Volume 3, Number 3, 2026

Pages 662-669

Bioscience Insights

ISSN: 3080-3489 (Print)
ISSN: 3080-3497 (Online)

BiosciInsights.com

INTEGRATED COMPUTATIONAL NETWORK ANALYSIS IDENTIFIES CANDIDATE BIOMARKERS IN COLORECTAL CANCER

Hasham Sohail[1], Sundas Rania[2], Humna Farooqi[3], Wajdan Ali[4], Fatima Maqbool[*5], Sabtain Ahmad[6], Nimra Allah Ditta[*7]

[1,2,3,5*]*Department of Life Sciences, University of Management and Technology, Lahore Pakistan*
[4]*Department of Anaesthesia, Shaikh Zayed Hospital, Lahore, Pakistan*
[6]*Department of Biological Sciences, The Superior University, Lahore, Pakistan*
[*7]*Department of Biotechnology, University of Management and Technology, Lahore, Pakistan*

[5]fatimamaqbool604@gmail.com, [7]allahdittanimra4@gmail.com

https://doi.org/10.5281/zenodo.23527392

Article Information

Article History

Received:	June 19, 2026
Revised:	August 19, 2026
Accepted:	September 07, 2026
Available Online:	October 19, 2026

Keywords:
Colorectal cancer, biomarkers, differential expression, protein-protein interaction, Cytoscape, cytoHubba, miRNA, GEO, bioinformatics.

Copyright @Author

Corresponding Author: *
Fatima Maqbool & Nimra Allah Ditta

Abstract

Colorectal cancer (CRC) is one of the leading causes of cancer morbidity and mortality and the molecular biomarkers may be useful for early detection and biological stratification. In this study, a comprehensive bioinformatics pipeline was employed to determine differentially expressed genes (DEGs), hub protein interactions, enriched functions and potential microRNA/transcription factor regulatory networks in CRC. Gene-expression data sets for the CRC were obtained from the publicly available NCBI Gene Expression Omnibus (GEO) database. The study thresholds were p value < 0.05 and log2 fold change > 2 (|log2 fold change| >= 2) for differential expression analysis using GEO2R/Limma. The overlapping DEGs were analyzed with STRING for protein-protein interaction (PPI) and Cytoscape/cytoHubba for network centrality. The functional enrichment was performed using DAVID for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Functional and mRNA-miRNA-transcription factor relationships were explored using GeneMANIA and miRNet, whereas GEPIA was used for evaluating the expression of tumor versus normal. Three GEO datasets were kept for analysis. GSE130750 resulted in 63 upregulated genes and 433 downregulated genes, while GSE160988 resulted in 134 upregulated and 606 downregulated genes. The third data set provided a smaller set of differential genes, 16 of them down and 0 up, reported in the thesis. A list of highly-connected hub genes was obtained via network analysis. The genes that were reported were extracellular-matrix/collagen: COL4A1, COL4A2, COL5A1, COL5A2, COL6A1, COL6A3, COL12A1 and immune-inflammatory: CXCL10, CCL18, IDO1, ITGAX, NLRP3, P4HA3 and TNFRSF1B. The regulatory-network analysis linked hub genes to the candidate miRNAs and transcription factors. The integrated analysis provides valuable insight into candidate CRC-associated genes and regulatory relationships that could be

targeted for further validation as diagnostic, prognostic or therapeutic biomarkers. Before clinical interpretation, independent cohorts, correction for multiple testing and laboratory/clinical validation are required.

INTRODUCTION

Coraclectal cancer (CRC) is a heterogeneous cancer disease of the colon or rectum that develops over time as a result of the accumulation of genetic and epigenetic changes. It is one of the most common cancers globally, and cancer deaths are an important cause of mortality in this cancer [1, 2, 3, 4, 5]. The mutual interactions between inherited susceptibility, somatic mutations, ageing, lifestyle, diet, obesity, physical inactivity, smoking and alcohol exposure contribute to CRC development [2,3,9,10]. While patterns differ by country, some Asian countries have seen rising CRC burden due to demographic and lifestyle changes [5,7,8,34].

Prognosis is closely linked to stage at diagnosis and so early detection is vita [48]. There are a number of established methods used, such as fecal occult blood testing, fecal immunochemical testing, stool-based molecular testing, endoscopy, imaging, or serum markers, but none of these methods is either completely sensitive or specific, is invasive, available everywhere, cheap, or accepted by patients, which continues to stimulate research into biomarkers [19,20,23,24,37]. Molecular markers based on DNA, RNA, proteins, epigenetic modifications and immune markers may play a role in diagnosis, prognosis, monitoring for recurrence and in treatment decisions [1,18,22,25-27].

In the context of diseases, high-throughput transcriptomic datasets can be used to uncover expression patterns associated with the disease, and network-based bioinformatics can be applied to prioritize genes based on their central positions in molecular interaction networks [49, 50]. For CRC, similar strategies have been applied to identify differentially expressed genes, enriched pathways, hub genes and non-coding RNA regulatory networks, etc. [43, 52]. In the current study, a combination of differential-expression analysis, functional enrichment analysis, PPI networks, hub gene prioritization, GeneMANIA, miRNet and GEPIA is used to detect potential CRC-related biomarkers and relationships between them and their regulators. The analysis focuses on datasets that are particularly relevant to the Asian population as stated in the source thesis [44, 45].

Study objective

The goal was to discover CRC-associated DEGs, prioritize hub genes in PPI networks, characterize enriched biological functions and KEGG pathways and investigate candidate gene-miRNA-transcription factor regulatory relationships that may guide future biomarker investigations [46, 47].

2. MATERIALS AND METHODS

2.1 Study design and dataset selection

Using the publicly available microarray gene-expression data from the NCBI GEO database, a retrospective in silico study was undertaken. The following terms were used in a combination search: colorectal cancer, Homo sapiens, CRC carcinoma, mRNA, and datasets. CRC and normal/control tissue data sets were screened for the presence of CRC, availability of normal tissue samples, and expression data suitable for differential-expression analysis. Thesis: An initial screening of up to 40 CRC datasets/samples, followed by prioritization of the datasets that are relevant for Asian populations and finally retaining three datasets for further downstream analysis.

2.2 Differential expression analysis

The Limma framework was used to assess the differential expression between CRC and normal tissues using GEO2R. Genes were deemed significant at $p<0.05$ and log2FC>=2 (up-regulated genes) or log2FC<= -2 (down-regulated genes). To visualize the distribution of DEGs, volcano plots were created. The number of overlaps of DEG lists was determined by an online Venn-diagram tool offered by Ghent University.

2.3 Functional enrichment analysis

Enriched Gene Ontology categories and KEGG pathways of DEGs were identified using Database for Annotation, Visualization and Integrated Discovery (DAVID). The source analysis considered enrichment terms with $p < 0.05$ as statistically significant. The leading functional categories and pathways have been visualized using R.

The construction of a PPI network and prioritization of hub genes were performed.A PPI network was built and hub genes were prioritized.

DEGs were submitted to STRING to build PPI networks with high score (0.90). STRING outputs were imported into Cytoscape for visualization of networks and downstream analyses. Multiple topological metrics are used with cytoHubba plugin to prioritize hub genes: Degree, Edge Percolated Component (EPC), Maximum Neighborhood Component (MNC), Density of Maximum Neighborhood Component (DMNC), Maximal Clique Centrality (MCC), Closeness, Bottleneck, Eccentricity, Radiality, and Betweenness centrality. The study also outlines a module level analysis performed inside of Cytoscape.

2.4 Functional and regulatory network analysis of genes

The functional relationships between prioritized genes were explored using GeneMANIA. To build mRNA-miRNA-transcription factor networks, we used miRNet, and to simplify networks while preserving the informative relationships we applied a Steiner forest approach. Selected hub genes were then compared for expression between colon adenocarcinoma and normal samples using GEPIA.

2.5 Statistical considerations

Analysis of differential expression and enrichment in the source thesis employed nominal $P < 0.05$. A p-value from a multiple-testing correction such as the false-discovery-rate or another method. Correction was not consistently documented; therefore, the findings should be interpreted as exploratory and hypothesis-generating rather than confirmatory.

Table 1. Summary of GEO differential-expression results reported in the thesis

GEO accession	Upregulated DEGs	Downregulated DEGs	Note
GSE130750	63	433	Reported explicitly
GSE160988	134	606	Reported explicitly
GSE161007*	0	16	Accession inferred only from the thesis figure caption; verify against original analysis files before submission

The body text describes a third dataset without consistently naming it, while a figure caption identifies GSE161007. This accession should be verified before journal submission.

3. RESULTS

3.1 Differentially expressed genes in CRC

Three GEO expression datasets were kept for the final analysis. According to the study thresholds, GSE130750 had 63 genes that were upregulated and 433 genes that were downregulated. In GSE160988 there were 134 genes that were upregulated and 606 genes that were downregulated. The third data set was reported as having 16 genes downregulated and no genes upregulated. Volcano plots for each comparison were created to visualize differential-expression patterns. Overall, the source analysis results showed that the number of genes that were downregulated was greater than the number of genes that were upregulated in the datasets.

3.2 Functional enrichment of DEGs

The biological functions of the DEG sets were characterized by DAVID and KEGG analyses. The source thesis lists functional categories and pathways that have been statistically significantly enriched at $p < 0.05$ and graphically displays top enriched annotations. The present manuscript does not reproduce

unreported pathway level statistics as the thesis text does not provide a complete tabulation of enrichment terms, effect sizes, gene counts, or adjusted p values.

3.3 PPI network and Cytoscape analysis

DEGs showed both up- and down-regulation were submitted to STRING separately and the resulted interaction networks were imported to Cytoscape. The networks were then combined to allow assessment of interactions across the combined DEG set. To prioritize genes with influential or highly connected positions in CRC associated PPI network, the network topology was analyzed by 10 cytoHubba centrality algorithms.

3.4 Candidate hub genes

After cytoHubba-based prioritization and network analysis, 25 hub genes were identified, which are reported in the thesis. The gene names that appear in the Results column supplied contain AGT, CCL18, CD80, COL4A1, COL4A2, COL5A1, COL5A2, COL5A3, COL6A1, COL6A3, COL7A1, COL8A1, COL12A1, COL14A1, COL19A1, CX3CL1, CXCL2, CXCL10, IDO1, IL1A, ITGAX, NLRP3, P4HA3, and TNFRSF1B. This collection of candidates is representative of the broad categories of extracellular-matrix remodeling, collagen organization, inflammatory signaling, chemokine activity, and immune regulation. Since one of the 25 reported hub-gene labels is not recoverable from the thesis text provided, the final gene set should be verified against the output of the original Cytoscape/cytoHubba output prior to submission.

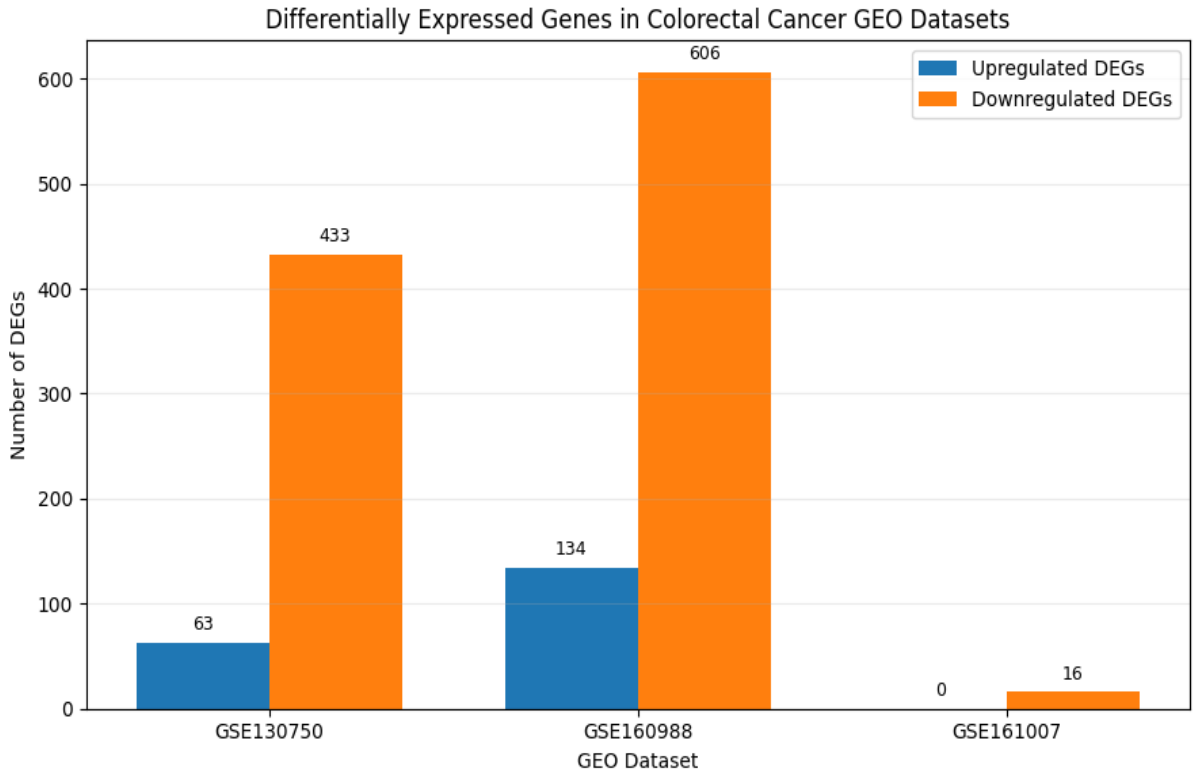
The evaluation of the 3.5 miRNA/transcription factor network and GEPIA.

The prioritized mRNAs were then mapped to predicted and/or curated miRNAs and transcription factors using miRNet. A Steiner forest network was used to minimize the density of edges and highlight important regulatory relationships. The expression levels of selected hub genes were then visualized using GEPIA in the colon adenocarcinoma and normal tissues. The source thesis includes several plots of gene expression levels but lacks full numerical effect estimates or survival data; these plots are treated as "exploratory" validation of the source thesis, but not as full statistical confirmation.

Table 2. Selected candidate hub genes reported in the thesis and their broad biological grouping

Group	Genes reported	Interpretive theme
Extracellular matrix/collagen	COL4A1, COL4A2, COL5A1, COL5A2, COL5A3, COL6A1, COL6A3, COL7A1, COL8A1, COL12A1, COL14A1, COL19A1	Matrix organization and stromal remodeling
Chemokine/inflammatory	CCL18, CX3CL1, CXCL2, CXCL10, IL1A	Immune-cell recruitment and inflammatory signaling
Immune-regulatory	CD80, IDO1, ITGAX, TNFRSF1B	Antigen presentation and immune regulation
Other network candidates	AGT, NLRP3, P4HA3	Signaling, inflammasome biology, and collagen maturation

Interpretive themes are broad biological descriptions and do not establish diagnostic or prognostic performance.



4. DISCUSSION

The integrated bioinformatics analysis involved transcriptomic data and network analysis to prioritize genes which may be involved in the biology of CRC. One thing that stood out was the fact that in the reported differential expression results, genes were almost always down-regulated, though the extent and direction of the change differed between the various datasets. This heterogeneity is anticipated in CRC, as tumors vary by anatomic location, molecular phenotype, stage, stromal composition, presence of immune infiltrates, and technical platform [35].

A striking characteristic of the prioritised gene set is the representation of the genes coding for collagen and extracellular-matrix proteins such as COL4A1, COL4A2, COL5A1, COL5A2, COL6A1, COL6A3 and COL12A1. The remodeling of the extracellular matrix may affect the invasion of tumors, increase tumor stiffness, promote the formation of new blood vessels and alter the relationship between the tumor and cancer-associated fibroblasts. The high centrality of genes involved in the matrix in a network analysis is therefore a plausible biological signal, but network centrality does not indicate a gene to be a clinically useful biomarker[36, 37].

These candidate set includes immune and inflammatory genes including CCL18, CXCL10, IDO1, ITGAX, NLRP3, CD80 and TNFRSF1B. These genes could represent the immune microenvironment, as well as the intricate dynamics between different tumor cells, myeloid cells, lymphocytes and inflammatory mediators. CRC is molecularly and immunologically diverse and immune-associated transcription can be affected by the biology of the tumor as well as by the composition of bulk tissues [38 , 39].

Several cytoHubba algorithms are a strength as genes that are prioritized by multiple topological criteria may be more robust than the genes prioritized by a single centrality measure. Likewise, PPI analysis in combination with GeneMANIA, miRNet and GEPIA offers complementary perspectives of functional connectivity, regulatory relationships and expression differences. These are nonetheless combined from heterogeneous public databases and/or predictions, and the combined networks must be seen as hypothesis-generating [40].

There are a number of caveats that need to be noted. First, the source thesis does not give a full sample-level description for each GEO dataset, and the third accession number is not always present in the

narrative. Secondly, no clear documentation of multiple testing correction was provided, which meant that there was an increased risk of false-positive results [41]. Third, the batch effect and platform heterogeneity in GEO datasets were not explained adequately to harmonize the data. Fourth, reported hub genes were not tested in a separate independent external cohort and prespecified model, or diagnostic performance metrics (e.g., area under the receiver-operating-characteristic curve) reported [51, 52]. Last but not least, there was no wet-lab or prospective clinical validation. The limitations of these studies conclude that the genes identified are candidate biomarkers and not clinically validated biomarkers [42]. Future studies should adopt fully-documented accession numbers and sample metadata, incorporate false-discovery-rate control, perform batch-aware meta-analysis when applicable, confirm the hub-gene signature in independent CRC cohorts and assess diagnostic or prognostic performance using prespecified statistical models. Biological and clinical relevance would require experimental validation by quantitative PCR, immunohistochemistry or other molecular assays.

5. CONCLUSION

An integrated GEO-based bioinformatics workflow successfully analysis and prioritized a network of candidate hub genes related to extracellular matrix, inflammatory and immune regulatory processes for CRC-associated DEGs. The analysis identifies a number of genes, such as collagen-family members, CXCL10, CCL18, IDO1, ITGAX, NLRP3, P4HA3 and TNFRSF1B, for further investigation. These genes should be regarded as candidates for further studies on biomarkers for the future, since the study is exploratory and has not been validated by independent experiments.

REFERENCES

Das V, Kalita J, Pal M. Predictive and prognostic biomarkers in colorectal cancer: a systematic review of recent advances and challenges. Biomed Pharmacother. 2017;87:8-19.

Sawicki T, et al. A review of colorectal cancer in terms of epidemiology, risk factors, development, symptoms and diagnosis. Cancers. 2021;13(9):2025.

Marmol I, et al. Colorectal carcinoma: a general overview and future perspectives in colorectal cancer. Int J Mol Sci. 2017;18(1):197.

Sung JJY, et al. Increasing trend in young-onset colorectal cancer in Asia: more cancers in men and more rectal cancers. Am J Gastroenterol. 2019;114(2):322-329.

Onyoh EF, et al. The rise of colorectal cancer in Asia: epidemiology, screening, and management. Curr Gastroenterol Rep. 2019;21:1-10.

Schetter AJ, Okayama H, Harris CC. The role of microRNAs in colorectal cancer. Cancer J. 2012;18(3):244-252.

Favoriti P, et al. Worldwide burden of colorectal cancer: a review. Updates Surg. 2016;68(1):7-11.

Yee YK, et al. Epidemiology of colorectal cancer in Asia. J Gastroenterol Hepatol. 2009;24(12):1810-1816.

Center MM, et al. Worldwide variations in colorectal cancer. CA Cancer J Clin. 2009;59(6):366-378.

Wong MCS, et al. Prevalence and risk factors of colorectal cancer in Asia. 2019;17(3):317-329.

Siegel RL, et al. Colorectal cancer statistics, 2023. CA Cancer J Clin. 2023;73(3):233-254.

Idrees R, et al. Cancer prevalence in Pakistan: meta-analysis of various published studies to determine variation in cancer figures resulting from marked population heterogeneity in different parts of the country. World J Surg Oncol. 2018;16(1):129.

Schetter AJ, Okayama H, Harris CC. The role of microRNAs in colorectal cancer. Cancer J. 2012;18(3):244-252.

Yang L, Belaguli N, Berger DH. MicroRNA and colorectal cancer. World J Surg. 2009;33:638-646.

Boussios S, et al. The developing story of predictive biomarkers in colorectal cancer. J Pers Med. 2019;9(1):12.

Faber C, Kirchner T, Hlubek F. The impact of microRNAs on colorectal cancer. Virchows Arch. 2009;454:359-367.

O'Brien J, et al. Overview of microRNA biogenesis, mechanisms of actions, and circulation. Front Endocrinol. 2018;9:402.

Huang MY, et al. Clinical implications and future perspectives of circulating tumor cells and biomarkers in clinical outcomes of colorectal cancer. Transl Oncol. 2016;9(4):340-347.
Zou J, et al. Noninvasive fecal testing for colorectal cancer. Clin Chim Acta. 2022;524:123-131.
Kronborg O, et al. Randomised study of screening for colorectal cancer with faecal-occult-blood test. Lancet. 1996;348(9040):1467-1471.
Chen B, et al. Emerging microRNA biomarkers for colorectal cancer diagnosis and prognosis. R Soc Open Sci. 2019;9(1):180212.
Ogunwobi OO, Mahmood F, Akingboye A. Biomarkers in colorectal cancer: current research and future prospects. Int J Mol Sci. 2020;21(15):5311.
Aghagolzadeh P, Radpour R. New trends in molecular and cellular biomarker discovery for colorectal cancer. World J Gastroenterol. 2016;22(25):5678.
Luo X, et al. MicroRNA signatures: novel biomarker for colorectal cancer? Cancer Epidemiol Biomarkers Prev. 2011;20(7):1272-1286.
Lech G, et al. Colorectal cancer tumour markers and biomarkers: recent therapeutic advances. World J Gastroenterol. 2016;22(5):1745.
Zarkavelis G, et al. Current and future biomarkers in colorectal cancer. Ann Gastroenterol. 2017;30(6):613.
Vacante M, et al. Biomarkers in colorectal cancer: current clinical utility and future perspectives. World J Clin Cases. 2018;6(15):869.
Saif MW, Chu E. Biology of colorectal cancer. Cancer J. 2010;16(3):196-201.
Boyle P, Leon ME. Epidemiology of colorectal cancer. Br Med Bull. 2002;64(1):1-25.
Potter JD. Colorectal cancer: molecules and populations. J Natl Cancer Inst. 1999;91(11):916-932.
Calvert PM, Frucht H. The genetics of colorectal cancer. Ann Intern Med. 2002;137(7):603-612.
Gu MJ, et al. Attributable causes of colorectal cancer in China. BMC Cancer. 2018;18:1-9.
Center MM, Jemal A, Ward E. International trends in colorectal cancer incidence rates. Cancer Epidemiol Biomarkers Prev. 2009;18(6):1688-1694.
Pourhoseingholi MA. Increased burden of colorectal cancer in Asia. World J Gastrointest Oncol. 2012;4(4):68.
Araghi M, et al. Global trends in colorectal cancer mortality: projections to the year 2035. Int J Cancer. 2019;144(12):2992-3000.
Rentsch M, et al. Surgery for colorectal cancer-trends, developments, and future perspectives. Visc Med. 2016;32(3):184-191.
Newton K, Newman W, Hill J. Review of biomarkers in colorectal cancer. Colorectal Dis. 2012;14(1):3-17.
Coghlin C, Murray GI. Biomarkers of colorectal cancer: recent advances and future challenges. Proteomics Clin Appl. 2015;9(1-2):64-71.
Walker AS, et al. Future directions for the early detection of colorectal cancer recurrence. J Cancer. 2014;5(4):272.
Falzone L, et al. Integrated analysis of colorectal cancer microRNA datasets: identification of microRNAs associated with tumor development. Aging. 2018;10(5):1000.
Fatima SA, et al. In silico analysis and experimental validation shows negative correlation between miR-1183 and cell cycle progression gene 1 expression in colorectal cancer. PLoS One. 2023;18(8):e0289082.
Zhang X, et al. Identification of key microRNAs and genes between colorectal adenoma and colorectal cancer via deep learning on GEO databases and bioinformatics. Contrast Media Mol Imaging. 2023;2023.
Tabatabaei ES, et al. Integrated bioinformatics and wet-lab analysis revealed cell adhesion prominent genes CDC42, TAGLN and GSN as prognostic biomarkers in colonic-polyp lesions. Sci Rep. 2023;13(1):10307.
Liu X, et al. Identification of crucial genes and pathways associated with colorectal cancer by bioinformatics analysis. Oncol Lett. 2020;19(3):1881-1889.

Xiong Y, et al. Prediction and validation of hub genes associated with colorectal cancer by integrating PPI network and gene expression data. Biomed Res Int. 2017;2017.

Wang Y, Zheng T. Screening of hub genes and pathways in colorectal cancer with microarray technology. Pathol Oncol Res. 2014;20:611-618.

Sun G, et al. Identification of differentially expressed genes and biological characteristics of colorectal cancer by integrated bioinformatics analysis. J Cell Physiol. 2019;234(9):15215-15224.

Ebadfardzadeh J, et al. Employing bioinformatics analysis to identify hub genes and microRNAs involved in colorectal cancer. Med Oncol. 2021;38(9):114.

Eo H. Discovery of novel genetic alteration using meta-analysis of colorectal cancer. Int J High Sch Res. 2023;5(1).

Wu Y, Xu Y. Integrated bioinformatics analysis of expression and gene regulation network of COL12A1 in colorectal cancer. Cancer Med. 2020;9(13):4743-4755.

Pruseth B, et al. Analysis of overexpressed miRNA in circulation and cancer tissue to develop a potential microRNA panel for the diagnosis of colorectal cancer. MicroRNA. 2021;10(4):250-262.

Guo L, et al. Construction and analysis of a ceRNA network reveals potential prognostic markers in colorectal cancer. Front Genet. 2020;11:418.