion, and tumorigenicity of cells in a nude mouse model of ESCC [1]. In our study, we both identified the hyper-methylation and down-regulation of *ZNF132* in CRC tumors, especially in KRAS mutated samples, suggesting its biological implications in CRC tumorigenesis. *ESR1* (estrogen receptor alpha) has been recognized as a tumor-suppressor gene and an estrogen receptor gene, encodes the main mediator of estrogen effect in breast epithelia and has also been shown to be activated by epidermal growth factor (EGF). The hyper-methylation status of *ESR1* has been reported previously in lung adenocarcinoma, breast cancer, prostate cancer, squamous cell cervical cancer and colorectal cancer [27-32]. Meanwhile, the *ESR1* hyper-methylation is also correlated with poor prognosis and drug response in breast cancer [33, 34]. Additionally, the hyper-methylation of *ESR1* promoter has been found to be associated with KRAS mutation, which was in accordance with our results [35]. *ZNF229* (zinc finger protein 229) is a protein-coding gene and few studies have suggested the hyper-methylation status of *ZNF229* in the diagnosis of cancers. The biological functions and its implications in colorectal cancers should be further explored. *ZNF542* (zinc finger protein 542) has been found to be involved in the epigenetic regulation of puberty through transcriptional repression [36]. Moreover, a CpG site located at *ZNF542* has been found to be a promising biomarker for esophageal squamous cell carcinoma [37]. Meanwhile, a pan-cancer study revealed that *ZNF542* was significantly hyper-methylated in ten kinds of cancers [38]. *ZNF677* (zinc finger protein 677) is located at the chromosomal region 19q13 and was found to regulate the putative tumor cell growth suppressor in non-small cell lung cancers through hyper-methylation [39]. In addition, *ZNF677* is frequently downregulated by promoter methylation in primary papillary thyroid cancers (PTC) and the decreased expression of *ZNF677* is significantly correlated with poor survival [40].

Until now, the majority of CRC patients were still diagnosed at later stages, especially at the developing countries. As a result, better diagnostic markers and models are urgently required. Compared with the other kinds of biomarkers, DNA methylation alterations may occur in advance of the mRNA and protein changes and thus could be ideal early diagnostic biomarkers for CRC. However, the high heterogeneity of the CRC makes it challenging to present a panel of DNA methylation-based biomarkers with high robustness and accuracy. Currently, *SEPT9* has been identified as CRC-specific DNA methylation-based diagnostic biomarkers and the first release of the cfDNA *SEPT9* DNA methylation assay achieved considerable sensitivity (72%) and specificity (86%) in CRC detection using plasma [41]. After that, an updated version of the assay (Epi proColon 2.0) showed better sensitivities (68–95%) and specificities (80–99%) in CRC diagnosis [42]. However, as shown in this study, *ZNF132* and *ESR1* showed comparable or even better diagnostic ability than that of *SEPT9*, indicating that these two genes may be of great potential for further utilization. Furthermore, we found that the diagnostic abilities of all five candidate biomarkers as well as the *SEPT9* were significantly affected by the KRAS mutation status, and none of which achieved satisfactory sensitivities and specificities in the diagnosis of KRAS- CRC patients. Therefore, it is suggested that the accurate diagnosis of CRC should be conducted according to the mutation landscape of the patients. Further studies are required to identify better DNA-methylation based biomarkers in KRAS- CRC patients. In addition, the integration of cfDNA mutation detection and methylation detection might be of great potential for CRC early diagnosis.

**Materials and Methods**

**Integration of public datasets and biomarker discovery**

Public high-throughput DNA methylation microarray datasets (Illumina HumanMethylation 450K) were searched exhaustedly from TCGA (The Cancer Genome Atlas) and GEO (Gene expression Omnibus) databases. In summary, two datasets from TCGA and nine datasets from GEO were included, yielding a total of 1,104 CRC tumor samples, 268 control and 54 adenoma samples (Supplementary Table 2). Meanwhile, the comprehensive list of genes in zinc finger family (n=1,594) were also obtained from HGNC (Supplementary Table 1).

To identify the robust DNA methylation-based biomarkers, we conducted the differential methylation region (DMR) analysis for all the included samples. As shown in previous study, the methylation linkage equilibrium decreased significantly when the block is longer than 1,000 bp, we defined the methylation region (MR) to have at least 6 CpG sites and should be less than 1,000 bp. We then arranged all the CpG sites in the high-throughput microarray according to their genomic coordinates and performed the sliding window method to capture the candidate DMRs. In total, 6,166 methylation regions were identified. Furthermore, differential methylation analyses were conducted for all these MRs. In summary, we extracted 85 candidate DMRs fulfilled our standard (McaM > 0.50, MadM > 0.50 and McoM < 0.30). In addition, the DMRs should be reliable and robust for liquid biopsy when the noises of the DNAs originated from the peripheral blood were mixed, it is necessary that the methylation rate of the DMRs should be very low in the peripheral blood. Therefore, we integrated the public high-throughput microarray datasets of the whole blood (WB, n = 1438), peripheral blood mononuclear cells (PBMC, n = 111) and peripheral blood leucocytes (PBL, n = 529) as reference for DMR identification. As a result, 32 DMRs were retained due to their low methylation rates in the WB, PBMC or PBL (mean methylation rate < 0.10). It is found that several DMRs were located at the same gene and we then selected the DMR with the most significant differences between CRC and control tissues for each gene (N = 10). Meanwhile, to obtain the DMRs that may regulate the expression of neighboring genes, we further selected the DMRs with transcription factor binding sites (TFBS) and correlate significantly with the expression of neighboring genes. In total, seven out of the 10 candidate DMRs were selected for validation. However, due to the difficulties in the primers design due to CG percent, PolyT, and the number of SNPs, two candidate DMRs (*SALL1*, *ZSCAN23*) were removed. Finally, we obtained the top five candidate DMRs (*ESR1*, *ZNF123*, *ZNF229*, *ZNF542* and *ZNF677*) for further validation.

**Patients, samples, and DNA**

The CRC tumor and paired adjacent normal tissues were obtained from Affiliated Hospital of Nantong University between the years from 2016 to 2018. The patients recruited have not been treated with any neo-adjuvant therapy before. At least two professional pathologists evaluated all tumor samples carefully. All procedures performed in this study were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments. The study was approved by the institutional review boards of Affiliated Hospital of Nantong University. Written informed consent was obtained from each participant of the study. All tumor and adjacent normal tissues were immediately frozen at – 80 °C after surgical resection.

**Targeted bisulfite sequencing assay and detection of KRAS mutation status**

DNA extraction and bisulfite conversion were conducted as previously described. After carefully evaluated the CG percent, PolyT, and the occurrence of SNPs in the targeted regions of the candidate DMRs, we designed the primers to detect them in a panel for NGS sequencing (Supplementary Table 7). The detailed procedure for bisulfite sequencing was shown in previous study [37]. BSseeker2 was applied for reads mapping and methylation calling [61]. After that, we removed the samples with bisulfite conversion rate < 98%. Meanwhile, the average coverage and missing rate for each CpG site was calculated and utilized for quality control (average coverage > 20X, missing rate < 20%). In addition, the samples with missing rate > 30% were also filtered out. Tumor DNA was detected for the mutation in codon 12, 13, 59, 61, 117, and 146 of KRAS and NRAS gene, and the mutation in codon 600 of BRAF gene [ref 38] by the FastTarget next generation sequencing [ref 39].

[ref 38] Sepulveda, A.R., et al., *Molecular Biomarkers for the Evaluation of Colorectal Cancer: Guideline From the American Society for Clinical Pathology, College of American Pathologists, Association for Molecular Pathology, and American Society of Clinical Oncology.* Arch Pathol Lab Med, 2017. **141**(5): p. 625-657.

[ref 39] Zhang, R., et al., *An ERCC4 regulatory variant predicts grade-3 or -4 toxicities in patients with advanced non-small cell lung cancer treated by platinum-based therapy.* Int J Cancer, 2018. **142**(6): p. 1218-1229.

**Statistical analysis and machine learning methods**

The Wilcoxon rank-sum test was performed in the discover stage to identify the differential methylation sites and regions between CRC, adenomas and adjacent normal tissues. Moreover, the differential methylation status (odds ratios) between tumor and adjacent normal tissues of the DMRs were calculated with logistic regression. The Benjamini-Hochburg correction was utilized for multiple test correction. In addition, the logistic regression (Package stats), support vector machine (SVM, Package e1071), random forest (Package randomForest), Naïve Bayes (Package e1071), neural network (Package nnet), linear discriminant analysis (LDA, Package mda), mixture discriminant analysis (MDA, Package mda), flexible discriminant analysis (FDA, Package mda), gradient boosting machine (Package gbm), catboost (Package catboost) and XGBoost (Package xgboost) were used for classification along with five-fold cross-validation. In addition, sensitivity, specificity, and accuracy were obtained from the logistic regression model. All statistical analyses were all conducted using R (v3.4.3).

**Abbreviations**

ZFP: Zinc finger protein

ZFG: Zinc finger gene

SEN: Sensitivity

SPE: Specificity

ACC: Accuracy

**Funding**

The work was supported by the National Natural Science Foundation of China (81201350) ，Postdoctoral Science Foundation of China (2018M641919，2019M651930)，Jiangsu Province's Key Young Medicine Talents Program (QNRC2016688) and Nantong People's Livelihood Science and Technology Plan (MS12018032).

**Author contributions**

All authors are involved in the following contribution to the paper: (1) conception and design the study, or analysis and interpretation of data; (2) drafting the article or revising it critically for important intellectual content; (3) final approval of the version to be published.

**Conflict of interests**

The authors declare that they have no competing interests

**Reference**

1. Jiang, D., et al., *Epigenetic silencing of ZNF132 mediated by methylation-sensitive Sp1 binding promotes cancer progression in esophageal squamous cell carcinoma.* Cell Death Dis, 2018. **10**(1): p. 1.

2. He, Y., et al., *Hypomethylation of the hsa-miR-191 locus causes high expression of hsa-mir-191 and promotes the epithelial-to-mesenchymal transition in hepatocellular carcinoma.* Neoplasia, 2011. **13**(9): p. 841-53.

3. Schubeler, D., *Function and information content of DNA methylation.* Nature, 2015. **517**(7534): p. 321-326.

4. Guo, S., et al., *Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA.* Nat Genet, 2017. **49**(4): p. 635-642.

5. Guo, S., et al., *Identification and validation of the methylation biomarkers of non-small cell lung cancer (NSCLC).* Clin Epigenetics, 2015. **7**: p. 3.

6. Koch, A., et al., *Analysis of DNA methylation in cancer: location revisited (vol 15, pg 459, 2018).* Nature Reviews Clinical Oncology, 2018. **15**(7): p. 467-467.

7. Chen, S., et al., *Genome-Wide DNA Methylation Profiles Reveal Common Epigenetic Patterns of Interferon-Related Genes in Multiple Autoimmune Diseases.* Front Genet, 2019. **10**: p. 223.

8. Guo, S., et al., *Genome-wide DNA methylation patterns in CD4+ T cells from Chinese Han patients with rheumatoid arthritis.* Mod Rheumatol, 2017. **27**(3): p. 441-447.

9. Ding, W., et al., *Genome-Wide DNA Methylation Analysis in Systemic Sclerosis Reveals Hypomethylation of IFN-Associated Genes in CD4(+) and CD8(+) T Cells.* J Invest Dermatol, 2018. **138**(5): p. 1069-1077.

10. Freitas, M., et al., *A novel DNA methylation panel accurately detects colorectal cancer independently of molecular pathway.* J Transl Med, 2018. **16**(1): p. 45.

11. Semaan, A., et al., *SEPT9 and SHOX2 DNA methylation status and its utility in the diagnosis of colonic adenomas and colorectal adenocarcinomas.* Clin Epigenetics, 2016. **8**: p. 100.

12. Wills, B., E. Gorse, and V. Lee, *Role of liquid biopsies in colorectal cancer.* Curr Probl Cancer, 2018. **42**(6): p. 593-600.

13. Okugawa, Y., W.M. Grady, and A. Goel, *Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers.* Gastroenterology, 2015. **149**(5): p. 1204-1225 e12.

14. Barault, L., et al., *Discovery of methylated circulating DNA biomarkers for comprehensive non-invasive monitoring of treatment response in metastatic colorectal cancer.* Gut, 2018. **67**(11): p. 1995-2005.

15. Sun, J., et al., *The role of (m)SEPT9 in screening, diagnosis, and recurrence monitoring of colorectal cancer.* BMC Cancer, 2019. **19**(1): p. 450.

16. Song, L., et al., *The performance of the SEPT9 gene methylation assay and a comparison with other CRC screening tests: A meta-analysis.* Sci Rep, 2017. **7**(1): p. 3032.

17. Church, T.R., et al., *Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer.* Gut, 2014. **63**(2): p. 317-25.

18. Ahlquist, D.A., et al., *The stool DNA test is more accurate than the plasma septin 9 test in detecting colorectal neoplasia.* Clin Gastroenterol Hepatol, 2012. **10**(3): p. 272-7.e1.

19. Krishna, S.S., I. Majumdar, and N.V. Grishin, *Structural classification of zinc fingers: survey and summary.* Nucleic Acids Res, 2003. **31**(2): p. 532-50.

20. Jen, J. and Y.C. Wang, *Zinc finger proteins in cancer progression.* J Biomed Sci, 2016. **23**(1): p. 53.

21. Gazin, C., et al., *An elaborate pathway required for Ras-mediated epigenetic silencing.* Nature, 2007. **449**(7165): p. 1073-7.

22. Serra, R.W., et al., *A KRAS-directed transcriptional silencing pathway that mediates the CpG island methylator phenotype.* Elife, 2014. **3**: p. e02313.

23. Tommerup, N. and H. Vissing, *Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders.* Genomics, 1995. **27**(2): p. 259-64.

24. Lleras, R.A., et al., *Hypermethylation of a cluster of Kruppel-type zinc finger protein genes on chromosome 19q13 in oropharyngeal squamous cell carcinoma.* Am J Pathol, 2011. **178**(5): p. 1965-74.

25. Abildgaard, M.O., et al., *Downregulation of zinc finger protein 132 in prostate cancer is associated with aberrant promoter hypermethylation and poor prognosis.* Int J Cancer, 2012. **130**(4): p. 885-95.

26. Stefansson, O.A., et al., *A DNA methylation-based definition of biologically distinct breast cancer subtypes.* Mol Oncol, 2015. **9**(3): p. 555-68.

27. Li, L.C., et al., *Age-dependent methylation of ESR1 gene in prostate cancer.* Biochem Biophys Res Commun, 2004. **321**(2): p. 455-61.

28. Martinez-Galan, J., et al., *ESR1 gene promoter region methylation in free circulating DNA and its correlation with estrogen receptor protein expression in tumor tissue in breast cancer patients.* BMC Cancer, 2014. **14**: p. 59.

29. Lin, Q., et al., *RASSF1A, APC, ESR1, ABCB1 and HOXC9, but not p16INK4A, DAPK1, PTEN and MT1G genes were frequently methylated in the stage I non-small cell lung cancer in China.* J Cancer Res Clin Oncol, 2009. **135**(12): p. 1675-84.

30. Kim, M.S., J. Lee, and D. Sidransky, *DNA methylation markers in colorectal cancer.* Cancer Metastasis Rev, 2010. **29**(1): p. 181-206.

31. Kirn, V., et al., *ESR1 promoter methylation in squamous cell cervical cancer.* Anticancer Res, 2014. **34**(2): p. 723-7.

32. Elliott, G.O., et al., *Quantitative profiling of CpG island methylation in human stool for colorectal cancer detection.* Int J Colorectal Dis, 2013. **28**(1): p. 35-42.

33. Ramos, E.A., et al., *Simultaneous CXCL12 and ESR1 CpG island hypermethylation correlates with poor prognosis in sporadic breast cancer.* BMC Cancer, 2010. **10**: p. 23.

34. Mastoraki, S., et al., *ESR1 Methylation: A Liquid Biopsy-Based Epigenetic Assay for the Follow-up of Patients with Metastatic Breast Cancer Receiving Endocrine Treatment.* Clin Cancer Res, 2018. **24**(6): p. 1500-1510.

35. Horii, J., et al., *Methylation of estrogen receptor 1 in colorectal adenomas is not age-dependent, but is correlated with K-ras mutation.* Cancer Sci, 2009. **100**(6): p. 1005-11.

36. Lomniczi, A., et al., *Epigenetic regulation of puberty via Zinc finger protein-mediated transcriptional repression.* Nat Commun, 2015. **6**: p. 10195.

37. Pu, W., et al., *Targeted bisulfite sequencing identified a panel of DNA methylation-based biomarkers for esophageal squamous cell carcinoma (ESCC).* Clin Epigenetics, 2017. **9**: p. 129.

38. Gevaert, O., R. Tibshirani, and S.K. Plevritis, *Pancancer analysis of DNA methylation-driven genes using MethylMix.* Genome Biol, 2015. **16**: p. 17.

39. Heller, G., et al., *DNA methylation transcriptionally regulates the putative tumor cell growth suppressor ZNF677 in non-small cell lung cancers.* Oncotarget, 2015. **6**(1): p. 394-408.

40. Li, Y., et al., *ZNF677 Suppresses Akt Phosphorylation and Tumorigenesis in Thyroid Cancer.* Cancer Res, 2018. **78**(18): p. 5216-5228.

41. deVos, T., et al., *Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer.* Clin Chem, 2009. **55**(7): p. 1337-46.

42. Payne, S.R., *From discovery to the clinic: the novel DNA methylation biomarker (m)SEPT9 for the detection of colorectal cancer in blood.* Epigenomics, 2010. **2**(4): p. 575-85.