

87526_Auto_Edited-check.docx

Name of Journal: *World Journal of Gastrointestinal Oncology*

Manuscript NO: 87526

Manuscript Type: ORIGINAL ARTICLE

Basic Study

Early results of the integrative epigenomic-transcriptomic landscape of colorectal adenoma and cancer

Lu YW *et al.* Epigenomic-transcriptomic landscape of CRC

Abstract

BACKGROUND

Aberrant methylation is common during the initiation and progression of colorectal cancer (CRC), and detecting these changes that occur during early adenoma (ADE) formation and CRC progression has clinical value.

AIM

To identify potential DNA methylation markers specific to ADE and CRC.

METHODS

Here, we performed SeqCap targeted bisulfite sequencing and RNA-seq analysis of colorectal ADE and CRC samples to profile the epigenomic-transcriptomic landscape.

RESULTS

Comparing 22 CRC and 25 ADE samples, global methylation was higher in the former, but both showed similar methylation patterns regarding differentially methylated gene positions, chromatin signatures, and repeated elements. High-grade CRC tended to exhibit elevated methylation levels in gene promoter regions compared to those in low-grade CRC. Combined with RNA-seq gene expression data, we identified 14 methylation-regulated differentially expressed genes, of which only *AGTR1* and *NECAB1* methylation had prognostic significance.

CONCLUSION

Our results suggest that genome-wide alterations in DNA methylation occur during the early stages of CRC and demonstrate the methylation signatures associated with colorectal ADEs and CRC, suggesting prognostic biomarkers for CRC.

Key Words: Colorectal cancer; Epigenomic alteration; Transcriptome; Methylation-regulated differentially expressed genes

Lu YW, Ding ZL, Mao R, Zhao GG, He YQ, Li XL, Liu J. Early results of the integrative epigenomic-transcriptomic landscape of colorectal adenoma and cancer. *World J Gastrointest Oncol* 2023; In press

Core Tip: Despite enhanced global methylation in colorectal cancer (CRC), both adenoma and CRC share similar methylation patterns. High-grade CRC tends to exhibit elevated methylation levels in gene promoter regions compared to those in low-grade CRC. The integration of gene expression data identified 14 methylation-regulated differentially expressed genes, with *AGTR1* and *NECAB1* methylation being prognostically significant.

INTRODUCTION

Mortality due to colorectal cancer (CRC) has increased worldwide^[1]. In China, this disease is the third most common malignancy and the fourth leading cause of cancer-related deaths^[2]. CRC is caused by the accumulation of genetic and epigenetic modifications, and the adenoma (ADE)-carcinoma progression pathway accounts for 65%-70% of CRC cases^[3]. Many epigenetic alterations have been identified using next-generation sequencing; however, only a few have been translated into clinical settings.

Increasing evidence suggests that epigenetic factors play important roles in CRC initiation and progression. Among the various epigenetic modifications, methylation has been the most extensively studied in this disease. Several studies have employed genome-wide methylation analyses to identify methylation biomarkers in CRC^[4-9]. ADEs are precursors of CRC, and Luo *et al*^[10] found that methylation modifications are early events during the progression from tubular ADEs to CRC. Moreover, ADEs share certain methylation modifications with CRC^[11]. As such, the detection of methylation modifications occurring in ADE is crucial for identifying novel markers, which could improve CRC detection and prognosis. However, comprehensive studies on colorectal

ADE and CRC with respect to DNA methylation and gene expression profiles are still lacking.

In this study, we used SeqCap targeted bisulfite sequencing to identify common epigenetic patterns in ADE and CRC. Compared to the widely used Illumina HumanMethylation EPIC BeadChip, this technology covers more CpG sites, thus achieving higher coverage^[12]. Moreover, we used RNA-seq to detect the methylation signatures that orchestrate transcriptional dysregulation. Finally, we explored the clinical significance of these methylation markers using The Cancer Genome Atlas Colon Adenocarcinoma (TCGA-COAD) database.

MATERIALS AND METHODS

Clinical samples

Twenty-six colorectal tumors, 29 colorectal ADEs, and 18 adjacent normal tissues were collected from patients at the First Affiliated Hospital of Kunming Medical University (Kunming, China). Written informed consent was obtained from all patients for genomic analysis. This study was approved by the Ethics Committee of the First Affiliated Hospital of Kunming Medical University. Adjacent normal tissue samples (at least 5 cm away from the tumor) were obtained through surgical resection of the intestine. The clinical characteristics of the patients are shown in Supplementary Table 1. All samples used in this study were obtained at the time of diagnosis, before treatment.

DNA extraction

The genomic DNA of these tissues was extracted using the TIANamp Genomic DNA Kit (TIANGEN, China), and total RNA was extracted from the tissue samples using the RNA Easy Fast Tissue/Cell Kit (TIANGEN, China), according to the manufacturer's instructions. The concentration and quality of DNA and RNA were assessed using a Qubit fluorometer (Thermo Fisher Scientific).

SeqCap targeted bisulfite sequencing

An EZ Methylation-Direct Kit (Zymo Research, CA, United States) was used to perform bisulfite conversion. To construct the target bisulfite sequencing library, we used the NimbleGen SeqCap Epi Enrichment System to capture the target genomic regions according to the manufacturer's protocols. Paired-end sequencing (100 bp) was performed on an Illumina HiSeq 2500 platform (Illumina, San Diego, CA, United States) according to the manufacturer's protocol.

Bisulfite sequencing data preprocessing

FastQC (v0.11.3) was used to assess sequence quality and generate quality reports. Bisulfite sequencing reads were pre-processed using Trimmomatic (v0.39). Both 6 bp Illumina adapters and low-quality sequences (base quality < 20) were filtered out prior to the analysis. The remaining reads were aligned against the human genome (GRCh38) using Bismark (v0.23.1)^[13] with the following parameter settings: -bowtie2 -parallel 4 -N 0 -L 20 -quiet -un. The methylation status of each CpG was determined using the Bismark methylation extractor function, and only methylation within the CpG context was retained for further analysis (Supplementary Table 2).

The R package methylKit (v1.20.0)^[14] was used for the differential methylation analysis. CpG sites with low (< 5') or extremely high (> 99th percentile) coverage were excluded. CpG sites mapped to X and Y chromosomes or mitochondria were removed, and those shared by all samples were retained for further analysis. We merged the CpG site methylation profiles of the 61 samples based on CpG coordinates, yielding 822353 CpG sites covered by all samples. CpG sites that overlapped with an SNP from dbSNP build 151 were removed, and 811441 CpG sites were retained for further analysis.

For the analysis of differentially methylated positions (DMPs), an absolute methylation difference > 20% and a false discovery rate (FDR) < 0.05 were used. For the analysis of differentially methylated regions (DMRs), the genome was tiled based on a window size of 200 bp with a step size of 200 bp, and the CpG methylation rates were averaged across multiple sites within the region. Next, the differences between the two

groups were tested using the calculated DiffMeth function provided by methylKit. The DMRs were determined using the following thresholds: > 5 CpGs in the DMRs, absolute methylation difference > 10%, and FDR < 0.05.

ChromHMM analysis

Chromatin state annotations were generated using ChromHMM (v1.18)^[15]. We generated an 18-state model with 11 healthy human tissues and one cell line using the LearnModel function to predict the chromatin state^[15].

Genomic enrichment analysis

The LOLA package (v1.24.0)^[16] was used to annotate the methylation data using histone marks, chromatin states, and repetitive DNA element databases. Genomic enrichment analyses were performed as previously described^[17]. Histone and chromatin states were derived from 11 healthy human tissues and one cell line from the NIH Roadmap Epigenomics region data^[18]. Repetitive DNA element databases were obtained using the UCSC table browser.

Annotation to genomic locations

DMPs and DMRs were annotated to hg38 CpG and gene features using the annotatr package (v1.20.0). TCGA COAD-derived DMPs were annotated in the IlluminaHumanMethylation450kanno.ilmn12.hg19 package (v1.40.0).

RNA-seq data analysis

RNA-sequencing libraries were constructed using a TruSeq RNA Sample Preparation Kit (Illumina). Transcriptome data were mapped using hisat2 (v2.2.1). The read counts were quantified using featureCounts (v2.0.0). Finally, differentially expressed genes (DEGs) were identified using DESeq2 (v1.36.0) with an FDR < 0.05.

HM450K DNA data analysis

TCGA data were obtained using the TCGA biolinks package (beta values at the CpG sites were obtained). HM450K DNA data were processed using the ChAMP package^[19].

TCGA gene expression data analysis

TCGA mRNA expression data for colon cancer were also obtained using the TCGAbiolink package. The expression matrix was normalized using DESeq2 (v1.36.0). Finally, the DEGs were screened using DESeq2.

Identification of methylation-regulated DEGs

First, we identified differentially methylated genes (DMGs) located in the gene promoter. We then overlapped the DMGs with the DEGs, and the remaining genes were candidate methylation-regulated DEGs (meDEGs). Candidate meDEG expression and methylation analyses were performed using TCGA HM450 microarray and RNA-seq data from COAD. Hypermethylated genes and those for which expression was downregulated were selected for further analysis. Spearman rank correlation analysis was performed on the beta values and gene expression levels for these gene-probe pairs using the psych package. The threshold was set at rho values < -0.2 and an FDR < 0.05 .

Survival analysis

Survival analysis was performed using survminer and survivor packages. Kaplan-Meier curves were used to compare the survival between the high- and low-methylation groups. The median methylation level of the selected gene was used as the cutoff point.

RESULTS

Methylation alterations in ADE and CRC

To investigate the methylation changes associated with ADE and CRC, we compared the methylation patterns in these tissues with those in normal samples. The average methylation values of ADE and normal samples were 0.42 and 0.41, respectively, and

we found no significant differences in methylation at CpG sites between these groups (Wilcoxon rank-sum $P = 0.75$; Figure 1A). However, principal component analysis (PCA) clustering suggested differences between the ADE and control groups, as they were separated across the most meaningful dimensions (Figure 1B). The average methylation values of CRC and normal samples were 0.45 and 0.40, respectively, and CRC samples tended to have higher methylation values (Wilcoxon rank-sum test, $P = 0.12$; Figure 1C). Moreover, PCA clustering showed some overlap between the CRC and normal groups (Figure 1D).

We then performed differential methylation analysis with a specific threshold (absolute methylation difference $> 20\%$ and $FDR < 0.05$) to define the DMPs between ADE and CRC. We identified substantial alterations in both conditions, identifying 48291 DMPs associated with ADE (30009 hypermethylation and 18282 hypomethylation; Figure 1E; a list of ADE DMPs is available in Supplementary Table 3) and 95887 associated with CRC (63107 hypermethylation and 32780 hypomethylation; Figure 1F; a list of CRC DMPs is available in Supplementary Table 4). Hypermethylation was more frequently observed at CpG sites in both ADE and CRC samples.

DMP locations in ADE and CRC

Hypermethylation was the predominant event in both ADE and CRC [accounting for 62% (30009/48291) and 66% (63107/95887) of the changes, respectively; Figure 2A]. To investigate the possible functional consequences of methylation alterations, we first annotated the DMPs with CpG island features. We observed that hypermethylation occurred more commonly at CpG islands in ADE and CRC, with Fisher's $P < 0.001$ and odds ratios (ORs) = 2.7 and 5.3, respectively. In contrast, hypomethylation in ADE and CRC was more common at open sea and shelf locations, with Fisher's $P < 0.001$ and ORs of 2.4 and 2.6 for open sea and ORs = 2.6 and 2.3 for shelf locations, respectively (Figure 2B).

Regarding gene locations, CRC and ADE displayed similar patterns. DMPs comprising hypermethylation occurred preferentially at 1-5k regions and 5' untranslated regions (UTRs), with Fisher's $P < 0.001$, ORs = 1.7 and 1.4 for 1-5k regions, and ORs = 1.8 and 1.4, respectively, for 5'UTRs. DMPs comprising hypomethylation were enriched at 1-5k and intergenic and intronic sequences, with Fisher's $P < 0.001$, ORs = 2.3 and 2.0 for 1-5k regions, ORs = 2.1 and 2.2 for intergenic sequences, and ORs = 1.6 and 1.6 for intronic sequences, respectively. Subsequently, we found that the hypermethylated DMP count was higher in both ADE and CRC, but the number of CpGs that mapped to genes was similar. These results suggest that hypomethylation is more widespread across different genes under both conditions (Figure 2D, $P < 0.001$).

Next, we performed an intersection analysis of DMPs in ADE and CRC. To this end, we applied the SuperExactTest^[20] to compare the co-occurrence patterns of DMPs between these samples (Figure 2E and Supplementary Table 5). These intersections were all significantly over-enriched (all $P < 0.001$), with common CRC- and ADE hyper-DMPs and common CRC-hypo and ADE-hypo DMPs displaying the highest degree of overlap. Specifically, 82% (24639/30009) of the hypermethylated DMPs in ADE were shared with CRC.

Chromatin landscape of CRC and ADE methylation alterations

To assess the enrichment of regulatory regions, we first performed over-enrichment analyses using the LOLA package based on CRC and ADE-related DMPs with different histone modifications. Regarding ADE-associated methylation changes, we observed that hypermethylation sites were enriched in H3K27me3 modifications and active enhancer/promoter-associated H3K4me1 (Figure 3A). Furthermore, we assessed the over-enrichment of ADE DMPs in different chromatin states (Figure 3B). The ADE-associated hypermethylated DMPs preferentially occurred at bivalent transcription start sites (states 14) and zinc-finger (ZNF) and enhancer sites (states 10 and 11). In addition, ADE-associated hypermethylated sites were enriched in heterochromatic regions and ZNF genes/repeats (states 12 and 13). Regarding ADE-associated hypomethylation, the

most enriched chromatin states (15-18) involved repressive functions, such as polycomb-associated sites. CRC exhibited similar methylated regulatory regions as those enriched in ADE (Figures 3C and D). Furthermore, hypomethylated DMPs were significantly enriched at H3K27me3 in normal colorectal tissues.

Next, we performed over-enrichment analyses of CRC- and ADE-related DMPs with repetitive DNA elements (Figure 3E). In ADE and CRC, hypomethylation preferentially occurred at short interspersed retrotransposable elements, long interspersed retrotransposable elements (LINEs), long terminal repeat elements (LTRs), DNA repeat elements, rolling circles, and satellite sites. Additionally, the hypomethylation of DMPs was associated with tRNA.

Finally, we performed HOMER analysis to identify the enrichment of DMPs in the transcription factor (TF) motifs. We observed both similarities and differences between the ADE and CRC groups (Figure 3F). The hypermethylation of DMPs in ADE was associated with homeobox TF motifs (Fisher's $P < 0.001$, OR = 7.4) and T-box motifs (Fisher's $P = 0.001$, OR = 2.7). The hypomethylation of DMPs in ADE was enriched in bHLH TF motifs (Fisher's $P < 0.001$, OR = 6.9). In CRC, hypermethylation particularly affected MADS and homeobox TF motifs (Fisher's $P < 0.001$, ORs = 33.2 and 4.8, respectively), whereas hypomethylation was involved in ETS and bZIP TF sites (Fisher's $P < 0.01$, ORs = 13.2 and 3.5). These findings highlight the clear functional discrepancy between ADE and CRC in terms of TF-motif enrichment.

Hypermethylated DMRs occur primarily at promoters in ADE and CRC

To identify DMRs in ADE and CRC, the WGBS dataset was analyzed using methylkit with a tile window size of 200 bp. We detected 6219 hypermethylated DMRs (hyperMe-DMRs) and 3365 hypoMe-DMRs in ADE (Supplementary Table 6) and 10424 hyperMe-DMRs and 4817 hypoMe-DMRs in CRC (Supplementary Table 7). We then used CIRCOS to visualize the data as a multilayer circular plot, shown in Figures 4A and B. The outer layer contained information including the chromosome, DMR position, and

change in the methylation value (-1 to 1). The inner layer lists changes in CpG island methylation.

In this study, hypermethylation was the predominant event at the DMR level in both ADE and CRC. The majority (75.5%) of the hyperMe-DMRs were located in the promoter regions of the genes, whereas hypoMe-DMRs were more likely to be located in the intergenic and intron regions (Figure 4C). To identify convergent DMRs located in gene promoter regions, we applied UpSet plots to determine the overlap in hypoMe-DMRs and hyperMe-DMRs between ADE and CRC. We found 2265 hypoMe-DMRs and 452 hyperMe-DMRs that were shared between the two conditions (Figure 4D).

Functional enrichment analyses of DMRs in promoters

To reveal the biological functions of DMRs, we performed Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) analyses of the related genes. GO enrichment analysis revealed that these genes were mainly enriched in homophilic cell adhesion *via* plasma membrane adhesion molecules and the neuronal cell body and DNA-binding transcription activator activity (Figure 5A). The KEGG enrichment analysis results showed that these genes were mainly enriched in the calcium signaling pathway, neuroactive ligand-receptor interactions, and other pathways (Figure 5B).

Gene expression alterations in ADE and CRC

We next performed RNA-seq analysis of 13 ADE, 10 CRC, and nine normal samples to detect transcriptional differences. First, we analyzed the transcriptional differences between ADE and normal samples. By applying t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis to the expression matrix of the top 2000 most variable genes, we classified the samples into two groups (Figure 6A). We used the “DESeq2” R package for differential expression analysis, with cut-off criteria of an adjusted $P < 0.05$ and $|\log_2FC| > 1$. We identified 2399 DEGs between the ADE and normal samples, of which the expression levels of 1311 were upregulated and those of 1088 were downregulated (Figures 6B and C).

We then analyzed the transcriptional differences between CRC and normal samples. t-SNE analysis of the expression matrix of the top 2000 most variable genes revealed mixed samples between the CRC and normal groups (Figure 6D). We identified 460 DEGs between CRC and normal samples, including 173 for which expression was upregulated and 287 for which expression was downregulated (Figures 6E and F).

By integrating the DEG analysis of ADE and CRC samples, we identified 235 genes (Figure 6G), including 177 for which expression was downregulated and 58 for which expression was upregulated in both CRC and ADE samples. To identify meDEGs, we applied UpSet plots to identify the overlap between hypoMe-DMRs and genes for which expression was upregulated, as well as hyperMe-DMRs and genes for which expression was downregulated. We finally identified 24 meDEGs, including 21 hypermethylated and three hypomethylated genes (Figure 6H).

Integrative analysis identifies meDEGs in CRC

To validate the meDEGs, we analyzed their methylation and expression levels using the COAD dataset of TCGA. Of the 24 meDEGs, all 24 genes were found to be expressed in the COAD RNA-seq dataset, with expression levels of three genes being upregulated and those of 21 being downregulated in CRC (Figure 7A). All gene expression patterns were consistent with those in our dataset.

We further examined the methylation status of meDEGs in the COAD 450K dataset and found methylation data for 18 of the 24 genes. Among these, 11 genes were hypermethylated in CRC tissues (Figure 7B). All methylation patterns were consistent with those observed in our dataset. Methylation levels of the seven meDEGs showed a significant and highly negative correlation with their expression ($FDR < 0.05$, Spearman's $\rho < -0.2$; Figure 7C). Next, to explore the functional relationships among the meDEGs, we used Gene-MANIA to construct a protein-protein interaction)/network for the meDEGs, and the results are shown in Figure 7D.

To further explore the clinical significance of the meDEGs, we performed a survival analysis based on the seven meDEGs in the COAD dataset. Our results showed that the

AGTR1-low-methylation group had a higher survival rate than the high-methylation group (Figure 7E). Similarly, the NECAB1-low-methylation group had a higher survival rate than the high-methylation group (Figure 7F).

Comparative analyses of methylation in CRC and ADE of different grades

We divided ADE into high-grade adenoma (HGA) and low-grade adenoma (LGA) groups base on the criterion provided by Lieberman *et al*^[21]. We divided CRC into high-grade cancer (HGC) at stage \geq IIB and low-grade cancer (LGC) at stage $<$ IIB. More than 60% of DMRs that were observed in both LGA (75.4%, 1535/2035) and HGA (56.1%, 1499/2674) were hyper-methylated compared to the status in normal tissues (Figure 8A, Supplementary Tables 8 and 9). However, with LGA as the reference, most DMRs observed in HGA were hypo-methylated (64.6%, 737/1140; Figure 8A, Supplementary Table 10). Additionally, there were substantial overlaps between genes with DMRs in LGA compared to normal tissues and those compared to HGA, suggesting a similar epigenetic process (Figure 8C). In CRC, more than 70% of DMRs that were observed in both LGC (71.6%, 2245/3136) and HGC (78.5%, 2920/3721) were hyper-methylated compared to the status normal tissues (Figure 8B, Supplementary Tables 11 and 12). Moreover, with LGC as the reference, most DMRs observed in HGC were hyper-methylated (88.3%, 1105/1251; Figure 8B, Supplementary Table 13). In addition, there were substantial overlaps between genes with DMRs in LGC compared to normal tissues and those compared to HGC, suggesting a similar epigenetic process (Figure 8D).

The meDEG methylation patterns between ADE of different grades (Figure 8E) and CRC (Figure 8F) of different grades were similar. NECAB1 was methylated in LGA and HGA compared to the status in normal tissues (Figure 8G). AGTR1 was hyper-methylated in LGA, but not in HGA, compared to the status in normal tissues (Figure 8G). In CRC, NECAB1 was methylated in LGC and HGC compared to the status in normal tissues (Figure 8H). AGTR1 was hyper-methylated in HGC, but not in LGC, compared to the status in normal tissues (Figure 8G).

To assess the reproducibility of our study, we evaluated the Ten-Gene Methylation Signature developed by Patai *et al*^[22]. In ADE, 8/10 methylation signatures were hyper-methylated in LGA and HGA compared to the status in normal tissues (Supplementary Table 14). In CRC, 8/10 methylation signatures were hyper-methylated in LGC and HGC compared to the status in normal tissues (Supplementary Table 15). Our results were thus highly consistent with the findings previously reported^[22].

Finally, we utilized the dataset GSE32323 to investigate the effects of restoring the methylation status on *AGTR1* and *NECAB1* expression, based on 5-aza-2'-deoxycytidine treatment. In the HT29 cell line, demethylation treatment partially restored the mRNA expression levels of *AGTR1* and *NECAB1*, indicating that DNA hypermethylation seems to play a role in the regulation of these genes (Supplementary Figure 1). A similar tendency was also observed in the HCT116 cell line (Supplementary Figure 1).

DISCUSSION

We profiled the global methylation status of 22 CRC and 25 ADE samples and found that the methylation level in CRC was higher than that in ADE. Moreover, these sample types showed similar methylation patterns with respect to the DMP gene locations, chromatin signatures, and repeated elements. Combined with RNA-seq gene expression data, we identified 14 meDEGs, of which only *AGTR1* and *NECAB1* methylation patterns had prognostic significance.

In our study, we found that hypermethylation at CpG sites is more prevalent in CRC and ADE than in normal tissues, which is consistent with the results of previous studies^[23,24]. McInnes *et al*^[23] found that 73% of differentially methylated CpGs were hypermethylated in CRC at the CpG-dinucleotide level. Further, Kibriya *et al*^[24] found that 600 of 875 autosomal loci were hypermethylated.

However, some other studies have reported contradictory results^[4,6,10]. Gu *et al*^[4] found that 87% of methylated CpG sites are hypomethylated at the CpG-dinucleotide level. These discrepancies could be attributed to differences in detection methods and sample sizes. The overall methylation level in ADE is lower than that in CRC, indicating

that aberrant methylation occurs at the earliest stages of tumor initiation^[6]. Luo *et al*^[10] reported that nearly 40% DMPs were hypermethylated, and 60% of DMPs were hypomethylated in ADEs, compared to the status in the normal colon mucosa. Moreover, genome-wide hypermethylation appears to continuously increase during ADE-cancer sequencing^[25].

The methylation patterns observed in the ADE and CRC samples were similar. Hypermethylated DMPs were mainly located at CpG islands, whereas hypomethylated DMPs were largely located in the open sea region. These findings were consistent with those reported by Naumov *et al*^[7]. Both CRC and ADE showed hypermethylation enrichment in the bivalent enhancer state regions (dominant H3K27me3) and heterochromatin state regions (dominant H3K9me3). H3K27me3 is often associated with transcriptionally repressed chromatin^[26], which often occurs during tumorigenesis^[27]. Some methylation aberrations were enriched exclusively in the normal colorectum. Bormann *et al*^[28] reported that ADEs retain cellular methylation signatures, which is consistent with our results. We also observed the hypomethylation of short interspersed retrotransposable, LIN, LTR, and satellite regions in both ADE and CRC. LINE-1 expression is a hallmark of several types of malignancy^[29], and hypomethylation within repetitive elements has been implicated in CRC initiation^[30]. Pathway enrichment analysis of the DMR in promoters showed that these genes were enriched in the neuroactive ligand-receptor interaction pathway, which is consistent with previous results^[11]. Our study suggests that the methylation of molecules involved in the gut-brain axis could play a crucial role in the initiation and progression of CRC, even in the early stages of ADE.

Seven meDEGs were identified as follows: *NECAB1*, *PCSK2*, *AGTR1*, *NLGN1*, *LRAT*, *RXRG*, and *TMEFF2*. Among these genes, the methylation of *PCSK2*^[31], *AGTR1*^[32], *LRAT*^[33], *RXRG*^[34], and *TMEFF2*^[35] has been previously reported in CRC. *NECAB1* encodes a neuronal calcium-binding protein^[36], and its methylation has been observed in acute myeloid leukemia^[37]. *NLGN1* is involved in the regulation of glutamatergic transmission^[38], and studies have shown that upregulated *NLGN1* expression predicts

poor survival in CRC^[39]. Moreover, Pergolizzi *et al*^[40] found that *NLGN1* promotes CRC progression. In addition, *NLGN1* might act as a tumor suppressor gene during early CRC initiation but could promote CRC progression at later stages. Furthermore, we found that *AGTR1* and *NECAB1* methylation levels were associated with survival. Previous studies have shown that *AGTR1* expression serves as a prognostic biomarker in various cancers^[41-43], whereas the prognostic value of *AGTR1* methylation has not been reported. There are still some limitations of our study: (1) The sample size of the present study was relatively limited hampering the statistical power of the study; and (2) Not all samples were subjected to both RNA-Seq and WGBS analysis due to the availability of RNA or DNA.

CONCLUSION

In summary, in this study, the global methylome patterns were characterized and genome-wide similarities in methylation patterns between ADE and CRC tissues were identified. Our results support the potential use of *AGTR1* and *NECAB1* methylation statuses as prognostic biomarkers for CRC.

ARTICLE HIGHLIGHTS

Research background

Aberrant methylation is a common occurrence in colorectal cancer (CRC) initiation and progression, with clinical implications for early detection during CRC formation.

Research motivation

Addressing the critical need to identify specific DNA methylation markers for both adenoma (ADE) and CRC, the aim of this study was to unravel key molecular signatures associated with ADE and CRC.

Research objectives

Utilizing SeqCap targeted bisulfite sequencing and RNA-seq analysis, the goal of this study was to comprehensively profile the epigenomic-transcriptomic landscapes of colorectal ADE and CRC samples.

Research methods

The research involved a detailed examination, through SeqCap targeted bisulfite sequencing and RNA-seq analysis, shedding light on the distinctive epigenomic and transcriptomic characteristics of colorectal ADE and CRC. Public The Cancer Genome Atlas datasets were mined to explore the clinical implications of those methylation makers.

Research results

The comparative analysis of 22 CRC and 25 ADE samples revealed higher global methylation in CRC, with both exhibiting similar methylation patterns in differentially methylated positions, gene locations, chromatin signatures, and repeated elements. Integration with RNA-seq data identified 14 methylation-regulated differentially expressed genes, highlighting the prognostic significance of *AGTR1* and *NECAB1* methylation.

Research conclusions

This study revealed genome-wide alterations in DNA methylation during early CRC stages. It further identified distinctive methylation signatures associated with colorectal ADEs and CRC. Notably, *AGTR1* and *NECAB1* methylation emerged as potential prognostic biomarkers for CRC.

Research perspectives

The findings provide a foundation for future investigations into the clinical utility of identified methylation markers and the development of new targets for CRC treatment.

ORIGINALITY REPORT

2%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet	42 words — 1%
2	bmcbioinformatics.biomedcentral.com Internet	17 words — < 1%
3	www.frontiersin.org Internet	17 words — < 1%
4	wjso.biomedcentral.com Internet	15 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	< 15 WORDS
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 15 WORDS