Identifying Epigenetic Biomarkers in Colorectal Cancer: A Bioinformatics Analysis

Olajumoke Bisola Oladapo[1*†], Ferial NajianTabri[2†], and Ujwala Vasireddy[2†]

[1]Department of Biomedical Engineering, University of Oklahoma.

[2]Department of Computer Science, University of Oklahoma.

[†]These authors contributed equally to this work.

December 9th, 2024

Abstract

Colorectal cancer (CRC) is a term that refers to the combination of colon and rectal cancer as they are being treated as a single tumor. In CRC, 72% of tumors are colon cancer while the other 28% represent rectal cancer. CRC is a multifactorial disease caused by both genetic and epigenetic changes in the colon mucosal cells, affecting the oncogenes, DNA repair genes, and tumor suppressor genes. Currently, two DNA methylation based biomarkers for CRC have received FDA approval: SEPT9, used in blood-based screening tests, and a combination of NDRG4 and BMP3 for stool-based tests. Although DNA methylation biomarkers have been explored in colorectal cancer (CRC), the identification of robust and clinically valuable biomarkers remains a challenge, particularly for early-stage detection and precancerous lesions. Patients often receive diagnoses at the locally advanced stage, which limits the potential utility of current biomarkers in clinical settings. This study aims to address this gap by employing a bioinformatics pipeline to identify novel DNA methylation-regulated genes associated with CRC. Datasets used in this study were retrieved from the Gene Expression Omnibus (GEO) database. The Limma R package was used to identify differentially methylated CPG sites (DMCs) and identify differentially expressed genes (DEGs). From the overlap of DMCs with DEGs in rectal cancer, we identified seven MRGs (GNG7, ZCCHC14, HKDC1, AZGP1, ALG1L, PITX2, and PDX1) as biomarkers for CRC. The KEGG pathway enrichment analysis revealed the involvement of

these methylation-regulated genes in two distinct pathways, namely Type II Diabetes Mellitus, and the biosynthesis of Neomycin, Kanamycin, and Gentamicin, which are related to human disease, and metabolism, respectively. Also, through gene ontology, we identified that the MRGs were involved in in endocrine system development, intracellular glucose homeostasis, glucose metabolic process, metabolic process, and some other biological processes.

1 INTRODUCTION

Colorectal cancer (CRC) is a term that refers to the combination of colon and rectal cancer as they are being treated as a single tumor. In CRC, 72% of tumors are colon cancer while the other 28% represents the rectal cancer [1]. CRC thrives as the third leading type of cancer and second in mortality, with a prediction of over 60% incidence rate in 2030 [1, 2]. There is a 5-year survival rate for 90% of patients diagnosed with CRC in early stages and localized stages but only 13.1% in later cancer stages and in cases of CRC metastasis [3]. Early detection is critical in the survival of patients diagnosed with CRC, and biomarkers play a very important role in its diagnoses and survival, and only a few biomarkers have been translated to clinical practice; hence the need to identify biomarkers in CRC [4]. Currently, microRNAs, DNA mutations and methylation, proteins, which are all epigenetic functions, and the gut microbiomes are areas that have been explored in identifying CRC biomarkers [5].

Epigenetics is a term that refers to the alteration of gene expression without modifying the underlying DNA sequences. Epigenetics is an important biological process that plays a major role in cellular functions such as cell differentiation, embryogenesis, and cellular metabolism, among others, and has significant importance in diseases such as cancer, diabetes, and obesity, among others [6],[7]. DNA methylation is the addition of a methyl group $-CH_3$ and carbon-5 of a cytosine that is adjacent to a guanine in a CpG site, majorly leading to the in-expression of the gene. In many diseases, abnormal methylation patterns have been seen at the CpGs of some very important genes. These genes can be either hypermethylated or hypomethylated, which changes how they are translated and expressed [7]. Hypermethylation has been linked to slowed down transcription in abnormal tissues, while unmethylation has been linked to increased gene expression in normal tissues[8]. Figure 1 below shows the different types of epigenetic modification and their downstream effects in gene expression and suppression.

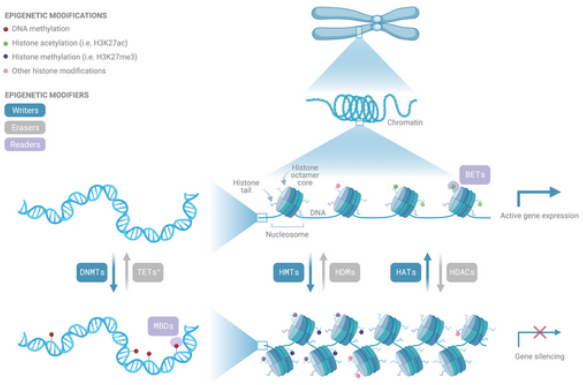


Figure 1: Epigenetic Modifications [6]

Many cancers exhibit global and focal alterations at CpG-rich sites, characterized by hypomethylation and hypermethylation, respectively. Brain tumors, breast cancer, and gastric cancer have been characterized with hypomethylation, while in cancers like glioblastoma, esophageal adenocarcinoma, and colorectal cancer [9]. Figure 2 below shows how DNA methylation is involved in oncogenesis, highlighting the difference in expression of oncogenes between normal and tumor cells.

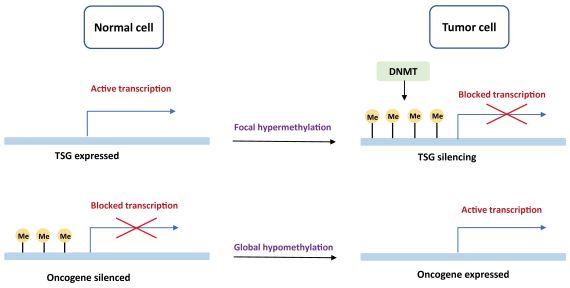


Figure 2: DNA Methylation in Oncogenesis[9]

CRC is a multifactorial disease caused by both genetic and epigenetic changes in the colon mucosal cells, affecting the oncogenes, DNA repair genes, and tumor suppressor genes. Low folate intake has been linked with increased cancer risks from impaired DNA methylation, synthesis, and repair in polymorphic individuals. In CRC, MTHFR polymorphism can be regulated by sufficient folate intake, hence reducing cancer risks. Also, in CRC cells with activated KRAS and SLC25A22 genes, there is an accumulation of succinic acid, which increases DNA methylation. In 80% of CRC cases, there is hypermethylation of the RASSF1 promoter, and in approximately 30% of CRC cases, p16INK4a is methylated, and its mutation leads to BRAF mutation [10]. DNA methylation patterns in normal cells and tumor-specific cells show significantly different profiles and can be used to identify

DNA from tumor samples, making it a promising biomarker[11]. Currently, two DNA methylation-based biomarkers for CRC have received FDA approval: SEPT9, used in blood-based screening tests, and a combination of NDRG4 and BMP3 for stool-based tests [12]. Recently, many studies have published promising DNA methylation biomarkers for CRC. Shen *et al.* [12] identified two candidate CpG site biomarkers for CRC: cg13096260 and cg12993163 from 76 pairs of CRC and adjacent normal tissue samples, 348 stool samples, and 136 blood samples. Similarly, through the Stool ColoDefense test, Zhao *et al.* [13] identified DNA methylation of SEPT9 and SDC2 as an integrated biomarker for CRC. Although DNA methylation biomarkers have been explored in colorectal cancer (CRC), the identification of robust and clinically valuable biomarkers remains a challenge, particularly for early-stage detection and precancerous lesions. Current biomarkers often lack the sensitivity and specificity needed for early diagnosis; patients are often diagnosed at the locally advanced stage, limiting their potential utility in clinical settings [4, 14]. This study aims to address this gap by employing a bioinformatics pipeline to identify novel DNA methylation-regulated genes associated with CRC. Following the approach used by Li *et al.* [15] to identify methylation-regulated genes in varicose vein samples, this study analyzes publicly available methylation and expression datasets of CRC to identify candidate biomarkers that exhibit consistent epigenetic alterations in CRC samples. The ultimate goal is to identify biomarkers that can be further validated for their diagnostic and prognostic potential in CRC and precancerous lesions.

2 MATERIALS AND METHODS

2.1 Experimental Design

We aim to identify DNA methylation biomarkers through methylation-regulated genes in colorectal cancer (CRC) following a bioinformatics pipeline adopted from Li *et al.* [15] study on the identification of biomarkers in varicose vein disease. All analyses were carried out on R. [16].

2.2 Data Collection

The datasets used in this study were retrieved from GEO Database using the GEOquery package [17]. Specifically, the GSE75548 and GSE75546 datasets, which represent expression profiling by array and methylation profiling by genome tiling array, respectively, were collected from six tissue

samples of patients with rectal cancer with paired normal tissues.

2.3 Identifying and Mapping Differentially Methylated CpG Sites

Differentially methylated CpG sites (DMCs) between the normal tissues and rectal cancer tissue samples were identified using the Limma package[18]. The results were considered statistically significant if $P < 0.05$. and $\log_2 \mathrm{FC} > 0.4$.We identified differentially methylated regions (DMRs) between normal and rectal cancer samples. A design matrix was constructed using metadata to group samples, and CpG sites were annotated using the DMRcate package [19] with an FDR threshold of 0.001. DMRs were classified as hypermethylated or hypomethylated based on mean methylation differences. Genomic coordinates were validated using the BSgenome.Hsapiens.UCSC.hg19 package[20], ensuring all regions were within standard chromosomes. A karyogram visualizing hypermethylated (red) and hypomethylated (blue) regions was generated using karyoploteR[21].

2.4 Identification of Differentially Expressed Genes (DEGs)

The differentially expressed genes (DEGs) were identified using the Limma package[18]. The results were considered statistically significant if $P < 0.05$. and $\log_2 \mathrm{FC} > 0.5$. Results were visualized using a volcano plot to highlight upregulated, downregulated, and non-significant genes.

2.5 Identification and Analysis of Methylation-Regulated Genes (MRGs)

Gene annotations were linked to genomic regions using the methylKit package[22]. A gene annotation BED file was utilized for transcript features, ensuring accurate identification of overlapping and nearby genes. Gene symbols from the annotated DMRs were compared with significantly differentially expressed genes (DEGs). This integration identified common genes: methylation-regulated genes (MRGs) that showed both methylation alterations and differential expression patterns. The overlapping genes were visualized using a Venn diagram created with the VennDiagram package[23].

2.6 KEGG Pathway and Gene Ontology (GO) Enrichment Analysis

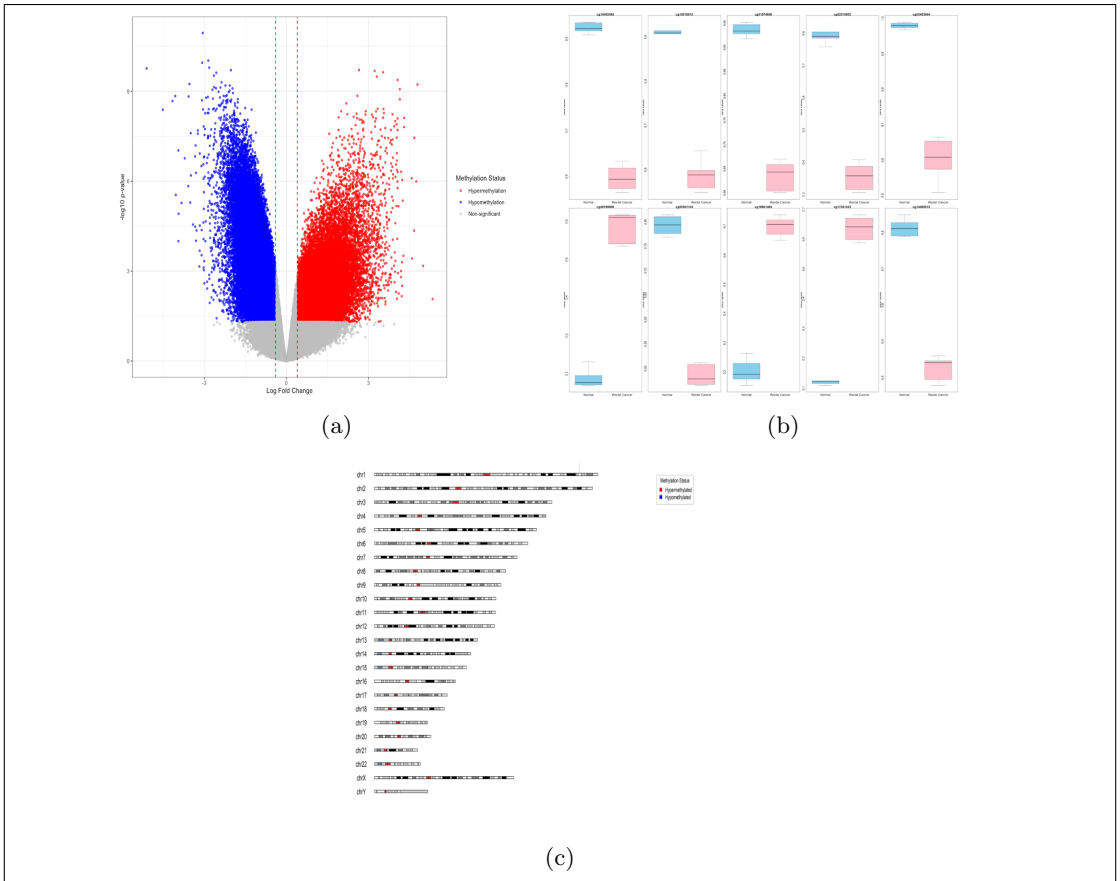
Entrez gene IDs for the identified MRGs were retrieved using the org.Hs.eg.db package [24]. These IDs were subjected to pathway and functional enrichment analysis using the clusterProfiler package[25]. KEGG pathway analysis for enrichment of biological pathways was performed with signif-

icance determined by a p-value cutoff of 0.05; results were visualized as dot plots. Similarly, Gene Ontology analysis focused on biological processes, identifying functional categories enriched in the MRGs; results were visualized in high-resolution dot plots.

3 RESULTS

3.1 Differentially Methylated CpG sites (DMCs)

Figure3a below shows the differentially methylated CpG sites (DMCs) in a volcano plot. The volcano plot highlights the hypomethylated CpG sites on the left wing in blue, the hypermethylated CpG sites on the right wing in red, and the ash color represents the CpG sites that do not significantly differ in methylation between cancer and normal tissues. The boxplot in figure3b is a representation of the top ten DMCs in the normal and cancer tissues. Figure 3c is a karyogram showing differentially methylated regions (DMRs).



(a) (b)

(c)

Figure 3: The DMCs from the GSE75546 dataset.(a) Volcano plot showing the DMCs. (b) Boxplot showing the top ten DMCs. (c) Karyogram showing the DMRs

3.2 Differentially Expressed Genes (DEGs)

Figure 4a below represents a heatmap depicting the clustering of samples based on their gene expression profiles. Figure 4b shows the results of Principal Component Analysis (PCA). This scatter plot illustrates the separation of samples along the first two principal components (PC1 and PC2). Group-specific coloring of the points confirms the distinct global expression profiles of rectal cancer and normal samples. Figure4c is a volcano plot visualizing the DEGs between rectal cancer and normal tissues. Each point represents a gene, with the x-axis showing the log2 fold change (log2FC) and the y-axis representing the -log10 of the p-value. Genes with significant upregulation (red points) and downregulation (blue points) are highlighted. Figure4d is a heatmap displaying the expression profiles of the top DEGs. Each row represents a gene, and each column represents a sample.

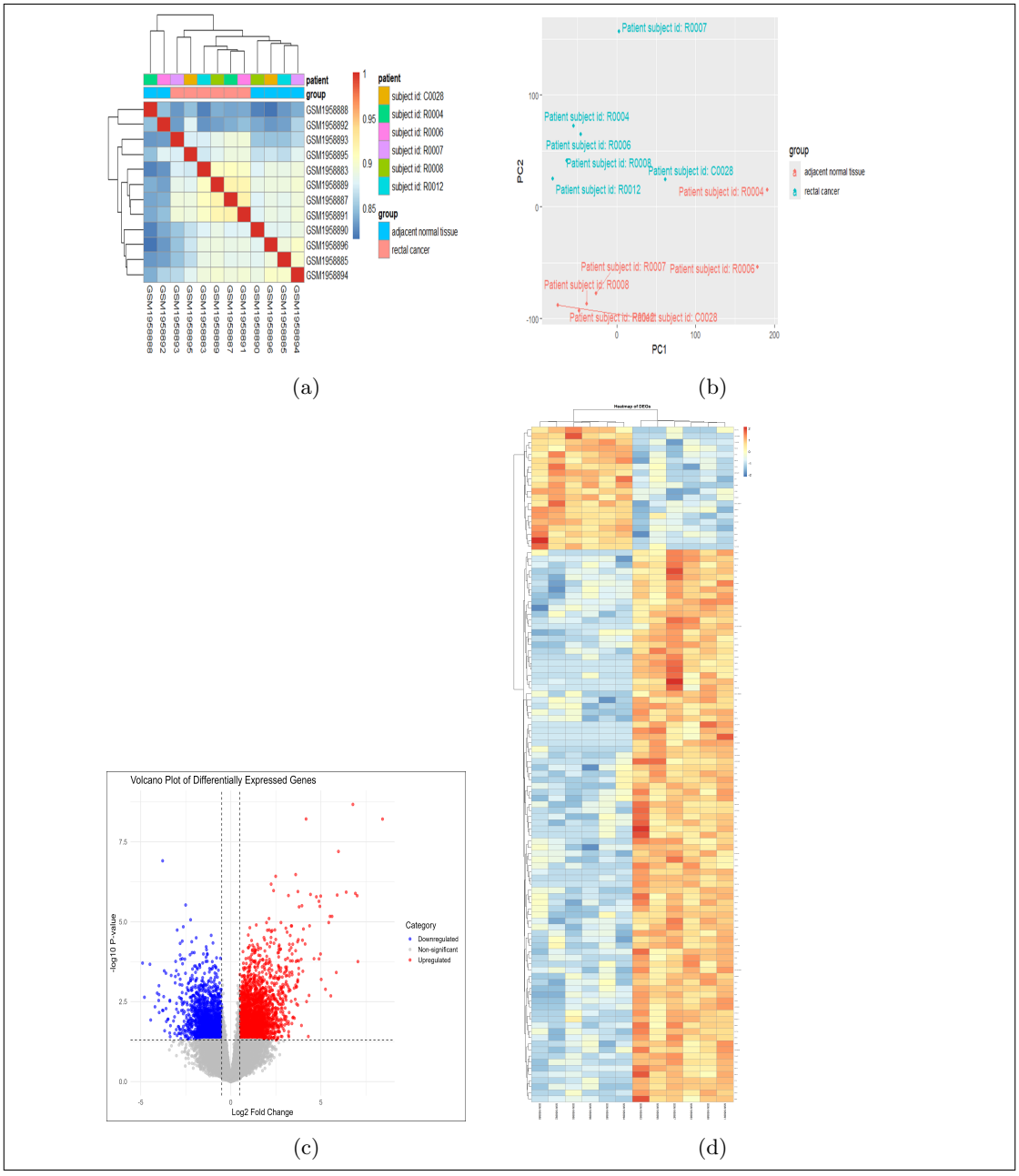


Figure 4: Heatmap showing expression profiles of samples. (a) Heatmap showing expression profiles of samples. (b) Principal Component Analysis (PCA) of Gene Expression Profiles. (c) Volcano plot of DEGs. (d) Heatmap showing DEGs among samples.

3.3 Methylation Regulated Genes (MRGs)

Out of the 776 differentially methylated genes, seven were found to overlap with the 110 differentially expressed genes filtered based on an adjusted P-value set to 0.05 (Figure 5). These eight overlapping

genes are considered MRGs and represent candidates for further functional and pathway analysis. The identified MRGs, including their log2 fold change, average expression, and adjusted p-values, are summarized in Table1.

Table 1: Methylation Regulated Genes

S/N	Gene_ID	logFC	AveExpr	t	P.Value	adj.P.Val	B	Category	Symbol
1	ILMN_1728107	-2.026332817	4.841524142	-5.95184717	9.20067667716203e-05	0.044555626	1.721789137	Downregulated	GNG7
2	ILMN_1743456	0.965532	6.99894855	6.197012885	6.4695045483695e-05	0.038755022	2.035504649	Upregulated	ZCCHC14
3	ILMN_1752502	3.600531267	5.90876605	6.044561274	8.04560154945336e-05	0.041503233	1.841603933	Upregulated	HKDC1
4	ILMN_1797154	3.602733083	2.103766309	10.73251601	3.33757378247011e-07	0.002551628	6.294997674	Upregulated	AZGP1
5	ILMN_2131293	3.058635319	1.815741357	7.235632876	1.5891225794823e-05	0.020284935	3.256655865	Upregulated	ALG1L
6	ILMN_2391400	3.707354795	2.942343519	6.502738955	4.21783649682205e-05	0.033764139	2.412817715	Upregulated	PITX2
7	ILMN_3249216	4.37360105	3.777353792	7.3696865	1.3388867845757e-05	0.018719227	3.402093842	Upregulated	PDX1

The Venn diagram in Figure 5 is a representation of the MRGs that shows the overlap between differentially methylated CpG sites and differentially expressed genes (DEGs).

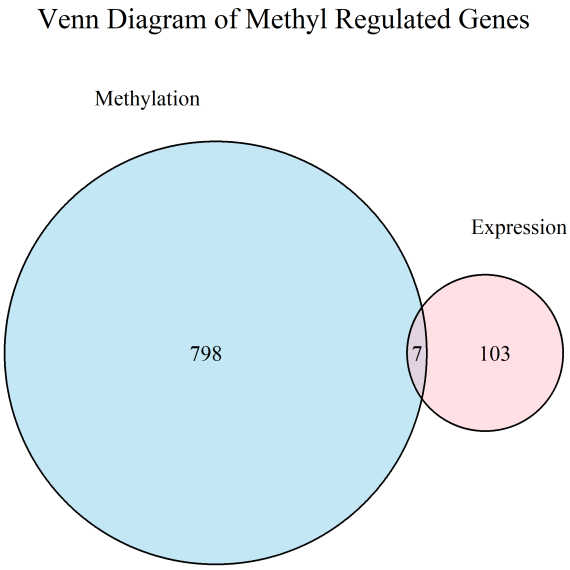
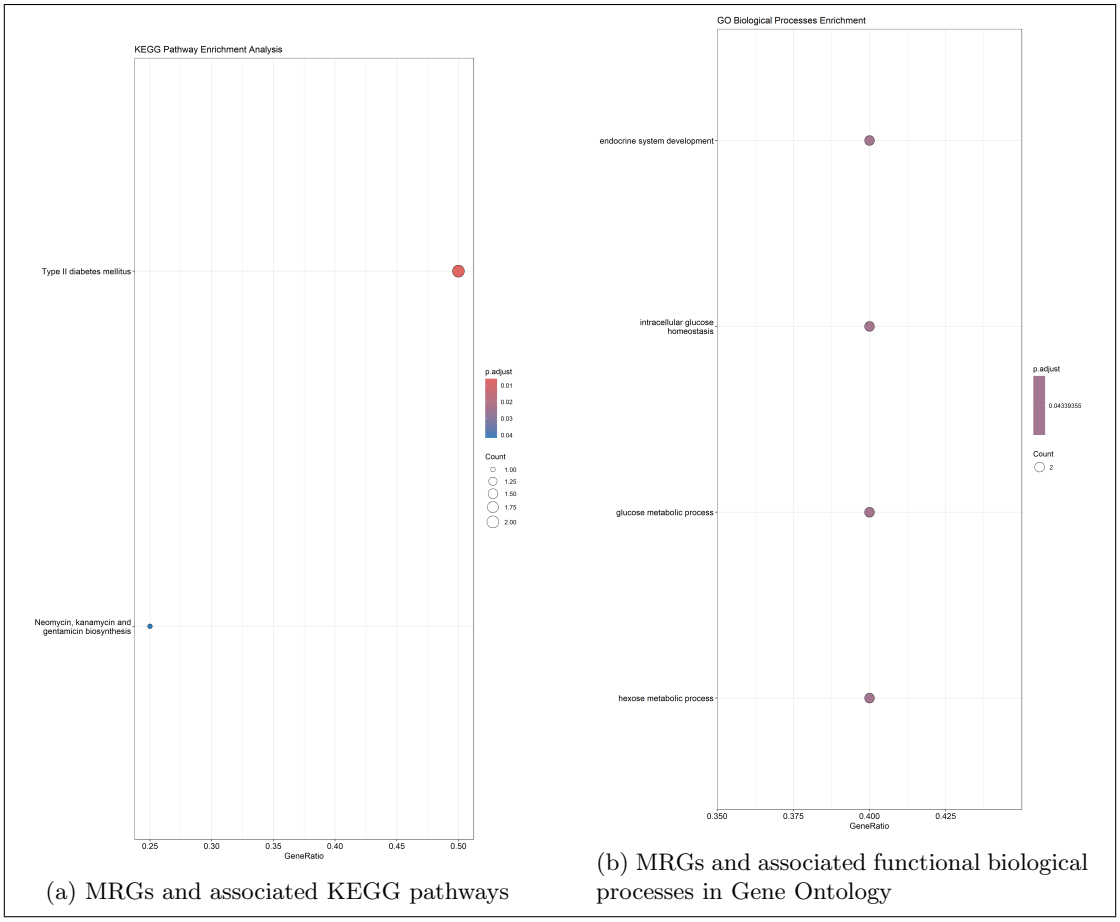


Figure 5: Venn Diagram showing MRGs

3.4 KEGG Pathway and Gene Ontology

Figure6a illustrates the KEGG pathway enrichment analysis, showing the top enriched pathways associated with MRGs. Figure 6b displays the GO enrichment analysis results, highlighting the top biological processes associated with MRGs.



(a) MRGs and associated KEGG pathways

(b) MRGs and associated functional biological processes in Gene Ontology

Figure 6: KEGG pathways and GO biological processes associated with MRGs.

4 DISCUSSION AND CONCLUSION

CRC develops when the normal colon and rectal epithelium is transformed into a precancerous lesion and eventually into an advanced carcinoma, which can metastasize to different organs [1]. Risks of developing CRC can be linked to age, environmental, behavioral, and genetic determining factors[26]. Raut *et al.* [27] identified two fecal DNA methylation markers to detect stages in CRC. Bach et al. [28], through urine-based DNA methylation analysis, identified SEPT9 and SDC2 as key markers for non-invasive colorectal cancer (CRC) detection. DNA methylation has been extensively studied in CRC; Huang *et al.*[29] identified specific tumor clusters with methylated CpG islands that were associated with metabolic pathways, increased ATP production, and tumor aggressiveness in CRC. In this research, we analyzed datasets from a publicly available dataset on rectal cancer samples and

carried out differential methylation and expression analysis on these datasets. A total of 384,933 CpG sites were screened with 98,209 differentially methylated CpG sites, of which 30,613 were hypermethylated and 67,596 were hypomethylated. DMCs were annotated to assess their genomic distribution relative to genic regions. Among the 874 differentially methylated regions (DMRs), the majority (95.77%) were located in intergenic regions, with smaller proportions overlapping with introns (3.89%), exons (0.80%), and promoters (0.57%). When prioritizing annotations (promoter > exon > intron), the overlap percentages were consistent, with 95.77% in intergenic regions and slight redistribution among genic parts. There were 0.40% of promoter boundaries, 0.07% of exon boundaries, and 0.31% of intron boundaries overlapped with DMRs. The distances to the nearest transcription start site (TSS) showed a median distance of 27,990 bp and a mean distance of 68,348 bp, with values ranging from 0 to 565,929 bp. These results indicate that the majority of DMRs are in intergenic regions, suggesting their potential involvement in distal regulatory functions. The smaller fraction of DMRs overlapping with promoters, exons, and introns points to possible roles in gene regulation, transcription initiation, and splicing. Diez-Villanueva et al. [30] explored the role of DNA methylation in the malignant progression of colon cancer and found distinct patterns of correlation between CpG methylation and gene expression depending on their genomic context. CpGs located in promoter regions were predominantly negatively correlated with gene expression, suggesting a repressive effect on transcription. In contrast, CpGs within gene bodies and intergenic regions exhibited predominantly positive correlations, indicating potential roles in enhancing transcription or other regulatory functions. Findings from Liu *et al.*[31] showed 411 upregulatedgenes thath were significantlyhypomethylated andd 239 downregulated genesthath were hypermethylated. The hub genes identified from, this study includes CAD, CCND1, ATM, RB1, MET, EGFR, ACTA1, SST, ESR1 and DNM2 and were suggested to be important biomarkers for CRC. Similarly, Sun *et al.*[32]identified PPBP,, CCL28, CXCL12, INSL5, CXCL3, CXCL10, andCXCL11 ass hub genesthath were differentially expressed in CRC analysis and suggested these hub genes as biomarkers of CRC. In this present study, a total of 4,414 differentially expressed genes (DEGs) were identified, of which 2,518 were upregulated and 1,896 weredownregulated. We compared the gene symbols from the annotated DMRs with the significantly downregulated DEGs. This integration identified common genes regarded asmethylation-regulated.dWe identified genes (MRGs) that exhibited both methylation alterations and differential expression patterns. Our analysis identifies these genes, GNG7, ZCCHC14, HKDC1, AZGP1, ALG1L, PITX2, PDX1 as methylation-regulated genes.

GNG7 (G Protein Subunit Gamma 7), a component of heterotrimeric G proteins, is highly enriched in the striatum and plays a crucial role in the neuroprotective response mediated by A2A adenosine and D1 dopamine receptors. Previous studies have reported GNG7 downregulation in various cancers, including pancreatic, gastrointestinal tract, renal, and lung cancers [33]. In our study, we also identified GNG7 as being downregulated in colorectal cancer as a methylation-regulated gene. HKDC1 (Hexokinase domain component 1) is a member of the hexokinase family genes[34].HKDC1 has been shown to be an active oncogene in previous studies in cancers such as breast, lymphoma, liver, pancreatic, and lung cancer[35, 36],[37]. Other studies also confirm HKDC1 as an oncogene in CRC [38, 39],[40]. In this present study, HKDC1 was significantly upregulated in colorectal cancer samples as a methylation-regulated gene. AZGP1 (Alpha-2-Glycoprotein 1, Zinc-Binding) codes for the Zinc-α_2-Glycoprotein (ZAG)[41]. Fang *et al.* [42] found AZGP1 to be upregulated in cancerous colorectal tissue and identified it as a common target of colorectal cancer initiation. AZGP1 has also been found to be linked in pathways involved in CRC progression, such as regulating the pentose phosphate pathway through its interaction with circNOLC1, making it a promising target for CRC liver metastasis[43]. Our present study identified AZGP1 to be significantly upregulated in colorectal cancer samples as a methylation-regulated gene. Lin *et al.*[44] found that ALG1L was differentially expressed in hepatocellular carcinoma. Akbulut*et al.*[45] ALG1L was among the top ten genes that were upregulated in hepatocellular carcinoma tissues. ALG1L was upregulated in colorectal cancer samples as a methylation-regulated gene. Bioinformatics analysis identified a high level of expression of PITX2 in stromal cells in pancreatic cancer [46]. In our current study, there is also an upregulation of PITX2 in colorectal cancer samples. A previous study suggested that KCNQ1 might be a potential prognostic biomarker for CRC with liver metastasis when there is a low level of expression of KCNQ1. This is contrary to our research findings suggesting upregulation of KCNQ1 as a biomarker in CRC [47]. A study highlighted that the hypermethylation of PDX1 could serve as a biomarker for CRC[48]. This is in line with our current study that methylation could influence the expression pattern of PDX1 by upregulating it; thus we identify it as a methylation-regulated gene.

Our KEGG pathway enrichment analysis uncovered the involvement of these methylation-regulated genes in two distinct pathways: Type II Diabetes Mellitus, and the biosynthesis of Neomycin, Kanamycin, and Gentamicin. These pathways are associated with human disease, and metabolism, respectively. This is in line with the studies of Sun*et al.*[49], where PTPN11 was found to be a

possible biomarker for CRC in people with diabetes. From Gene Ontology, we identified that the MRGs were involved in endocrine system development, intracellular glucose homeostasis, glucose metabolic proce, metabolic process, hexose metabolic process, stem cell differentiation, monosaccharide metabolic process, glucose homeostasis, carbohydrate homeostasis, cardiac neural crest cell migration involved in outflow tract morphogenesis, carbohydrate mediated signaling, type B pancreatic cell apoptotic process, receptor guanylyl cyclase signaling pathway, cardiac neural crest cell, development involved in outflow tract morphogenesis, regulation of type B, pancreatic cell proliferation, left/right axis specification, response to leucine, cardiac neural crest cell differentiation involved in heart development, cardiac neural crest cell development involved in heart development, adenohypophysis development, negative regulation of endoplasmic, reticulum stress-induced intrinsic apoptotic signaling pathway, gland development, and cell migration involved in heart development. This is in line with the studies of Li *et al.*[50], which identified overexpression of KDM4B, which increases glucose metabolism.

In conclusion, we identify seven genes as methylation-regulated genes through a comprehensive bioinformatics analysis, suggesting that methylation affects their expression levels. These genes have been associated with a variety of tumors in literature studies, with some specifically being linked to colorectal cancer (CRC). We suggest that these genes could serve as biomarkers for CRC, and further wet lab procedures are needed to validate their functions as biomarkers for CRC. Limitations of this study include the inability to carry out further computational validation procedures and wet lab procedures, and we look forward to continuing that as a future direction for this project.

Acknowledgments

General

We appreciate and acknowledge the guidance of our professor, Dr. Moussa, throughout this project and the Bioinformatics class members for their useful reviews during the project proposal stage.

Author Contributions

"Oladapo, O. conceptualized the idea for this project and came up with the experimental design, worked on annotating differentially methylated regions and karyogram plots, also the KEGG pathway

and Gene Gntology, and drafted the manuscript for this project."

"NajianTabri, F., worked on the data collection and identification of methylation-regulated genes. NajianTabri, F. also worked on weighted gene coexpression network analysis (WCGNA) to identify modules associated with CRC, which was not reported in this study due to incomplete results and inadequacy of results to complement the study." "Vasireddy, U. worked on the identification of differentially methylated CpG sites and differentially expressed genes analysis and generated respective plots for these analyses."

Funding

"The authors acknowledge that they did not receive funding for this work."

Conflicts of Interest

"The authors declare that there is no conflict of interest regarding the publication of this article."

Data Availability

The datasets used in this study are publicly available at the GEO Database.

Supplementary Materials

All tables generated and codes used for this analysis are provided in the supplementary materials.

References

1. Alzahrani SM, Al Doghaither HA, and Al-Ghafari AB. General insight into cancer: An overview of colorectal cancer. Molecular and clinical oncology 2021;15:271.
2. Hoang T, Kim H, and Kim J. Dietary intake in association with all-cause mortality and colorectal cancer mortality among colorectal cancer survivors: a systematic review and meta-analysis of prospective studies. Cancers 2020;12:3391.
3. Housini M, Dariya B, Ahmed N, et al. Colorectal cancer: Genetic alterations, novel biomarkers, current therapeutic strategies and clinical trials. Gene 2024;892:147857.

4. Ogunwobi OO, Mahmood F, and Akingboye A. Biomarkers in colorectal cancer: current research and future prospects. International journal of molecular sciences 2020;21:5311.

5. Zygulska AL and Pierzchalski P. Novel diagnostic biomarkers in colorectal cancer. International journal of molecular sciences 2022;23:852.

6. Davalos V and Esteller M. Cancer epigenetics in clinical practice. CA: a cancer journal for clinicians 2023;73:376–424.

7. Xu W, Wang F, Yu Z, and Xin F. Epigenetics and cellular metabolism. Genetics & epigenetics 2016;8:GEG–S32160.

8. Scala G, Federico A, Palumbo D, Cocozza S, and Greco D. DNA sequence context as a marker of CpG methylation instability in normal and cancer tissues. Scientific reports 2020;10:1721.

9. Chen C, Wang Z, Ding Y, et al. DNA methylation: from cancer biology to clinical perspectives. Frontiers in Bioscience-Landmark 2022;27:326.

10. Wang Y, Wang C, Zhong R, Wang L, and Sun L. Research progress of DNA methylation in colorectal cancer. Molecular Medicine Reports 2024;30:154.

11. Vrba L and Futscher BW. DNA methylation changes in biomarker loci occur early in cancer progression. F1000Research 2020;8:2106.

12. Shen Y, Wang D, Yuan T, et al. Novel DNA methylation biomarkers in stool and blood for early detection of colorectal cancer and precancerous lesions. Clinical Epigenetics 2023;15:26.

13. Zhao G, Liu X, Liu Y, et al. Aberrant DNA methylation of SEPT9 and SDC2 in stool specimens as an integrated biomarker for colorectal cancer early detection. Frontiers in Genetics 2020;11:643.

14. Nikolouzakis TK, Vassilopoulou L, Fragkiadaki P, et al. Improving diagnosis, prognosis and prediction by using biomarkers in CRC patients. Oncology reports 2018;39:2455–72.

15. Li S, Liu Y, Liu M, Wang L, and Li X. Comprehensive bioinformatics analysis reveals biomarkers of DNA methylation-related genes in varicose veins. Frontiers in Genetics 2022;13:1013803.

16. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria, 2024. URL: `https://www.R-project.org/`.

17. Davis S and Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 2007;23:1846–7.

18. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research 2015;43:e47–e47.

19. Peters TJ, Buckley MJ, Statham AL, et al. De novo identification of differentially methylated regions in the human genome. Epigenetics & chromatin 2015;8:1–16.

20. Team T. BSgenome. hsapiens. UCSC. hg19: Full genome sequences for homo sapiens (UCSC version hg19, based on GRCh37. p 13) 2020.

21. Gel B and Serra E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics 2017;33:3088–90.

22. Akalin A, Kormaksson M, Li S, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome biology 2012;13:1–9.

23. Chen H and Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC bioinformatics 2011;12:1–7.

24. Carlson M, Falcon S, Pages H, Li N, et al. org. Hs. eg. db: Genome wide annotation for Human. R package version 2019;3:3.

25. Yu G, Wang LG, Han Y, and He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology 2012;16:284–7.

26. Lewandowska A, Rudzki G, Lewandowski T, Stryjkowska-Gora A, and Rudzki S. Risk factors for the diagnosis of colorectal cancer. Cancer control 2022;29:10732748211056692.

27. Raut JR, Guan Z, Schrotz-King P, and Brenner H. Fecal DNA methylation markers for detecting stages of colorectal cancer and its precursors: a systematic review. Clinical Epigenetics 2020;12:1–19.

28. Bach S, Paulis I, Sluiter N, et al. Detection of colorectal cancer in urine using DNA methylation analysis. Scientific reports 2021;11:2363.

29. Huang H, Li Q, Tu X, et al. DNA hypomethylation patterns and their impact on the tumor microenvironment in colorectal cancer. Cellular Oncology 2024:1–15.

30. Diez-Villanueva A, Sanz-Pamplona R, Carreras-Torres R, et al. DNA methylation events in transcription factors and gene expression changes in colon cancer. Epigenomics 2020;12:1593–610.

31. Liu J, Li H, Sun L, Wang Z, Xing C, and Yuan Y. Aberrantly methylated-differentially expressed genes and pathways in colorectal cancer. Cancer Cell International 2017;17:1–10.

32. Sun G, Li Y, Peng Y, et al. Identification of differentially expressed genes and biological characteristics of colorectal cancer by integrated bioinformatics analysis. Journal of cellular physiology 2019;234:15215–24.

33. Wei Q, Miao T, Zhang P, Jiang B, and Yan H. Comprehensive analysis to identify GNG7 as a prognostic biomarker in lung adenocarcinoma correlating with immune infiltrates. Frontiers in Genetics 2022;13:984575.

34. Irwin DM and Tan H. Molecular evolution of the vertebrate hexokinase gene family: Identification of a conserved fifth vertebrate hexokinase gene. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics 2008;3:96–107.

35. Wang X, Shi B, Zhao Y, et al. HKDC1 promotes the tumorigenesis and glycolysis in lung adenocarcinoma via regulating AMPK/mTOR signaling pathway. Cancer cell international 2020;20:1–12.

36. Pang Q, Huang S, Wang H, and Cao J. HKDC1 promotes autophagy and proliferation in pancreatic adenocarcinoma through interaction with PARP1 and poly (ADP-ribosyl) ation. Cellular Signalling 2024;124:111474.

37. Liu P, Luo Y, Wu H, et al. HKDC1 functions as a glucose sensor and promotes metabolic adaptation and cancer growth via interaction with PHB2. Cell Death & Differentiation 2024:1–16.

38. Lian H, Wang A, Shen Y, et al. Identification of novel alternative splicing isoform biomarkers and their association with overall survival in colorectal cancer. BMC gastroenterology 2020;20:1–12.

39. Yu C, Bao Tt, Jin L, Lu Jw, and Feng Jf. HKDC1 Silencing Inhibits Proliferation and Glycolysis of Gastric Cancer Cells. Journal of Oncology 2023;2023:3876342.

40. Chen HY, Li XN, Yang L, Ye CX, Chen ZL, and Wang ZJ. CircVMP1 promotes glycolysis and disease progression by upregulating HKDC1 in colorectal cancer. Environmental Toxicology 2024;39:1617–30.

41. Wei X, Liu X, Tan C, et al. Expression and function of zinc-α2-glycoprotein. Neuroscience bulletin 2019;35:540–50.

42. Fang YY, Huang JM, Wen JY, et al. AZGP1 up-regulation is a potential target for andrographolide reversing Radioresistance of colorectal cancer. Pharmacogenomics and Personalized Medicine 2022:999–1017.

43. Yuan M, Zhang X, Yue F, et al. CircNOLC1 Promotes Colorectal Cancer Liver Metastasis by Interacting with AZGP1 and Sponging miR-212-5p to Regulate Reprogramming of the Oxidative Pentose Phosphate Pathway. Advanced Science 2023;10:2205229.

44. Lin K, Shann Y, Chau G, Hsu C, and Huang C. Identification of latent biomarkers in hepatocellular carcinoma by ultra-deep whole-transcriptome sequencing. Oncogene 2014;33:4786–94.

45. Akbulut S, Kucukakcali Z, Sahin TT, Colak C, and Yilmaz S. Role of Epigenetic Factors in Determining the Biological Behavior and Prognosis of Hepatocellular Carcinoma. Diagnostics 2024;14:1925.

46. Wu D, Chen W, Yang Y, et al. PITX2 in pancreatic stellate cells promotes EMT in pancreatic cancer cells via the Wnt/β-catenin pathway: PITX2 promotes pancreatic cancer progression via Wnt/β-catenin pathway. Acta Biochimica et Biophysica Sinica 2023;55:1393.

47. Than BL, Goos J, Sarver AL, et al. The role of KCNQ1 in mouse and human gastrointestinal cancers. Oncogene 2014;33:3861–8.

48. Lee Y, Dho SH, Lee J, et al. Hypermethylation of PDX1, EN2, and MSX1 predicts the prognosis of colorectal cancer. Experimental & molecular medicine 2022;54:156–68.

49. Sun M, Han Z, Luo Z, et al. PTPN11 is a potential biomarker for type 2 diabetes mellitus complicated with colorectal cancer. Scientific Reports 2024;14:25155.

50. Li H, Lan J, Wang G, et al. KDM4B facilitates colorectal cancer growth and glucose metabolism by stimulating TRAF6-mediated AKT activation. Journal of Experimental & Clinical Cancer Research 2020;39:1–13.