

Utilizing an In-silico Approach to Pinpoint Potential Biomarkers for Enhanced Early Detection of Colorectal Cancer

Cancer Informatics
Volume 23: 1–11
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11769351241307163



Alireza Gharebaghi¹, Saeid Afshar^{2,3}, Leili Tapak^{4,5} ,
Hossein Ranjbar¹, Massoud Saidijam^{2,3†} and Irina Dinu⁶

¹Neurophysiology Research Center, Hamadan University of Medical Sciences, Hamadan, Iran.

²Department of Molecular Medicine and Genetics, School of Medicine, Hamadan University of Medical Sciences, Hamadan, Iran. ³Research Center for Molecular Medicine, Hamadan University of Medical Sciences, Hamadan, Iran. ⁴Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran. ⁵Modeling of Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran. ⁶School of Public Health, Health Academy, University of Alberta, Edmonton, AB, Canada.

ABSTRACT

OBJECTIVES: Colorectal cancer (CRC) is a prevalent disease characterized by significant dysregulation of gene expression. Non-invasive tests that utilize microRNAs (miRNAs) have shown promise for early CRC detection. This study aims to determine the association between miRNAs and key genes in CRC.

METHODS: Two datasets (GSE106817 and GSE23878) were extracted from the NCBI Gene Expression Omnibus database. Penalized logistic regression (PLR) and artificial neural networks (ANN) were used to identify relevant miRNAs and evaluate the classification accuracy of the selected miRNAs. The findings were validated through bipartite miRNA-mRNA interactions.

RESULTS: Our analysis identified 3 miRNAs: miR-1228, miR-6765-5p, and miR-6787-5p, achieving a total accuracy of over 90%. Based on the results of the mRNA-miRNA interaction network, CDK1 and MAD2L1 were identified as target genes of miR-6787-5p.

CONCLUSIONS: Our results suggest that the identified miRNAs and target genes could serve as non-invasive biomarkers for diagnosing colorectal cancer, pending laboratory confirmation.

KEYWORDS: Colorectal neoplasms, microRNA, smoothly clipped absolute deviation, least absolute shrinkage and selection operator, the minimax concave penalty, artificial neural networks

RECEIVED: June 26, 2024. **ACCEPTED:** November 28, 2024.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by the Hamadan University of Medical Sciences (Grant No. 9906254351).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Leili Tapak, Department of Biostatistics, School of Public Health and Modeling of Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Pajouhesh Square, Hamadan 6516717673, Iran. Email: l.tapak@umsha.ac.ir

Introduction

Colorectal cancer (CRC) is the third most prevalent cancer globally and the second leading cause of cancer-related mortality. In 2020 alone, there were approximately 1.93 million new CRC cases and 935 000 deaths.

Age is a major risk factor for CRC, with most cases occurring in individuals aged 50 or older. Other risk factors include a family history of CRC, inflammatory bowel disease, genetic mutations, poor dietary choices, obesity, and lack of physical activity.^{1–3} In recent decades, developing countries have experienced an epidemiological shift in CRC, marked by a concerning rise in its incidence.⁴ CRC has become a major contributor to cancer-related mortality worldwide.⁵ The early detection of CRC through screening plays a crucial role in enhancing treatment outcomes and improving patient survival rates. This is

primarily because early-stage CRC typically presents no noticeable symptoms. Consequently, individuals with early-stage CRC are often diagnosed at later stages, when the cancer is more advanced and treatment is more challenging.⁶ The overall survival of patients is intricately linked to the progression of cancer at the time of diagnosis. This is primarily due to the fact that the extent of cancer progression upon diagnosis serves as a robust predictor of overall survival.^{7,8}

Early diagnosis has the potential to significantly impact the trajectory of treatment.⁹ Traditional screening methods for CRC, such as fecal immunochemical testing (FIT) and guaiac-based fecal occult blood test (gFOBT), have become routine practices. However, these methods have inherent drawbacks, including low sensitivity and the inability to detect CRC in a timely manner. These limitations have spurred efforts to develop new screening methods that offer improved sensitivity and timely detection.^{6,10}

Biomarkers, as molecular signatures, hold the potential to serve as more effective tools for cancer screening compared to

[†]Massoud Saidijam passed away on July 20, 2020. His contributions to this study were invaluable.



traditional methods.⁶ The dysregulation of genes, both coding and non-coding, along with perturbed signaling pathways, plays a substantial role in cancer development. Recent research has highlighted the significance of leveraging these genes and signaling pathways for early cancer detection.¹¹ miRNAs have emerged as highly recognized biological molecules and genes that intricately regulate the pathways involved in the formation of cancer cells, specifically in CRC. These miRNAs engage in interactions with proteins and other non-coding RNAs, thereby contributing to the pathogenesis of CRC.¹² Extracellular miRNAs have been identified in serum and plasma, rendering them non-invasive biomarkers with potential applications in various disease conditions.^{12,13} Circulating miRNAs in the blood exhibit remarkable stability and reproducibility, rendering them a promising biomarker for CRC. Biological processes can influence the expression of miRNAs, and epigenetic changes can further contribute to alterations in miRNA expression specifically in CRC.¹⁴⁻¹⁶ In recent years, the study of Differentially Expressed miRNAs (DEmiRs) has gained traction in cancer research. DEmiRs are miRNAs whose expression levels significantly differ between normal and disease conditions, such as in cancerous vs. healthy tissues.

One significant challenge in identifying biomarkers associated with different clinical outcomes, such as distinguishing normal from cancerous tissue samples, is the high-dimensional nature of the data. The number of miRNAs often exceeds the sample size, requiring specialized methods to address this issue. Penalized regression models, including Penalized Logistic Regression (PLR), have garnered considerable attention for analyzing this type of data. These models enable simultaneous variable selection and coefficient estimation. As a result, non-informative miRNAs receive close to zero estimations, while the remaining miRNAs in the model are associated with the outcome and can reliably detect CRC.

In this study, we employed PLR with 3 different penalties: Smoothly Clipped Absolute Deviation (SCAD), Least Absolute Shrinkage and Selection Operator (LASSO), and the Minimax Concave Penalty (MCP), to identify miRNAs related to CRC. The primary objective of this article was to identify miRNAs capable of detecting CRC at an early stage. By leveraging systems biology and data mining techniques, we aimed to determine non-invasive biomarkers with high accuracy, facilitating timely treatment through early diagnosis of CRC.

Material and Methods

The bioinformatics strategy presented in Figure 1 involved the utilization of serum microarray datasets to identify miRNAs and key genes associated with CRC through systems biology methods. Initially, miRNAs were extracted from each sample's profile and subjected to evaluation using PLR. Subsequently, an ANN was developed to assess the accuracy of the selected miRNAs. The analysis resulted in the identification of Differentially Expressed miRNAs (DEmiRs) and their

respective target genes. To validate the findings, common genes were identified between the target genes and Differentially Expressed Genes (DEGs) using bipartite miRNA-mRNA interactions.

Notably, factors such as age, health status, and patient risk factors were not accounted for in this study.

miRNA expression profile dataset

Two miRNAs and gene expression datasets for CRC were acquired from the Gene Expression Omnibus (GEO) repository, namely GSE106817 and GSE23878. GSE106817 was generated using the "3D-Human miRNA V21_1.0.0" platform (GPL21263) and comprised 4043 samples including various disease conditions and healthy individuals. Among these, 115 samples were from CRC patients, while 2759 samples were from healthy individuals. In order to maintain balance, 115 healthy samples were randomly selected using R software. The expression levels of 2566 miRNAs were measured in each sample without any initial screening, providing data for subsequent analysis and modeling. Additionally, GSE23878, generated with the "Illumina HumanHT-12 V3.0" platform (GPL6947), consisted of 59 tissue specimens, including 35 CRC samples and 24 normal tissue samples. This dataset was used as a validation set to assess key genes identified in the study.

Statistical Analysis

miRNA selection through penalized model

PLR techniques are a class of statistical learning methods that can be used for variable selection. These techniques attach a penalty to the objective function of the PLR, which shrinks the estimates of the regression coefficients toward zero. In this way, penalized regression techniques can simultaneously perform variable selection and coefficient estimation. In this study, we used PLR models with (1) Smoothly Clipped Absolute Deviation (SCAD) and (2) Least Absolute Shrinkage and Selection Operator (LASSO) and (3) the Minimax Concave Penalty (MCP) to identify important miRNAs. Briefly, PLR is a shrinkage regression model that adds a penalty term to the regression coefficients in the likelihood function. The LASSO penalty considers an absolute value term for each variable in the likelihood function as the penalty term, more specifically. The SCAD penalty is a Smoothly Clipped Absolute Deviation penalty that is defined as follows:

$$p_{\lambda}(t) = \lambda \cdot |t|, \text{ if } |t| \leq \lambda$$

$$p_{\lambda}(t) = -(|t|^2 - 2a\lambda|t| + \lambda^2) / 2(a-1), \text{ if } \lambda < |t| \leq a\lambda$$

$$p_{\lambda}(t) = ((a+1)\lambda^2) / 2, \text{ if } |t| > a\lambda$$

Where t is the regression coefficient and λ is the tuning parameter. The MCP a concave penalty function used in penalized

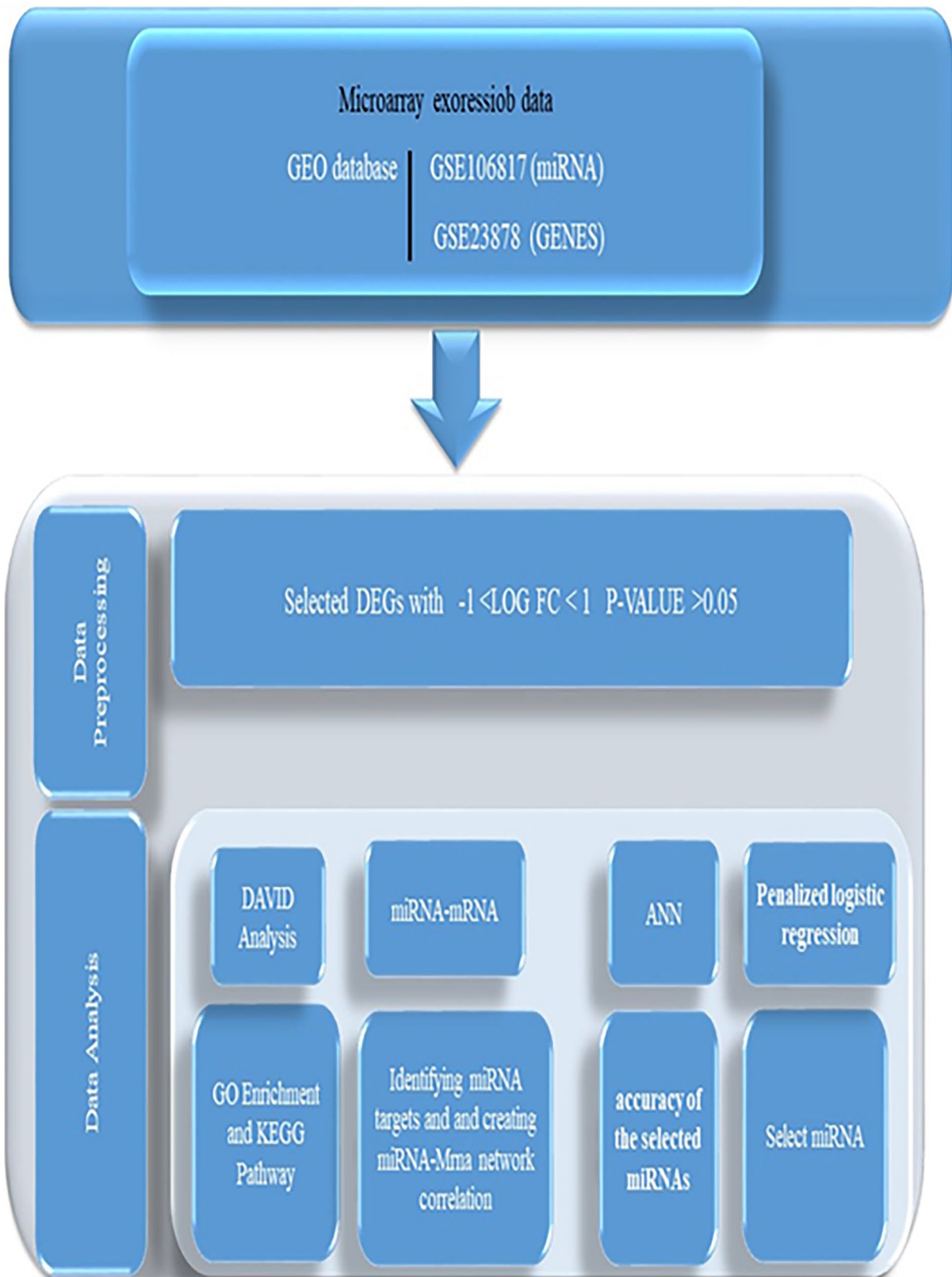


Figure 1. Flow chart of bioinformatics analysis.

regression for variable selection and coefficient estimation. It is defined as follows:

$$p'_\lambda(|\beta_j|) = \left(\lambda - \frac{|\beta_j|}{a}\right) I(|\beta_j| \leq a\lambda).$$

We used a 10-fold cross-validation strategy to select the optimal value of λ . The value of λ that minimized the Bayesian Information Criterion was chosen as the optimal value. The PLR models with the 3 types of penalties were repeated 1000

times and the miRNA that were selected at least by 2 penalties were considered as miRNA biomarkers. The “grpreg” package was used for gene selection in R software version 4.0.2.¹⁷⁻¹⁹ The source code used for the analysis is available on GitHub at <https://github.com/ARGHAREBAGHI>.

Artificial neural networks

The analysis involved utilizing the R package version 4.0.2 software to train an ANN. To prepare the data for training, it was normalized using the maximum and minimum values. Subsequently, an ANN model was designed in the R software package, incorporating the important variables. The model parameters were adjusted to construct a disease prediction model, taking into account the weight information derived from the expression of miRNAs. In this model, the pathogenicity score was computed by summing the weighted scores, which were multiplied by the significant miRNAs' disappearance. For gene selection, the “neuralnet” package (version 19) within R software version 4.0.2 was employed. To optimize the performance of the model, a 10-fold cross-validation strategy was employed, allowing for the fine-tuning of hyper-parameters.²⁰⁻²³

miRNA target prediction

The miRWalk 3.0 online database, available at <http://mirwalk.uni-hd.de/>, is a user-friendly and easily accessible resource that provides predictive data obtained through a machine learning algorithm. The database prioritizes accuracy, simplicity, and up-to-date information to facilitate efficient miRNA research. In the context mentioned, miRWalk was utilized as a tool to search for predicted target genes of miRNAs.²⁴

Protein-protein interaction (PPI) network analysis

In this study, an interactive network of proteins was employed to investigate gene interactions and identify hub genes. The protein-protein interaction (PPI) network for the selected genes was constructed using the STRING online tool, with an interaction score threshold of 0.4. To visualize and analyze the constructed network, Cytoscape software version 3.8.2 was utilized. The CytoHubba plugin version 1.6 within Cytoscape was employed to evaluate various network measures, including Maximum Neighborhood Component (MNC), Maximal Clique Centrality (EPC), and DEGREE, to identify the hub genes within the network. Furthermore, a Venn diagram was utilized to identify the common genes and select the hub genes that appeared consistently across the different measures.^{25,26}

DEGs' enrichment analyses

In this study, the function of DEGs was explored through Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analyses. The GO classification system, encompassing molecular function (MF), cellular

component (CC), and biological processes (BP), was utilized to gain insights into the functional characteristics of the DEGs. To conduct the functional enrichment analysis of the gene list, the Database for Annotation, Visualization, and Integrated Discovery (DAVID) program, accessible at <https://david.ncifcrf.gov>, was employed. The analysis involved determining significant enrichment of gene functions using an adjusted *P*-value cutoff threshold of $<.05$.^{27,28}

Potential miRNA-mRNA interactions

In this study, DE miRNAs were identified between CRC samples and normal tissues, considering an adjusted *P*-value $<.05$ and $|\log_{2}FC| > 1$ as the criteria for differential expression. Subsequently, the target genes of the DE miRNAs were determined using the miRWalk database. To understand the miRNA-mRNA regulatory interactions comprehensively, a bipartite miRNA-mRNA correlation network was constructed and analyzed using Cytoscape version 3.8.2 software. The interaction score threshold of 0.4 was employed to filter out weak interactions in the network. The choice of a bipartite network is appropriate for this study since mRNAs and miRNAs do not directly interact with each other. This network structure allows mRNAs and miRNAs to be connected solely through their interactions with target genes.

Hub gene validation by GEPIA

The Gene Expression Profiling Interactive Analysis (GEPIA) database (<http://gepia.cancer-pku.cn/>) is a web-based tool designed for fast and CHECK FOR PLAGIRISM : customizable analyses using data from The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) projects. In this study, GEPIA was used to validate the expression of key hub genes by comparing cancerous and normal tissue samples, specifically focusing on colorectal cancer. Differential gene expression was analyzed using ANOVA, with statistical significance set at *P*-value $<.05$ and a fold change greater than 2.

Result

Differentially expression analysis

The miRNAs expression data series (GSE106817) was utilized to identify miRNAs that were DE miRNAs, as well as DEGs. In order to validate the findings, a total of 3763 DEGs were identified by applying the criteria of an adjusted *P*-value $<.05$ and $|\log_{2}FC| > 1$. It was observed that these genes overlapped with the DEGs identified in the primary data series (GSE23878), which was utilized for comparison.

Identification of differentially expressed miRNAs

The miRNA expression data was utilized to train the PLR model, as outlined in the Methods section, with the aim of identifying DE miRNAs associated with CRC diagnosis. The PLR

Table 1. Frequencies of the selected miRNA over 1000 repetitions using penalized logistic regression by SCAD, MCP, and LASSO penalties.

MIRNA	SCAD	MCP	LASSO	TOTAL ACCURACY
MIMAT0005582	1	1000	1000	.966
MIMAT0019776	1000			.983
MIMAT0027430	1		1000	.966
MIMAT0027436	961			.966
MIMAT0027474	1		1000	.966
MIMAT0015079		305		.759
MIMAT0003320			1000	.845
MIMAT0004970			1000	.966
MIMAT0005922			1000	.931
MIMAT0015075			389	.879
MIMAT0018949			1000	.931
MIMAT0022259			1000	.966
MIMAT0019776			1000	.931
MIMAT0027392			1000	.931
No. selected miRNA	5	2	11	

model used the binary outcome variable, where 1 represented CRC and 0 denoted healthy controls. In Table 1, we present the names of the 14 selected DEmiR profiles and their respective frequencies, determined over 1000 repetitions using LASSO, SCAD, and MCP methods. LASSO selected 11 miRNA profiles, while SCAD and MCP identified 5 and 2 miRNA profiles, respectively. Notably, 3 miRNAs (miR-6765-5p, miR-6787-5p, and miR-1228) were confirmed as significant in at least 2 PLR methods.

The results of the univariate PLR analysis for the selected miRNAs are presented in Table 2, which includes the regression coefficient, standard error of the coefficient, odds ratio (OR), and corresponding *P*-values. Notably, the results demonstrate that all 13 miRNAs exhibited statistically significant associations with the diagnosis of CRC.

Table 2 presents the outcomes of unpenalized logistic regression for estimating the regression coefficients of the selected miRNAs. The table reveals that certain miRNAs exhibited a positive association with CRC, whereas others displayed a negative association with CRC.

- Positively associated miRNAs: miR-1228, miR-6765-5p, miR-6768, and miR-1268. This means that an increase in the expression of these miRNAs increases the chance of CRC.
- Negatively associated miRNAs: miR-1343, miR-6787-5p, miR-650, miR-920, miR-3190, miR-4433, miR-5100, miR-1343, and miR-6746. This means that a decrease in the expression of these miRNAs increases the chance of CRC.

The miRNAs identified through PLR were employed as inputs for an ANN model to develop classifiers capable of diagnosing patients. The ANN model was designed with a 1:1:1 architecture, comprising a single input layer, 1 hidden layer, and 1 output layer. The activation functions used in the model were sigmoid for the input layer, hyperbolic tangent for the hidden layer, and linear for the output layer.

The input variables for the ANN model were the miRNA expression values that were chosen in the preceding step. The model's output was a binary value, either 0 or 1, enabling the classification of patients as non-cancerous or cancerous, respectively. This classification holds the potential for early cancer detection, offering valuable diagnostic capabilities.

The outcomes of the ANN model are displayed in the final column of Table 1. Notably, a majority of the miRNAs exhibit a total accuracy greater than 90%, underscoring their significant potential for cancer detection.

Identification of key genes using PPI network analysis

In this study, an analysis was conducted using the PPI (Protein-Protein Interaction) network to explore the 3763 DEGs. The resulting PPI network consisted of 443 nodes and 8314 edges, as depicted in Figure 4. Additionally, the Venn diagram analysis of the 10 top genes, using the 3 methods, resulted in the identification of 7 hub genes: *CDC20*, *MAD2L1*, *UBE2C*, *CDK1*, *AURKB*, *CCNA2*, and *TOP2A*. These findings are illustrated in Figure 2.

Table 2. Results of fitting univariate logistic regression for the selected genes using penalized logistic regression by SCAD, MCP, and LASSO penalties.

MIRNA	SCAD			MCP			LASSO		
	β (S.E)	OR	P-VALUE	β (S.E)	OR	P-VALUE	β (S.E)	OR	P-VALUE
MIMAT0005582	10.95 (2.80)	56954	<.0001	10.95 (2.80)	56954	<.0001	10.95 (2.80)	56954	<.0001
MIMAT0019776	-3.04 (.56)	.048	<.0001						
MIMAT0027430	12.23 (2.46)	204843	<.0001				12.23 (2.46)	204843	<.0001
MIMAT0027436	1.81 (.34)	6.11	<.0001						
MIMAT0027474	-5.84 (1.24)	.003	<.0001				-5.84 (1.24)	.003	<.0001
MIMAT0015079				-1.44 (.28)	.237	<.0001			
MIMAT0003320							-2.01 (.26)	.134	<.0001
MIMAT0004970							-3.36 (.59)	.035	<.0001
MIMAT0005922							9.99 (1.60)	21807	<.0001
MIMAT0015075							-1.74 (.21)	.176	<.0001
MIMAT0018949							-4.69 (.59)	.009	<.0001
MIMAT0022259							-2.27 (.40)	.103	<.0001
MIMAT0019776							-3.04 (.56)	.048	<.0001
MIMAT0027392							-6.57 (1.38)	.001	<.0001

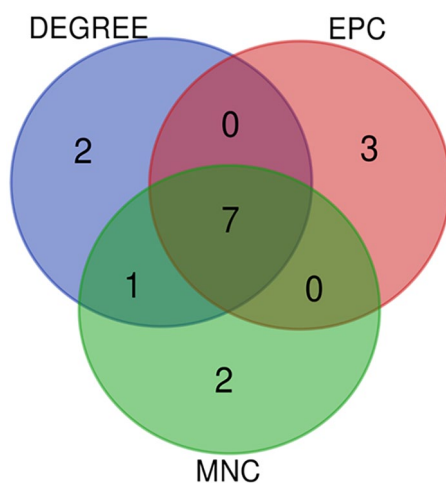


Figure 2. The overlap between the top 10 predicted target genes, ranked by MNC, EPC, and DEGREE illustrated in a Venn diagram. The number 7 in the image's center describes the 3 groups' commonalities.

Functional and pathway enrichment analysis

The results of the GO study, biological processes (BP), cellular components (CC) and molecular functions (MF) were significantly enriched:

- Top 10 terms BP: rRNA processing, cell division, translation, mitochondrial translation, mitotic spindle

organization, protein folding, cytoplasmic translation, ribosomal large subunit biogenesis, proteasomal ubiquitin-independent protein catabolic process, mitotic sister chromatid segregation.

- Top 10 terms CC: nucleoplasm, cytosol, membrane, extracellular exosome, cytoplasm, nucleus, endoplasmic reticulum, mitochondrion, chromosome, ribosome.
- Top 10 terms MF: protein binding, RNA binding, identical protein binding, structural constituent of ribosome, cadherin binding, enzyme binding, chaperone binding, ATPase activity, snoRNA binding, unfolded protein binding.

On other hand, KEGG pathway analysis indicated the following pathways involved: Nucleocytoplasmic transport, Proteasome, DNA replication, Spliceosome, Glutathione metabolism, Ribosome, Protein processing in endoplasmic reticulum, p53 signaling pathway (Figure 3).

BiPartite miRNA and mRNA network analysis

mRNA-miRNA network analysis is a valuable computational approach utilized for understanding the underlying mechanisms contributing to CRC pathogenesis. In this particular study, the MiRwalk database was employed to identify target genes of DEmiRs. By assessing the overlap between the

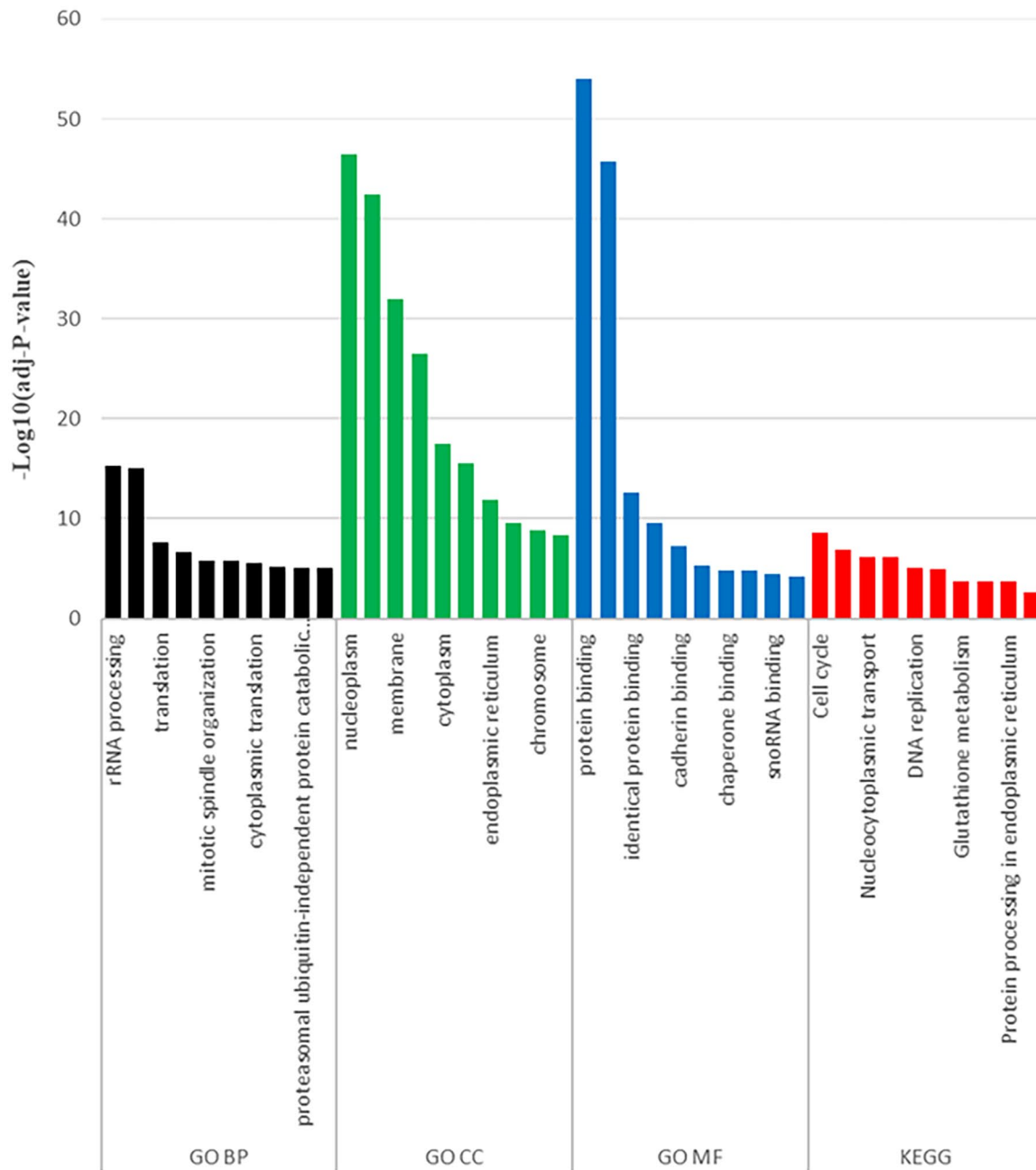


Figure 3. Gene Ontology (GO) and KEGG pathway enrichment analyses were performed for the module genes. The top 10 GO terms in Biological Process (BP), Molecular Function (MF), and Cellular Component (CC), along with significant KEGG pathways, are presented.

identified miRNA targets and the validated DE miGs, key hub genes such as *CDK1* and *MAD2L1* were identified as both targets of miR-6787 and pivotal players in CRC. Notably, the expression of miR-6787-5p was significantly downregulated in cancer tissue samples compared to normal tissue samples, with *CDK1* and *MAD2L1*, being identified as its target genes. These findings highlight the intricate regulatory network involving miRNAs and their target genes in CRC (Figure 4).

Gene expression analysis of the central hub genes

We used the GEPIA database to analyze the expression of 2 candidate genes in cancer tissues and normal samples from the

TCGA-COAD dataset. The results revealed that *CDK1* and *MAD2L1* were both significantly upregulated in tumors in comparison to normal tissues presented in Figure 5.

Discussion

CRC is a leading cause of global mortality, making early detection vital for improved treatment response and reduced mortality rates. Biomarkers play a critical role in CRC diagnosis and treatment, and bioinformatics tools facilitate the identification of CRC-related biomarkers and molecular interactions.²⁹⁻³¹ In this study, a bioinformatics approach was employed, utilizing 2 databases, GSE106817 and GSE23878, to identify DE miRs and hub genes associated with the progression of CRC. The

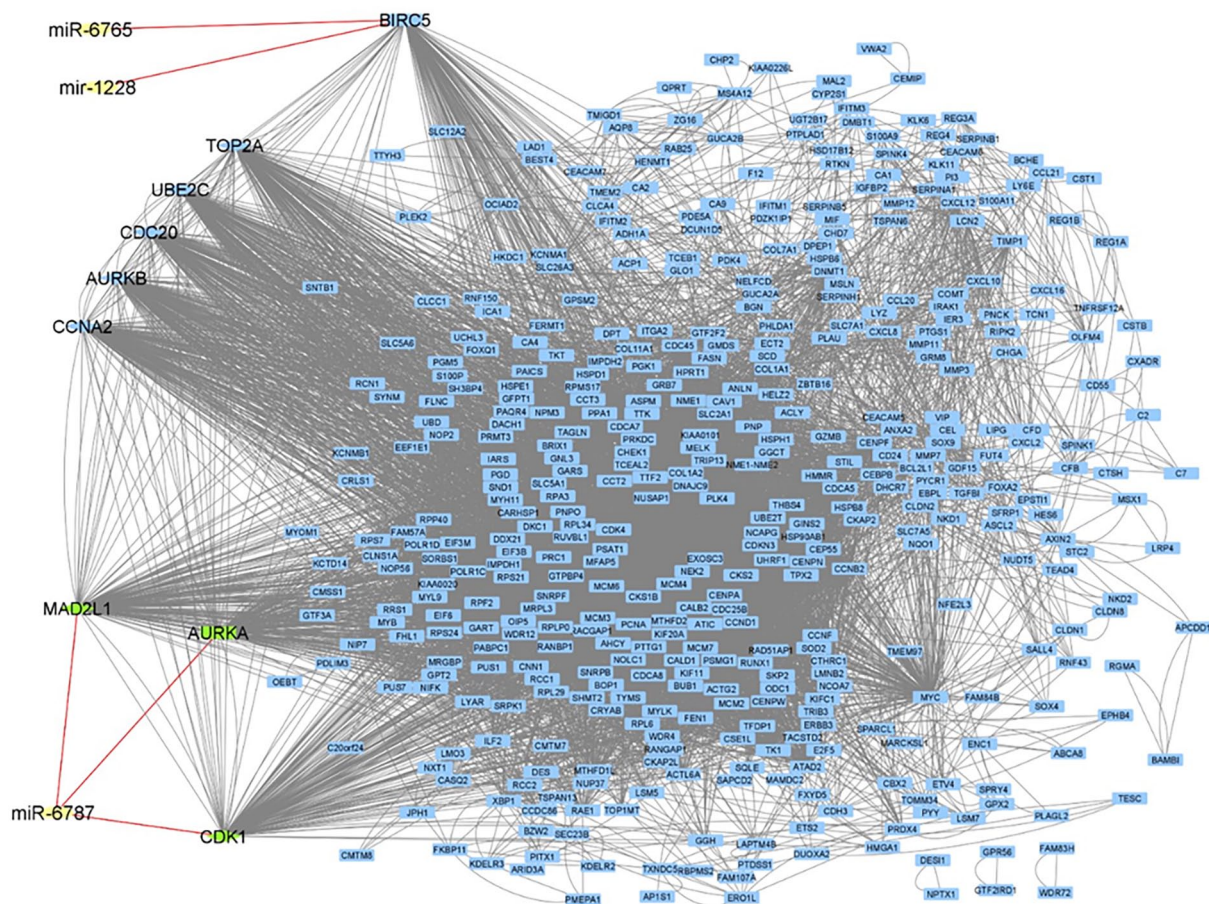


Figure 4. Bipartite mRNA-miRNA subnetwork for CRC. Blue diamonds consist of hub genes between CRC and normal tissues. Green diamonds consist of 2 hub genes targeting miR-6787. Cytoscape v.3.8.2 was used to visualize the network.

analysis of these databases enabled the identification of specific miRNAs and genes that play a crucial role in CRC progression. By investigating the expression patterns and interactions of these DE miRNAs and hub genes, valuable insights into the molecular mechanisms underlying CRC development and progression can be gained. miRNAs such as miR-6765-5p, miR-6787-5p, and miR-1228 were selected based on their intersection in LASSO, MCP, and SCAD regression methods. The overall accuracy of these 3 miRNAs exceeded 95%, underscoring their potential as promising biomarkers for stable plasma level determination in CRC patients. The study also demonstrated the utility of an ANN employing 3 different penalty functions to effectively identify miRNAs significantly associated with CRC.

miRNAs have emerged as key regulators in cancer biology, functioning as both tumor suppressors and oncogenes depending on their expression patterns and the cancer type. These small non-coding RNAs play a pivotal role in a range of cancer-related processes, including initiation, malignant transformation, progression, and metastasis. Recent research has demonstrated that certain cancers have unique miRNA signatures, making them valuable diagnostic and prognostic markers

as well as potential therapeutic targets. Advances in techniques such as microarray analysis, RT-PCR, and next-generation sequencing have facilitated the profiling of miRNAs in various cancer types, even from archived tumor tissues. Emerging detection methods, such as nanoparticle-based and hybridization chain reaction (HCR) amplification, aim to enhance miRNA detection sensitivity. miRNAs are also stable in body fluids, making them promising candidates for non-invasive cancer diagnostics. Their dysregulation in cancer cells, influenced by both genetic and epigenetic factors, highlights their role in tumorigenesis, and disruptions in the miRNA biogenesis process could significantly contribute to cancer development.^{32,33}

In CRC, miR-1228 is often downregulated. This downregulation is associated with poor prognosis. The exact role of miR-1228 in CRC is not fully understood, but it is thought to play a role in tumor growth and progression. miR-1228 targets a number of genes that are involved in cell proliferation, angiogenesis, and apoptosis. By targeting these genes, miR-1228 helps prevent cancer cells from growing and spreading. Numerous studies have shown that miR-1228 plays an essential role in the proliferation of cancer cells and

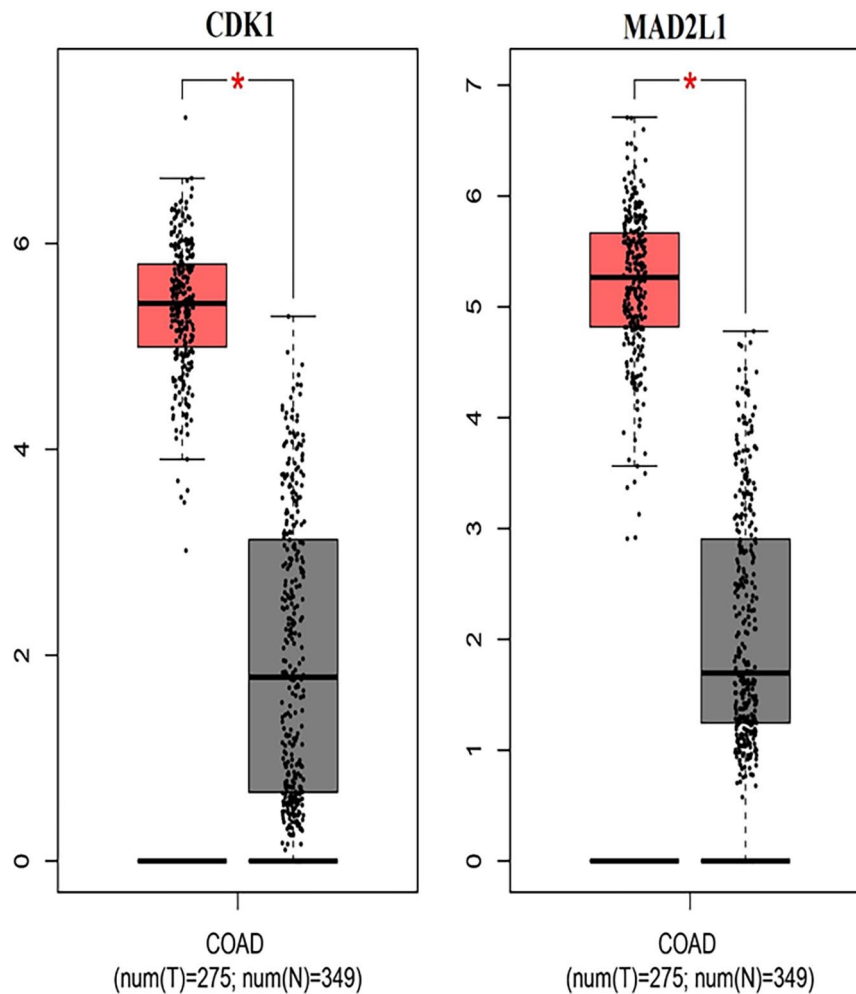


Figure 5. Validation of hub genes in colorectal cancer using TCGA-COAD. Two hub genes including *CDK1*, and *MAD2L1* were significantly upregulated in CRC tissues compared to normal tissues in TCGA- COAD data.

can be used for early detection of cancer.^{34,35} miR-1228 regulates stress-induced cellular apoptosis by targeting the MOAP1 protein.³⁶ In another report, the findings showed that miR-1228 has a role in metabolism, maintaining cell survival, regulating apoptosis, stimulus- response, and survival. However, some studies have investigated the target gene miR-1228 for CRC.^{37,38} LRP1 is the target gene of miR-1228 and is located on chromosome 12.^{39,40} This gene mainly plays a role in basic metabolism and cell structure, which is a key component of maintaining cell survival. In past research, the expression level of miR-1228-3p has been checked in drug resistance of breast cancer, chronic heart failure, endometrial carcinoma, prostate cancer, CRC, and cancer secretions. The expression level of miR-1228-3p is stable in blood circulation and can be used as a biomarker.⁴¹ In a study by Yang et al,³⁷ it was revealed that miR-1228 remained unaffected by surgical treatment, indicating its suitability as an optimal reference gene for treatment studies. Additionally, the circulating level of miR-1228 was found to be independent of tumor stage.

In CRC, miR-6787-5p is often downregulated. This downregulation is associated with poor prognosis. The exact role of miR-6787-5p in CRC is not fully understood, but it is thought to play a role in tumor growth and progression. miR-6787-5p targets a number of genes that are involved in cell proliferation, angiogenesis, and apoptosis. By targeting these genes, miR-6787-5p helps prevent cancer cells from growing and spreading.⁴² The exact role of miR-6765-5p in CRC is not fully understood.

Bioinformatics analysis was then performed using the MNC, EPC, and DEGREE tools in Cytoscape software. The functional and biological interactions between the DEGs were investigated using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses. In the present study, the nucleoplasm was identified as one of the significant enrichment pathways of DEGs in CRC. Network analysis demonstrated that 4 genes of DEGs are involved in this pathway. These findings suggest that the DEGs are involved in a number of biological processes that are important for the pathogenesis of CRC. Further research is needed to

confirm these findings and to identify new diagnostic targets for CRC.⁴³ Therapeutic modulation of cell membrane lipid composition and organization is an emerging field with potential applications in a variety of diseases, including cancer. Research has shown that this approach could be used to treat a variety of diseases, including cancer.⁴⁴ It has been shown that GO terms such as rRNA processing,⁴⁵ translation,⁴⁶ Mitochondrial translation,⁴⁷ mitotic spindle organization,⁴⁸ extracellular exosome.⁴⁹ and protein binding⁵⁰ were associated with CRC

By using miRNA-mRNA expression profiling, *CDK1* and *MAD2L1* were identified as the most important genes playing an important role in CRC. The *CDK1* gene encodes a protein known as cyclin-dependent kinase 1, which belongs to a family of enzymes involved in the regulation of the cell cycle. The cell cycle is a fundamental process responsible for cell growth, division, and the generation of new cells. In CRC, the *CDK1* gene can undergo mutations, resulting in abnormal functioning. These mutations can lead to excessive production of the cyclin-dependent kinase 1 protein. Scientific investigations have demonstrated that dysregulation of *CDK1* accelerates tumor growth and uncontrolled proliferation of cancer cells.^{51,52} Zhang et al⁵³ revealed that *CDK1*, in addition to being overexpressed and sensitive to apoptosis in CRC cells, plays a crucial role in controlling the cell cycle and contributes to the development of colorectal tumors through an iron-regulated signaling axis. Previous studies have established a link between *CDK1* overexpression and the development of colorectal, liver, and lung cancers, ultimately impacting patient survival.⁵⁴

MAD2L1 plays a crucial role as a tumor suppressor gene in regulating the cell cycle. Mutations in the *MAD2L1* gene can disrupt the normal control of cell growth and division, which can contribute to the development of cancer. Deletion of the *MAD2L1* gene has been found to impede the growth of CRC cells.^{55,56} Venugopal et al⁵⁷ revealed that there is a higher expression of *MAD2L1* in CRC cell lines and tissues, and this overexpression has been associated with poor prognosis. Li et al⁵⁵ revealed that *MAD2L1* gene has demonstrated potential as a biomarker for colorectal cancer, according to previous studies.

The present study introduced a novel set of gene expression profiles that are predictive of CRC patients using a miRNA-mRNA model. This model provides a different perspective than the traditional proportional point of view.

Conclusions

This study identified 3 novel miRNAs (miR-1228, miR-6765-5p, and miR-6787-5p) that are potentially associated with CRC and could serve as biomarkers. Additionally, the target genes related to these miRNAs, namely *CDK1* and *MAD2L1*, were found to be upregulated in CRC compared to normal tissues. The miRNAs associated with the hub genes in the mRNA-miRNA bipartite network played a pivotal role in

CRC. However, further molecular studies are warranted to validate the role of these genes in CRC tumorigenesis.

List of Abbreviations

ABBREVIATION	DEFINITION
miRNAs	microRNAs
CRC	Colorectal cancer
ANN	Artificial neural networks
PLR	Penalized logistic regression
SCAD	Smoothly clipped absolute deviation
LASSO	Least absolute shrinkage and selection operator
MCP	The minimax concave penalty
GEO	Gene Expression Omnibus
DEmiRs	Differentially Expressed miRNAs
PPI	Protein-protein interaction
DEGs	Differentially expressed genes

Acknowledgements

We acknowledge that this study was approved by the Ethics Committee of Hamadan University of Medical Sciences (IR.UMSHA.REC.1399.537) and is part of a Professional Doctorate in Medicine (MD) thesis conducted at Hamadan University of Medical Sciences, Hamadan, Iran.

Authors Contributions

S. A. and L. T. conceived the research topic, explored that idea, supervised the project, analyzed the data and drafted the manuscript.

L. T., S. A., A. GH., H. R., M. S. and I. D. provided critical review and participated in data analysis and writing. All authors reviewed the manuscript.

Availability of Data and Materials

The datasets analyzed during the current study are publicly available from <https://www.ncbi.nlm.nih.gov/geo> with accession series of GSE106817 and GSE23878.

Ethics Approval and Consent to Participate

All methods were carried out in accordance with relevant guidelines and regulations, and the study was approved by the Ethical Committee of the Hamadan University of Medical Sciences. This study was conducted on a public data set and this is a secondary data analysis. Therefore, a waiver of informed consent was awarded for this study by the ethical committee of Hamadan University of Medical Sciences (Ethical code: IR.UMSHA.REC.1399.537).

ORCID iD

Leili Tapak  <https://orcid.org/0000-0002-4378-3143>

REFERENCES

- Zhang S, Lian P, Huang T, Zhou J. Effect of Quxie capsule in patients with colorectal cancer: a systematic review and meta-analysis. *Medicine (Baltimore)*. 2021;100:e24322.
- Lee R, Holmes D. Barriers and recommendations for colorectal cancer screening in Africa. *Glob Health Action*. 2023;16:2181920.
- Morgan E, Arnold M, Gini A, et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut*. 2023;72:338-344.
- Høydahl Ø, Edna TH, Xanthoulis A, Lydersen S, Endreseth BH. Long-term trends in colorectal cancer: incidence, localization, and presentation. *BMC Cancer*. 2020;20:1077.
- Jian F, Yangyang R, Wei X, et al. The prognostic and predictive significance of circRNA CDR1as in tumor progression. *Front Oncol*. 2020;10:549982.
- Ogunwobi OO, Mahmood F, Akingboye A. Biomarkers in colorectal cancer: current research and future prospects. *Int J Mol Sci*. 2020;21:5311.
- Dashti S, Taheri M, Ghafouri-Fard S. An in-silico method leads to recognition of hub genes and crucial pathways in survival of patients with breast cancer. *Sci Rep*. 2020;10:18770.
- Dulskas A, Gaizauskas V, Kildusiene I, Samalavicius NE, Smailyte G. Improvement of survival over time for colorectal cancer patients: a population-based study. *J Clin Med*. 2020;9:4038.
- Rasmussen J, Langerman H. Alzheimer's disease - why we need early diagnosis. *Degener Neurol Neuromuscul Dis*. 2019;9:123-130.
- Kronborg O, Regula J. Population screening for colorectal cancer: advantages and drawbacks. *Dig Dis*. 2007;25:270-273.
- Koveitypour Z, Panahi F, Vakilian M, et al. Signaling pathways involved in colorectal cancer progression. *Cell Biosci*. 2019;9:97.
- Schee K, Fodstad Ø, Flatmark K. MicroRNAs as biomarkers in colorectal cancer. *Am J Pathol*. 2010;177:1592-1599.
- Giannopoulou N, Constantinou C. Recent developments in diagnostic and prognostic biomarkers for colorectal cancer: a narrative review. *Oncology*. 2023;101:675-684.
- Yan L, Zhao W, Yu H, Wang Y, Liu Y, Xie C. A comprehensive meta-analysis of microRNAs for predicting colorectal cancer. *Medicine (Baltimore)*. 2016;95:e2738.
- Pu XX, Huang GL, Guo HQ, et al. Circulating miR-221 directly amplified from plasma is a potential diagnostic and prognostic marker of colorectal cancer and is correlated with p53 expression. *J Gastroenterol Hepatol*. 2010;25:1674-1680.
- Clancy C, Joyce MR, Kerin MJ. The use of circulating microRNAs as diagnostic biomarkers in colorectal cancer. *Cancer Biomark*. 2015;15:103-113.
- Chintalapudi N, Angeloni U, Battineni G, et al. LASSO regression modeling on prediction of medical terms among seafarers' health documents using tidy text mining. *Bioengineering (Basel)*. 2022;9:124.
- Kim SM, Kim Y, Jeong K, Jeong H, Kim J. Logistic LASSO regression for the diagnosis of breast cancer using clinical demographic data and the BI-RADS lexicon for ultrasonography. *Ultrasonography*. 2018;37:36-42.
- McEligot AJ, Poynor V, Sharma R, Panangadan A. Logistic LASSO regression for dietary intakes and breast cancer. *Nutrients*. 2020;12:2652.
- Wang F, Casalino LP, Khullar D. Deep learning in medicine—promise, progress, and challenges. *JAMA Intern Med*. 2019;179:293-294.
- Zou J, Han Y, So SS. Overview of artificial neural networks. *Methods Mol Biol*. 2008;458:15-23.
- Zhang Z. A gentle introduction to artificial neural networks. *Ann Transl Med*. 2016;4:370.
- Han SH, Kim KW, Kim S, Youn YC. Artificial neural network: understanding the basic concepts without mathematics. *Dement Neurocogn Disord*. 2018;17:83-89.
- Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: an online resource for prediction of microRNA binding sites. *PLoS One*. 2018;13:e0206239.
- Scholes HM, Cryar A, Kerr F, et al. Dynamic changes in the brain protein interaction network correlates with progression of Aβ42 pathology in Drosophila. *Sci Rep*. 2020;10:18517.
- Jeon J, Nim S, Teyra J, et al. A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening. *Genome Med*. 2014;6:57.
- Hoque MN, Sarkar MMH, Khan MA, et al. Differential gene expression profiling reveals potential biomarkers and pharmacological compounds against SARS-CoV-2: insights from machine learning and bioinformatics approaches. *Front Immunol*. 2022;13:918692.
- Izadi F. Differential connectivity in colorectal cancer gene expression network. *Iran Biomed J*. 2019;23:34-46.
- Yang K, Yang P, Chen S, et al. Vibration identification based on Levenberg-Marquardt optimization for mitigation in adaptive optics systems. *Appl Opt*. 2018;57:2820-2826.
- Thomas DS, Fourkala EO, Apostolidou S, et al. Evaluation of serum CEA, CYFRA21-1 and CA125 for the early detection of colorectal cancer using longitudinal preclinical samples. *Br J Cancer*. 2015;113:268-274.
- Kim Y, Lee S, Park S, et al. Gastrointestinal tract cancer screening using fecal carcinoembryonic antigen. *Ann Clin Lab Sci*. 2003;33:32-38.
- Selvakumar SC, Preethi KA, Sekar D. MicroRNAs as important players in regulating cancer through PTEN/PI3K/AKT signalling pathways. *Biochim Biophys Acta Rev Cancer*. 2023;1878:188904.
- Thomas P, Preethi KA, Selvakumar SC, Ramani P, Sekar D. Relevance of micro-RNAs and their targets as a diagnostic and prognostic marker in oral squamous cell carcinoma. *J Oral Maxillofac Pathol*. 2023;27:364-373.
- Zhang N, Hu X, Du Y, Du J. The role of miRNAs in colorectal cancer progression and chemoradiotherapy. *Biomed Pharmacother*. 2021;134:111099.
- Duran-Sanchon S, Vila-Navarro E, Marcuello M, et al. Validation of miR-1228-3p as housekeeping for microRNA analysis in liquid biopsies from colorectal cancer patients. *Biomolecules*. 2019;10:16.
- Yan B, Zhao JL. miR-1228 prevents cellular apoptosis through targeting of MOAP1 protein. *Apoptosis*. 2012;17:717-724.
- Yang CL, Tsai FM, Chen CW, Hsiao KH, Chen JH, Kao WY. Comparing miR-16 and miR-1228 as an optimal endogenous control for quantification of circulating microRNAs in colorectal cancer patients. *Tzu Chi Med J*. 2022;34:318-322.
- Hu J, Wang Z, Liao BY, et al. Human miR-1228 as a stable endogenous control for the quantification of circulating microRNAs in cancer patients. *Int J Cancer*. 2014;135:1187-1194.
- Boulagnon-Rombi C, Schneider C, Leandri C, et al. LRP1 expression in colon cancer predicts clinical outcome. *Oncotarget*. 2018;9:8849-8869.
- Zhang Y, Dai J, Deng H, et al. miR-1228 promotes the proliferation and metastasis of hepatoma cells through a p53 forward feedback loop. *Br J Cancer*. 2015;112:365-374.
- Xue WX, Zhang MY, Rui L, Liu X, Yin YH, Qu YQ. Serum miR-1228-3p and miR-181a-5p as noninvasive biomarkers for non-small cell lung cancer diagnosis and prognosis. *Biomed Res Int*. 2020;2020:9601876.
- Chen Z, Xie Y, Chen W, Li T, Chen X, Liu B. microRNA-6785-5p-loaded human umbilical cord mesenchymal stem cells-derived exosomes suppress angiogenesis and metastasis in gastric cancer via INHBA. *Life Sci*. 2021;284:119222.
- Vychytilova-Faltejskova P, Radova L, Sachlova M, et al. Serum-based microRNA signatures in early diagnosis and prognosis prediction of colon cancer. *Carcinogenesis*. 2016;37:941-950.
- Preta G. New insights into targeting membrane lipids for cancer therapy. *Front Cell Dev Biol*. 2020;8:571237.
- Nait Slimane S, Marcel V, Fenouil T, et al. Ribosome biogenesis alterations in colorectal cancer. *Cells*. 2020;9:2361.
- Bash-Imam Z, Thérizols G, Vincent A, et al. Translational reprogramming of colorectal cancer cells induced by 5-fluorouracil through a miRNA-dependent mechanism. *Oncotarget*. 2017;8:46219-46233.
- Crisuolo D, Avolio R, Matassa DS, Esposito F. Targeting mitochondrial protein expression as a future approach for cancer therapy. *Front Oncol*. 2021;11:797265.
- Dalton WB, Yang VW. Mitotic origins of chromosomal instability in colorectal cancer. *Curr Colorectal Cancer Rep*. 2007;3:59-64.
- Kotelevets L, Chastre E. Extracellular vesicles in colorectal cancer: from tumor growth and metastasis to biomarkers and nanomedications. *Cancers (Basel)*. 2023;15:1107.
- Yu Z-L, Chen Y-F, Zheng B, et al. Protein-protein interaction analysis reveals a novel cancer stem cell related target TMEM17 in colorectal cancer. *Cancer Cell Int*. 2021;21:94.
- Liu F, Zhu C, Ma H, Yang Q. Curcumin targets miR-134-5p to suppress the progression of colorectal cancer through regulating the CDCA3/CDK1 pathway. *Naunyn Schmiedebergs Arch Pharmacol*. 2024;397(1):109-122.
- Li C, Gao Y, Lu C, Guo M. Identification of potential biomarkers for colorectal cancer by clinical database analysis and Kaplan-Meier curves analysis. *Medicine (Baltimore)*. 2023;102:e32877.
- Zhang P, Kawakami H, Liu W, et al. Targeting CDK1 and MEK/ERK overcomes apoptotic resistance in BRAF-mutant human colorectal cancer. *Mol Cancer Res*. 2018;16:378-389.
- Li J, Wang Y, Wang X, Yang Q. CDK1 and CDC20 overexpression in patients with colorectal cancer are associated with poor prognosis: evidence from integrated bioinformatics analysis. *World J Surg Oncol*. 2020;18:50.
- Li Q, Tong D, Jing X, et al. MAD2L1 is transcriptionally regulated by TEAD4 and promotes cell proliferation and migration in colorectal cancer. *Cancer Gene Ther*. 2023;30:727-737.
- Ding X, Fu Q, Chen W, et al. Targeting of MAD2L1 by miR-515-5p involves the regulation of cell cycle arrest and apoptosis of colorectal cancer cells. *Cell Biol Int*. 2022;46:840-848.
- Venugopal A, Subramaniam D, Balmaceda J, et al. RNA binding protein RBM3 increases β-catenin signaling to increase stem cell characteristics in colorectal cancer cells. *Mol Carcinog*. 2016;55:1503-1516.