Gene Expression **2024** vol. 23(3) | 157–166
DOI: 10.14218/GE.2023.00088

Original Article

Krüppel-like Factor 4, A Potential Therapeutic Agent for Colorectal Cancer: A Bioinformatics Analysis

Esen Çakmak*

Health Services Vocational School, Department of Medical Services and Techniques, Medical Laboratory Techniques Program, Kahramanmaraş Sütçü İmam University, Kahramanmaraş, Turkey

Received: August 10, 2023 | **Revised:** September 20, 2023 | **Accepted:** October 25, 2023 | **Published online:** December 18, 2023

Abstract

Background and objectives: Colorectal cancer is one of the most significant and deadliest malignant tumors among various types of cancers. Due to its generally low overall survival rate, the development of new treatment strategies for early detection and diagnosis, as well as the identification of prognostic markers, has become exceedingly crucial. The molecular mechanism of colorectal cancer remains uncertain and to address this, the aim is to identify key genes, determine in which pathways these genes are involved, explore their interactions with regulatory molecules, and investigate their overall relationship with survival and immune cell infiltration.

Methods: After selecting the databases related to colorectal cancer from the Gene Expression Omnibus database, differentially expressed genes were identified. Gene ontology and pathway analyses were then conducted for these genes, and interaction networks with proteins were constructed. Core genes were identified, and their relationship with regulatory molecules such as miRNAs and transcription factors was examined. Additionally, immune cell infiltration and survival analyses were performed.

Results: As a result of the bioinformatic analyses, 71 differentially expressed genes were identified, which were found to overlap in four distinct microarray datasets. Among these differentially expressed genes, Krüppel-like factor 4 (*KLF4*), *CLCA4*, *GUCA2B*, *GUCA2A*, *LGR5*, *SLC4A4*, *ZG16*, *CA7*, *CA2*, and *GCG* were determined as hub genes. Among the hub genes, *CA2*, *CLCA4*, *SLC4A4*, and *KLF4* genes showed a positive correlation with immune cells in immune cell infiltration analyses. The expression levels of these four genes were also confirmed using data from the Human Protein Atlas database. Additionally, only the *KLF4* gene was associated with poor prognosis in overall survival analyses.

Conclusion: The obtained results suggest that the *KLF4* gene may serve as a potential therapeutic agent.

Introduction

Colorectal cancer (CRC) is one of the leading malignant tumors worldwide, ranking third in cancer-related deaths. Current treatment modalities for CRC patients include surgery, chemotherapy, radiotherapy, and targeted therapy. These treatment strategies reduce the rate of disease recurrence and contribute to an increase in survival rates.[1,2] However, these approaches are most effective when applied during the early stages of the disease.[3] While the 5-year survival rate for CRC patients is approximately 90% in the early stages, it drops below 5% in cases with distant metastases.[4] Limitations in the existing treatment strategies and CRC's high metastatic potential contribute to the detection of the disease at advanced stages, leading to an increase in mortality rates.[5] Therefore, the development of new treatment strategies for early detection and diagnosis, as well as the identification of prognostic markers, has become exceedingly crucial in the pathogenesis of CRC.

In recent years, microarray data analysis and bioinformatic analyses applied in cancer genomics have contributed to the development of new treatment strategies and the discovery of novel biomarkers for cancer pathogenesis.[6,7] Biomarkers play a guiding role in the early diagnosis of cancer and personalized cancer therapy. Bioinformatic applications, allowing the processing of experimental data, have a powerful potential in deciphering the underlying molecular mechanisms of diseases and unraveling complex physiological events.[8]

By comparing gene expression profiles between healthy and cancerous tissues through bioinformatic analyses, valuable information about cancer progression and development can be obtained,

Keywords: *KLF4*; Therapeutic agent; Colorectal cancer; Bioinformatic analysis.
***Correspondence to:** Esen Çakmak, Health Services Vocational School, Department of Medical Services and Techniques, Medical Laboratory Techniques Program, Kahramanmaraş Sütçü İmam University, Kahramanmaraş 46100, Turkey. ORCID: https://orcid.org/0000-0001-8805-3315. E-mail: esencakmak1820@gmail.com
How to cite this article: Çakmak E. Krüppel-like Factor 4, A Potential Therapeutic Agent for Colorectal Cancer: A Bioinformatics Analysis. *Gene Expr* 2024;23(3):157–166. doi: 10.14218/GE.2023.00088.

leading to the identification of new biomarkers for diagnosis and treatment.[6] Bioinformatic analyses are used to define the functions and biological processes of relevant genes involved in cancer pathogenesis and recurrence. Publicly available databases such as The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) are widely used in identifying differentially expressed genes (DEGs) between healthy and cancerous tissues.[9]

Numerous studies have discovered diagnostic biomarkers related to CRC through bioinformatic analyses. These biomarkers include genes, their encoding proteins, long noncoding RNAs, and microRNAs, which have been shown to play regulatory roles in various physiological and pathological processes, including cell proliferation, differentiation, apoptosis, and metastasis.[10–12] However, the molecular mechanism of CRC is still not fully elucidated, and there is still a need for potential biomarkers for early diagnosis and detection.

In the current study, bioinformatic analyses were performed to identify new therapeutic targets for colorectal cancer treatment. Four different datasets from the GEO database were selected, and differentially expressed genes were determined. For the overlapping DEGs in the datasets, gene ontology (GO) functional annotation analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were conducted. Subsequently, a protein-protein interaction (PPI) network was constructed, and hub genes associated with colorectal cancer were selected using Cytoscape software. The relationship between the hub genes and their regulatory molecules, such as miRNAs and transcription factors (TFs), was identified. Immune infiltration analysis, validation of the gene expression levels, and survival analysis were performed to determine the potential of the hub genes as candidate biomarkers.

Materials and methods

Data source and processing

GEO (http://www.ncbi.nlm.nih.gov/geo) is an international genomic database that archives and freely distributes microarrays, next-generation sequencing, and other high-capacity functional genomic data. To obtain microarray data of colorectal cancer cells within *Homo sapiens*, a search of the GEO database was conducted using the keyword "colorectal." Four datasets, namely GSE113513 (14 cancer tissues and 14 normal tissues), GSE21510 (123 cancer tissues and 25 normal tissues), GSE21815 (11 cancer tissues and 5 normal tissues), and GSE32323 (17 cancer tissues and 17 normal tissues), were selected for further analysis. The microarray platforms were GPL15207 for GSE113513, GPL570 for GSE21510, GPL6480 for GSE21815, and GPL570 for GSE32323.

GEO2R (www.ncbi.nlm.nih.gov/geo/geo2r) was used to identify differentially expressed genes (DEGs) between cancer and healthy tissues in the microarray datasets. The selection of DEGs was set with an adjusted *p*-value cutoff of $p < 0.05$ and a log fold change >2. Subsequently, the overlapping DEGs in these datasets were analyzed, and Venn diagrams were drawn using the Venny online tool (https://bioinfogp.cnb.csic.es/tools/venny/index.html).

Gene ontology and pathway analysis

Database for Annotation, Visualization, and Integrated Discovery (DAVID) software (https://david.ncifcrf.gov/) was used to explore the potential functions of overlapping DEGs through GO and KEGG pathway enrichment analyses. The significance threshold was set at $p < 0.05$.

GO analysis was used to define the molecular functions, biological processes, and cellular components related to the DEGs while KEGG pathway analysis was used to examine the reference pathways of the DEGs.

Creating the PPI network

The PPI network was constructed for the DEGs using the Search Tool for the Retrieval of Interacting Genes (STRING) database (https://string-db.org/cgi/). The network was visualized using the Cytoscape tool (www.cytoscape.org), and interactions with a confidence score >0.4 were retained for visualization. Additionally, the DEGs were analyzed using the CytoHubba algorithm to identify hub genes that play a significantly important role (top >5 degrees) within the network.

Regulatory network analysis of hub genes

The hub gene-miRNA and hub gene-TF networks were created to identify the transcriptional and post-transcriptional regulators of hub genes. Significant miRNAs were obtained from the miRTarBase database, while TFs were obtained from the Encode database. The interactions with the highest number of miRNAs and TFs were selected to construct the networks. Moreover, a hub gene-drug network was created by analyzing drug interactions with hub genes. All analyses were performed using the NetworkAnalyst program (https://www.networkanalyst.ca/).

Immune infiltration and hub genes

The TIMER (http://timer.cistrome.org/) program was used to systematically analyze the relationship between immune infiltration and the expression of four selected hub genes. The expression levels of the hub genes and their correlation with immune cell infiltration were evaluated using the "gene module." The TIMER algorithm was utilized to investigate the infiltration levels of CD8+ T cells, B cells, and neutrophils. Risk scores and the correlation between immune infiltration were calculated using Pearson correlation, with a significance threshold set at $p < 0.05$.

The protein levels of the hub genes in the Human Protein Atlas database

The protein levels of the hub genes in tumor and normal tissues in colorectal cancer were evaluated using the Human Protein Atlas (HPA) database (https://www.proteinatlas.org/) which contains immunohistochemistry-based expression data specific to various human tissues.

Survival analysis based on the TCGA database

The impact of the hub genes in tumor and normal tissues on patient survival was analyzed using the Gene Expression Profiling Interactive Analysis (GEPIA) web server based on TCGA data. All parameters were set to their default values, and the cutoff value was set at the median = 50%. $p < 0.05$ was considered statistically significant in this analysis.

Results

Identification of differential genes

In four different microarray datasets related to colorectal cancer (GSE113513, GSE21510, GSE21815, and GSE32323), a total of 165 cancer tissue samples and 61 healthy tissue samples were analyzed. Using the GEO2R program, genes with differential expression were identified in each microarray dataset. In the GSE113513 dataset, 475 genes (366 down-regulated and 109 up-regulated) with differential expression were detected, while in the GSE21510 dataset, 1,349

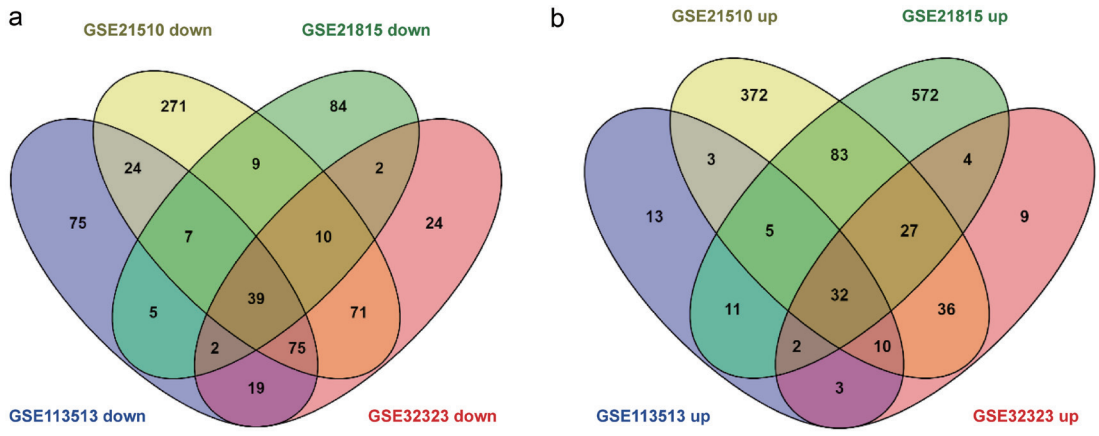

Fig. 1. Venn Diagram of common down-regulated and up-regulated DEGs from four different datasets. DEGs, differentially expressed genes.

genes (661 down-regulated and 688 up-regulated) were detected. In the GSE21815 dataset, 1,031 genes (185 down-regulated and 846 up-regulated) were detected, and finally, in the GSE32323 dataset, 440 genes (301 down-regulated and 139 up-regulated) with differential expression were detected. The DEGs were analyzed using the Venn program to identify the genes common in all four microarray datasets (Fig. 1). In the microarray dataset, a total of 39 DEGs were identified that were down-regulated and overlapped, and 32 DEGs were identified that were up-regulated and overlapped.

Enrichment analysis of DEGs

Detailed enrichment analyses of overlapping DEGs in colorectal cancer tissues were performed using the DAVID online program. The DEGs were enriched in terms of extracellular matrix organization, negative regulation of cell proliferation, plasma membrane, extracellular space, zinc ion binding, and hydrolase activity (Table 1). In KEGG pathway analyses, they were mainly enriched in metabolic pathways and nitrogen metabolism pathways.

PPI analysis of DEGs

A total of 71 genes (39 down-regulated and 32 up-regulated) with differential expression and overlapping in the microarray datasets were subjected to PPI network analysis using the String program. The network was constructed using the Cytoscape program and comprised 70 nodes and 40 edges (Fig. 2a).

Additionally, highly connected hub genes were identified using the Cytoscape program's CytoHubba application. Hub genes with a degree value of more than 6 were identified as Krüppel-like factor 4 (*KLF4*), *CLCA4*, *GUCA2B*, *GUCA2A*, *LGR5*, *SLC4A4*, *ZG16*, *CA7*, *CA2*, and *GCG* (Fig. 2b). Among these hub genes, *LGR5* was up-regulated, while the rest of the hub genes were down-regulated. Expression levels of hub genes were not found to be statistically significant ($p > 0.05$) (Figure 3).

Regulatory molecules of hub genes

Separate interaction networks with miRNAs and TFs were constructed for the hub genes (Fig. 4). Among the hub genes, *KLF4*, *LGR5*, *CLCA4*, and *GUCA2B* showed interactions with miRNAs (Fig. 4a). The miRNA *hsa-mir-335-5p* was identified as the most interacting miRNA, targeting *KLF4*, *GUCA2B*, and *CLCA4*. Other prominent miRNAs included *hsa-mir-124-3p* (targeting *LGR5* and *KLF4*) and *hsa-mir-128-3p* (targeting *KLF4* and *CLCA4*). The genes with the most interactions with transcription factors were found to be *ZG16*, *LGR5*, *KFL4*, and *CA2* (Fig. 4b). The transcription factors *GTF2F1* and *MTA1TF* were found to interact with *CA2* and *KLF4* genes, *ZNF76* and *PPARG* with *ZG16* and *KLF4* genes, and *DMAP1*, *SOX13*, *GATAD2A*, *ZBTB26*, and *SSRP1* with *ZG16* and *LGR5* genes. In the analyses of hub gene and drug interactions, only *CA7* and *CA2* genes were found to interact with the drugs in the database (Fig. 4c). Among the drugs, Acetazolamide, Zonisamide, Diclofenamide, Ellagic acid, Ethoxzolamide and Methazolamide were the common drugs used for both genes.

Immune cell infiltration analysis

The relationships between immune cell infiltration and the expression of hub genes were analyzed. The evaluations were conducted on 458 colorectal cancer (COAD data) patients. The cells considered in the analysis were CD8+, CD4+, B cells, neutrophils, macrophages, and dendritic cells. The results of the immune cell infiltration analysis for all genes are provided in Table 2.

According to the analysis results, *CA2*, *CLCA4*, *KLF4*, and *SLC4A4* genes were found to be more strongly correlated compared to other genes. Based on the Spearman correlation analysis, the *CA2* gene showed a positive correlation with CD8+ (rho = 0.23, p = 1.22e-04), B cells (rho = 0.14, p = 1.70e-02), neutrophils (rho = 0.27, p = 5.03e-06), and dendritic cells (rho = 0.20, p = 7.62e-04). On the other hand, negative correlations were observed between the following pairs: *CLCA4*-Macrophage, *LGR5*-CD8+, and *ZG16*-Macrophage cells (Fig. 5).

The validation of hub genes using HPA

The expression levels of *CA2*, *CLCA4*, *KLF4*, and *SLC4A4* hub genes were validated using immunohistochemical staining data from the HPA database. When compared between normal and tumor tissues, it was observed that the expression levels significantly decreased in tumor tissues (Fig. 6).

Survival analysis

To analyze the prognostic value of the hub genes in colorectal cancer patients, the GEPIA database was utilized. The survival analysis of patients in the GEPIA database was conducted using the Cox PH Model, with a significance threshold of 0.05 for the p-value. Based on the analysis results, only the *KLF4* gene (Hazard Ratio = 0.6; p = 0.0099) was found to be associated with poor prognosis in colorectal cancer patients. However, the other hub

Table 1. GO and KEGG pathway analysis of DEGs in colorectal cancer

Category	GO ID	GO Term	DEGs Count	*p*-value	Associated Genes
GOTERM_BP_DIRECT	GO:0006730	one-carbon metabolic process	5	7,6E-06	*CA1, CA12, CA2, CA4, CA7*
GOTERM_BP_DIRECT	GO:0030198	extracellular matrix organization	7	2,50E-05	*ABI3BP, COL11A1, MMP12, MMP28, MMP7, SPINK5, TGFBI*
GOTERM_BP_DIRECT	GO:0051453	regulation of intracellular pH	3	4,70E-03	*CA2, CA7, SLC4A4*
GOTERM_BP_DIRECT	GO:0008285	negative regulation of cell proliferation	7	5,80E-03	*ABI3BP, CXCL8, KLF4, CDKN2B, FABP6, INHBA, SFRP1*
GOTERM_BP_DIRECT	GO:0030574	collagen catabolic process	3	9,10E-03	*MMP12, MMP28, MMP7*
GOTERM_CC_DIRECT	GO:0005615	extracellular space	18	1,70E-04	*ABI3BP, CXCL8, CA2, CPM, CTHRC1, COL11A1, DPEP1, GCG, HILPDA, INHBA, MMP12, MMP28, MMP7, MUC2, SFRP1, SRPX2, TGFBI, ZG16*
GOTERM_CC_DIRECT	GO:0016324	apical plasma membrane	8	2,80E-04	*CA4, CLCA4, CLDN1, DPEP1, KCNMA1, SCNN1B, SLC6A6, SI*
GOTERM_CC_DIRECT	GO:0005886	plasma membrane	31	1,10E-03	*ATP11A, CD177, VSIG2, ACSL6, ADH1C, BEST2, CDH3, CA12, CA2, CA4, CPM, CLCA4, CLDN1, CLDN8, CNTN3, CPNE8, DPEP1, EPB41L3, LGR5, MUC12, MUC2, KCNMA1, SCARA5, SFRP1, SCNN1B, SLC22A3, SLC4A4, SLC6A6, SI, TGFBI, TRIB3*
GOTERM_CC_DIRECT	GO:0005576	extracellular region	17	1,50E-03	*ABI3BP, CXCL8, CPM, CLCA4, CTHRC1, COL11A1, CNTN3, GCG, GUCA2A,GUCA2B,INHBA, MMP12, MMP7, MUC2, SFRP1, SPINK5, TGFBI*
GOTERM_CC_DIRECT	GO:0031012	extracellular matrix	6	1,80E-03	*COL11A1, MMP12, MMP28, MMP7, MUC2, TGFBI*
GOTERM_MF_DIRECT	GO:0004089	carbonate dehydratase activity	5	2,40E-07	*CA1 CA12 CA2, CA4 CA7*
GOTERM_MF_DIRECT	GO:0016836	hydro-lyase activity	5	7,70E-07	*CA1 CA12 CA2, CA4 CA7*
GOTERM_MF_DIRECT	GO:0008270	zinc ion binding	13	4,40E-05	*KLF4, ADH1C CA1 CA12 CA2, CA4 CA7, CPM DPEP1 MMP12 MMP28 MMP7 NR5A2*
GOTERM_MF_DIRECT	GO:0005201	extracellular matrix structural constituent	5	1,40E-03	*ABI3BP CTHRC1 COL11A1 SRPX2 TGFBI*
GOTERM_MF_DIRECT	GO:0008201	heparin-binding	5	3,30E-03	*ABI3BP CXCL8 COL11A1 MMP7 SFRP1*
KEGG_PATHWAY	hsa00910	nitrogen metabolism	5	8,00E-07	*CA1 CA12 CA2, CA4 CA7*
KEGG_PATHWAY	hsa04964	proximal tubule bicarbonate reclamation	4	1,40E-04	*CA2, CA4 PCK1 SLC4A4*
KEGG_PATHWAY	hsa04972	pancreatic secretion	4	1,10E-02	*CA2, CLCA4, KCNMA1 SLC4A4*
KEGG_PATHWAY	hsa01100	metabolic pathways	13	3,30E-02	*UGT2A3 ACSL6 ADH1C CA1 CA12 CA2, CA4 CA7, CKMT2 GPAT3 PCK1 PSAT1 SI*
KEGG_PATHWAY	hsa03320	PPAR signaling pathway	3	4,50E-02	*ACSL6 FABP6 PCK1*

DEGs, differentially expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; *KLF4*, Krüppel-like factor 4; PPAR, peroxisome proliferator-activated receptor.

genes did not show any statistically significant association with overall patient survival (Fig. 7).

Discussion

The molecular mechanism of colorectal cancer is not yet fully understood.[13] Therefore, it is necessary to elucidate the molecular pathways of CRC and identify potential biomarkers. Bioinformatics analyses are widely used to explore the molecular mechanisms of malignant tumors and are commonly employed to identify candidate biomarkers, aiming to contribute to the diagnosis and treatment process.

In this study, four different datasets were analyzed from GEO. A total of 71 DEGs were identified, with 32 upregulated and 39

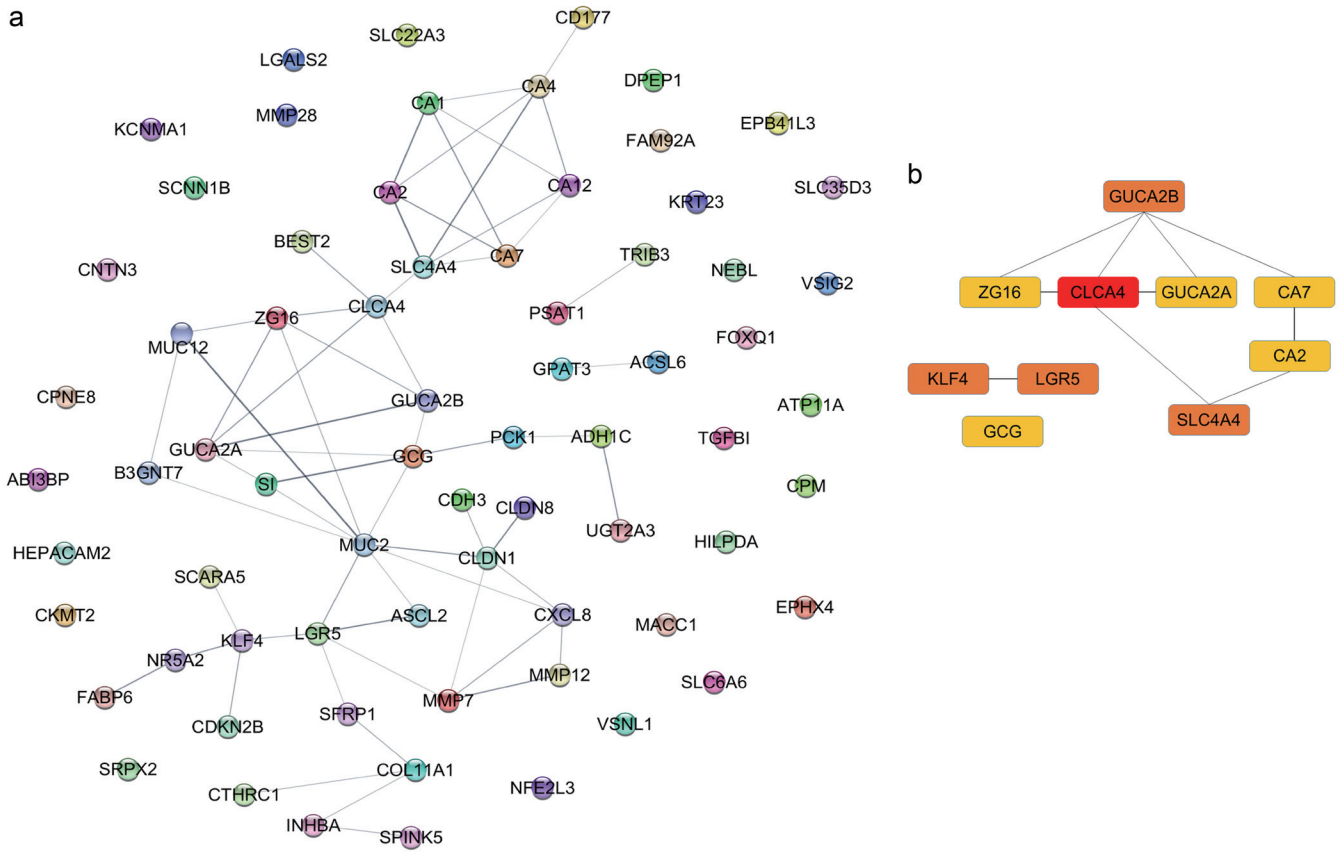

Fig. 2. Protein-protein interaction (PPI) networks. (a) Protein-protein interaction (PPI) network of interacting proteins. (b) PPI network of hub genes in the interaction network. *KLF4,* Krüppel-like factor 4.

downregulated genes, between CRC tissues and normal tissues. Functional enrichment analysis showed that these DEGs were enriched in biological processes such as extracellular matrix organization, negative regulation of cell proliferation, plasma membrane,

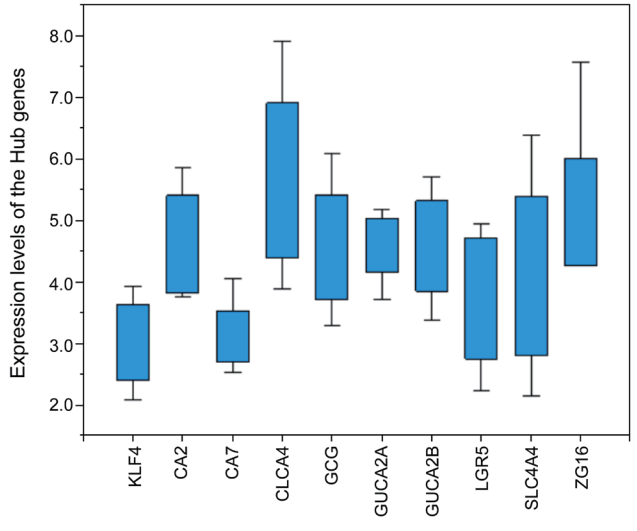

Fig. 3. The expression level of hub genes for four microarray datasets (*p* > 0.05). *KLF4,* Krüppel-like factor 4.

extracellular space, zinc ion binding, and hydrolase activity. Previous studies have reported that these functions play a crucial role in CRC formation and progression, and the findings of this study are consistent with those results.[14–16] Components of the plasma membrane in cancer cells differ significantly from healthy cells, with the plasma membrane composition playing a key role, particularly in the activation of the Wingless and Int-1 signaling pathway. Abnormal activation of the Wingless and Int-1 signaling pathway has been associated with cancer types such as colorectal cancer.[17] In KEGG pathway analysis, DEGs were found to be significantly associated with metabolic pathways and nitrogen metabolism pathways. Metabolic pathways in cancer cells are reprogrammed, and this reprogramming is a dynamic process regulated by oncogenes and tumor suppressor genes. Many metabolic pathways, such as glucose, glutamine, amino acid, serine/glycine, and lipid metabolism, significantly increase the cell proliferation rate in cancer cell management. Moreover, they interact with the microenvironment to change the phenotype.[18] The findings from this study suggest that the DEGs may play an active role in the formation and progression of CRC.

Among DEGs, a total of 10 hub genes were identified, namely *CLCA4*, *GUCA2B*, *GUCA2A*, *KLF4*, *LGR5*, *SLC4A4*, *ZG16*, *CA7*, *CA2*, and *GCG*. Some literature reviews are also consistent with our results, suggesting that these hub genes are associated with CRC.[19–24] The relationship between hub genes and their regulatory molecules, such as miRNAs and TFs, was examined. *KLF4*, *GUCA2B*, *CLCA4*, *ZG16*, *LGR5*, and *CA2* were identified as inter-

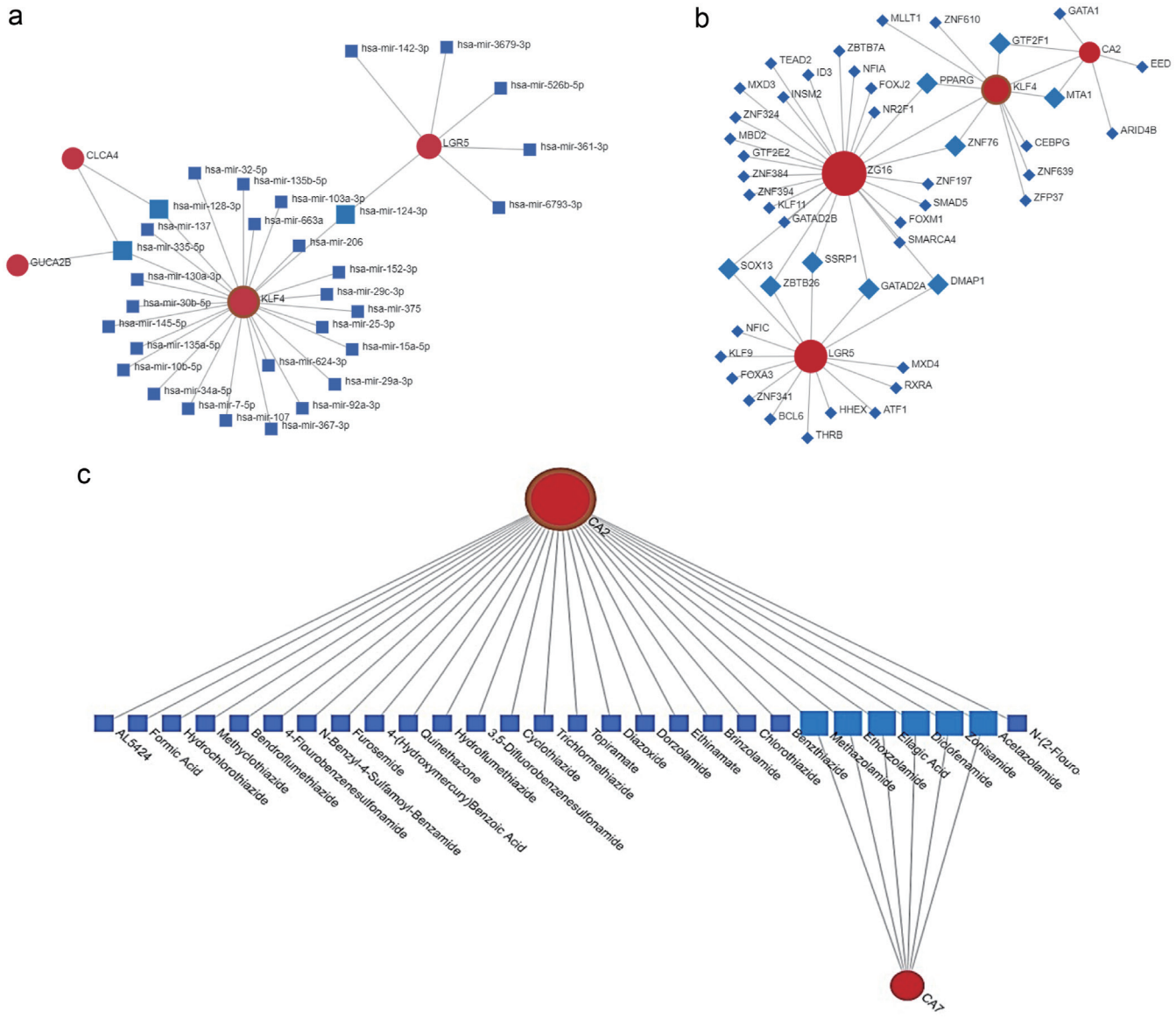

Fig. 4. Networks created via NetworkAnalyst. (a) Hub gene-miRNA (red nodes represent hub genes with the interactions, blue nodes represent miRNAs with the interactions, and big blue nodes represent miRNAs with the most interactions). (b) Hub gene-TF network (red nodes represent hub genes with the interactions, blue nodes represent TFs with the interactions, and big blue nodes represent TFs with the most interactions). (c) Hub gene-drug network (red nodes represent hub genes with the interactions, blue nodes represent drugs with the interactions, and big blue nodes represent drugs with the most interactions). KLF, Krüppel-like factor; TFs, transcription factors.

acting genes with both miRNAs and TFs. Among them, the *KLF4* gene stood out as the gene with the most interactions among both miRNAs and TFs. Notably, *hsa-mir-335-5p*, *hsa-mir-124-3p*, and *hsa-mir-128-3p* were identified as leading miRNAs in the analysis. These miRNAs are known to play important roles in the development and progression of various tumors, including colorectal cancer.[25–28] Additionally, several prominent transcription factors were identified, including *GTF2F1*, *MTA1TF*, *ZNF76*, *PPARG*, *DMAP1*, *SOX13*, *GATAD2A*, *ZBTB26*, and *SSRP1*. In recent years, carbonic anhydrase (CA) isozymes have been used as biomarkers for various diseases.[29] Drugs such as Acetazolamide, Metazolamide, Diclofenamide, Ethoxzolamide, and Zonisamide are available CA inhibitors used in the treatment of various diseases including colorectal cancer.[30] These findings align with the results obtained in the present study.

Immunotherapy has become a promising approach in many cancer types. Invasion of tumor cells into surrounding tissues or metastasis is a consequence of inducing the host's immune response. However, compared to other cancer types, CRC has a lower involvement of immune cells. Due to the heterogeneity of the tumor, the number and distribution of immune cells even vary in different pathological conditions of the same patient. Therefore, it is essential to enhance the effectiveness of immunotherapy in CRC, identify effective treatments, and discover new biomarkers.[31] In the current study, immune cell infiltration analyses were performed, and 4 hub genes (*CA2*, *CLCA4*, *KFL4*, and *SLC4A4*)

Table 2. Correlation results between hub genes and immune cells

Hub genes	Cor	Purity	CD8+	CD4+	B cell	Neutrofil	Macrophage	DC
CA2	rho	−0.312	0.23	−0.042	0.144	0.271	−0.115	0.202
	p	1.22e-10	1.22e-04	4.90e-01	1.70e-02	5.03e-06	5.65e-02	7.62e-04
CA7	rho	−0.098	0.001	0.116	0.077	−0.004	−0.057	0.015
	p	4.72e-02	9.80e-01	5.42e-02	2.02e-01	9.46e-01	3.50e-01	7.99e-01
CLCA4	rho	−0.164	0.131	0.019	0.101	0.168	−0.134	0.117
	p	9.20e-04	3.02e-02	7.53e-01	9.52e-02	518e-03	2.67e-02	5.36e-02
GCG	rho	−0.106	0.061	0.136	0.05	0.082	−0.036	0.082
	p	3.29e-02	3.16e-01	2.42e-02	4.09e-01	1.18e-01	5.54e-01	1.75e-01
GUCA2A	rho	−0.099	−0.045	0.137	0.55	−0.065	0.009	−0.015
	p	4.58e-02	4.53e-01	2.30e-02	3.65e-01	2.85e-01	8.78e-01	8.00e-01
GUCA2B	rho	−0.148	0.078	0.108	0.067	0.2	0.019	0.18
	p	2.81e-03	2.00e-01	7.29e-02	2.66e-01	8.75e-04	7.59e-01	2.79e-03
KLF4	rho	−0.196	0.314	0.1	0.168	0.305	−0.117	0.319
	p	6.63e-05	1.01e-07	9.79e-02	5.22e-03	2.43e-07	5.31e-02	6.07e-08
LGR5	rho	0.043	−0.144	0.127	−0.083	−0.07	0.079	−0.077
	p	3.85e-01	1.68e-02	3.48e-02	1.70e-01	2.46e-01	1.91e-01	2.02e-01
SLC4A4	rho	−0.265	0.232	0.052	0.068	0.248	0.054	0.243
	p	5.44e-08	1.01e-04	3.92e-01	2.61e-01	3.26e-05	3.72e-01	4.62e-05
ZG16	rho	−0.115	−0.031	0.093	0.093	−0.044	−0.158	−0.018
	p	2.05e-02	6.07e-01	1.24e-01	1.24e-01	4.67e-01	8.79e-03	7.64e-01

Cor, corelation; DC, dendritic cells; *KLF4*, Krüppel-like factor 4; rho, spearman rank correlation.

showed a strong positive correlation with immune cells. These results indicate that these cells are present in lower numbers in tumor tissues compared to healthy tissues. The immunohistochemical results of *CA2*, *CLCA4*, *KFL4*, and *SLC4A4* gene expression also confirmed these findings. The *CA2* gene showed a positive correlation with CD8+, Bcell, Neutrophil, and dendritic cell populations. This gene catalyzes the reversible hydration of carbon dioxide and is one of the isozymes of CA that plays an essential role in tissue pH homeostasis and the downregulated *CA2* gene has been shown to be involved in CRC's metastasis mechanism.[32] In another study, the relationship between the *SLC4A4* gene and immune cells infiltrating the tumor was examined, and a positive correlation was found. The importance of *SLC4A4* gene expression, especially its association with immune cells infiltrating the tumor, suggests that it could serve as a biomarker for the diagnosis and a target for treatment in colon adenocarcinoma.[33] In a study focusing on colitis-associated colon cancer, bioinformatics analyses revealed that the *KLF4* gene exhibited significantly low expression in tumor tissues. Additionally, survival analyses associated it with poor prognosis.[24] These findings are consistent with the results of the current study.

KLF4 is a zinc finger transcription factor that participates in cell proliferation, differentiation, and apoptosis. It also regulates the pathogenesis of inflammation and tumor formation.[34] *KLF4* plays a dual role as an oncogene or a tumor suppressor in the development and progression of various cancer types. However, in CRC, it is downregulated in cancer tissues compared to healthy tissues and is known as a tumor suppressor. Thus, low expression of *KLF4* is clearly associated with poor overall survival. In fact, the higher the malignancy of the cancer, the lower the expression of *KLF4*.[35] Therefore, *KLF4* has a high potential as a biomarker for the diagnosis and prognosis in CRC. The data obtained in the current study also indicate that the *KLF4* gene may serve as a potential biomarker in CRC.

Conclusions

This study analyzed multiple datasets from the GEO database to identify differentially expressed genes and interaction networks in CRC. Hub genes were identified, and their potential as biomarkers was explored through immune cell infiltration analysis and overall survival analysis. The bioinformatics analyses highlighted the *KLF4* gene as a strong candidate for potential drug targets and a biomarker in CRC patients. However, further in vitro and in vivo experiments are needed to validate the *KLF4* gene as a molecular biomarker. Considering the importance of early diagnosis and treatment of CRC pathogenesis, small-molecule inhibitors designed to target the *KLF4* gene in pre-cancerous lesions may prove to be an effective strategy. Furthermore, understanding the molecular mechanism of the *KLF4* gene could lay a strong foundation for the development of new treatment approaches in CRC.

Acknowledgments

There is nothing to declare.

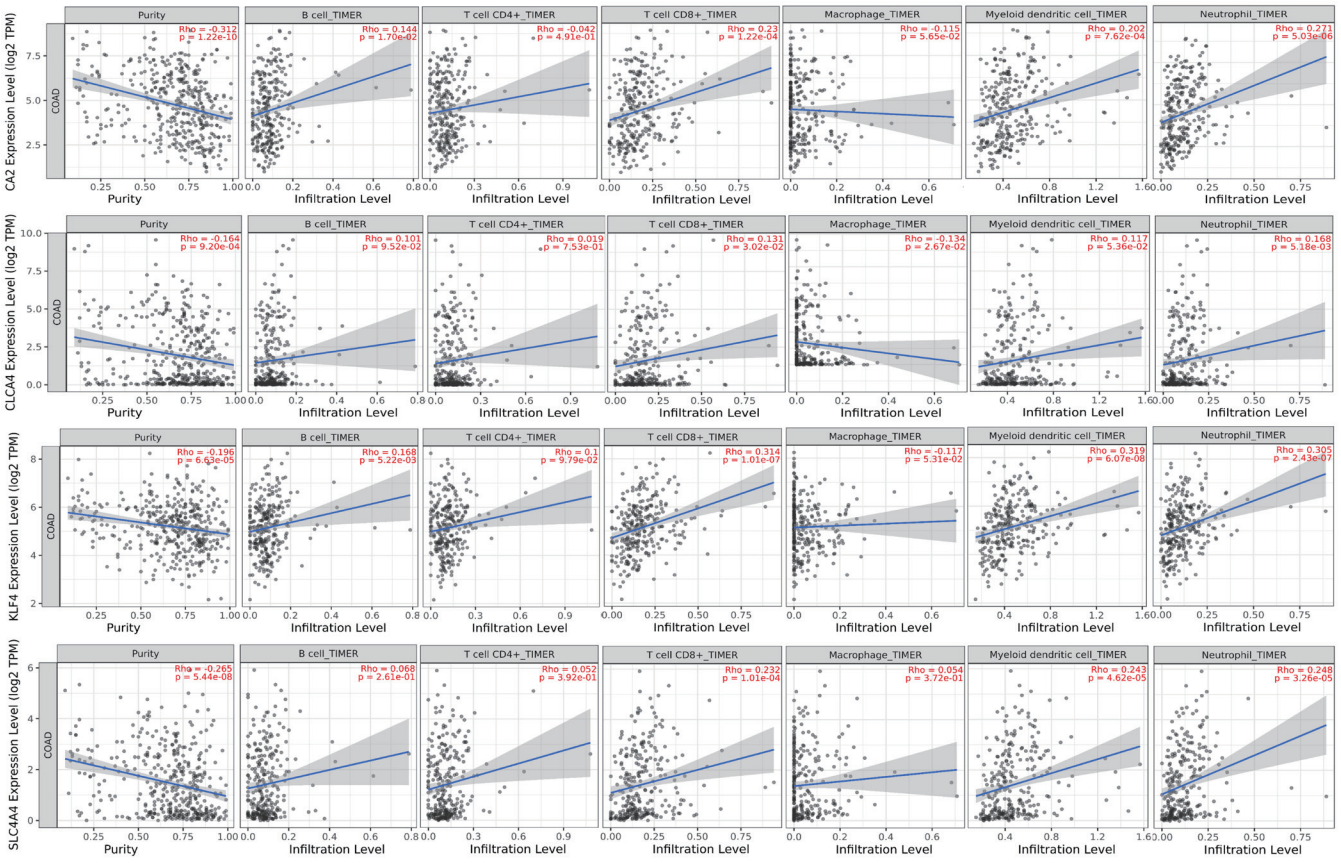

Fig. 5. Expression correlations of the four hub genes (*CA2*, *CLCA4*, *KLF4*, and *SLC4A4*) with immune cell infiltration. *KLF4*, Krüppel-like factor 4.

Funding

The author declared that this study received no financial support.

Conflict of interest

The author has no conflict of interest related to this publication.

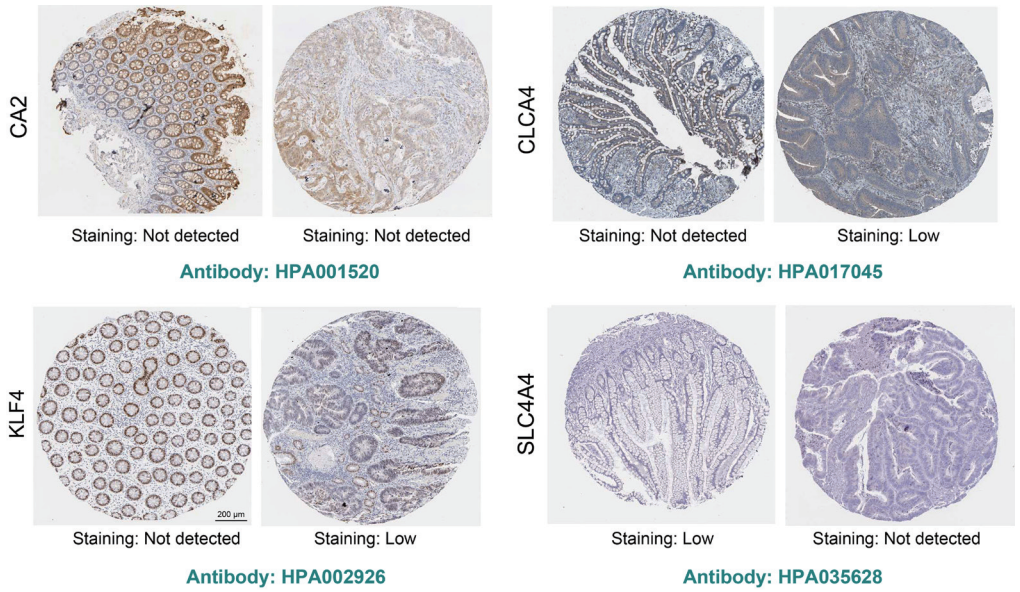

Fig. 6. Validation of the four hub genes (*CA2*, *CLCA4*, *KLF4*, and *SLC4A4*) in normal and tumor tissues using immunohistochemical staining data from the Human Protein Atlas database. *KLF4*, Krüppel-like factor 4.

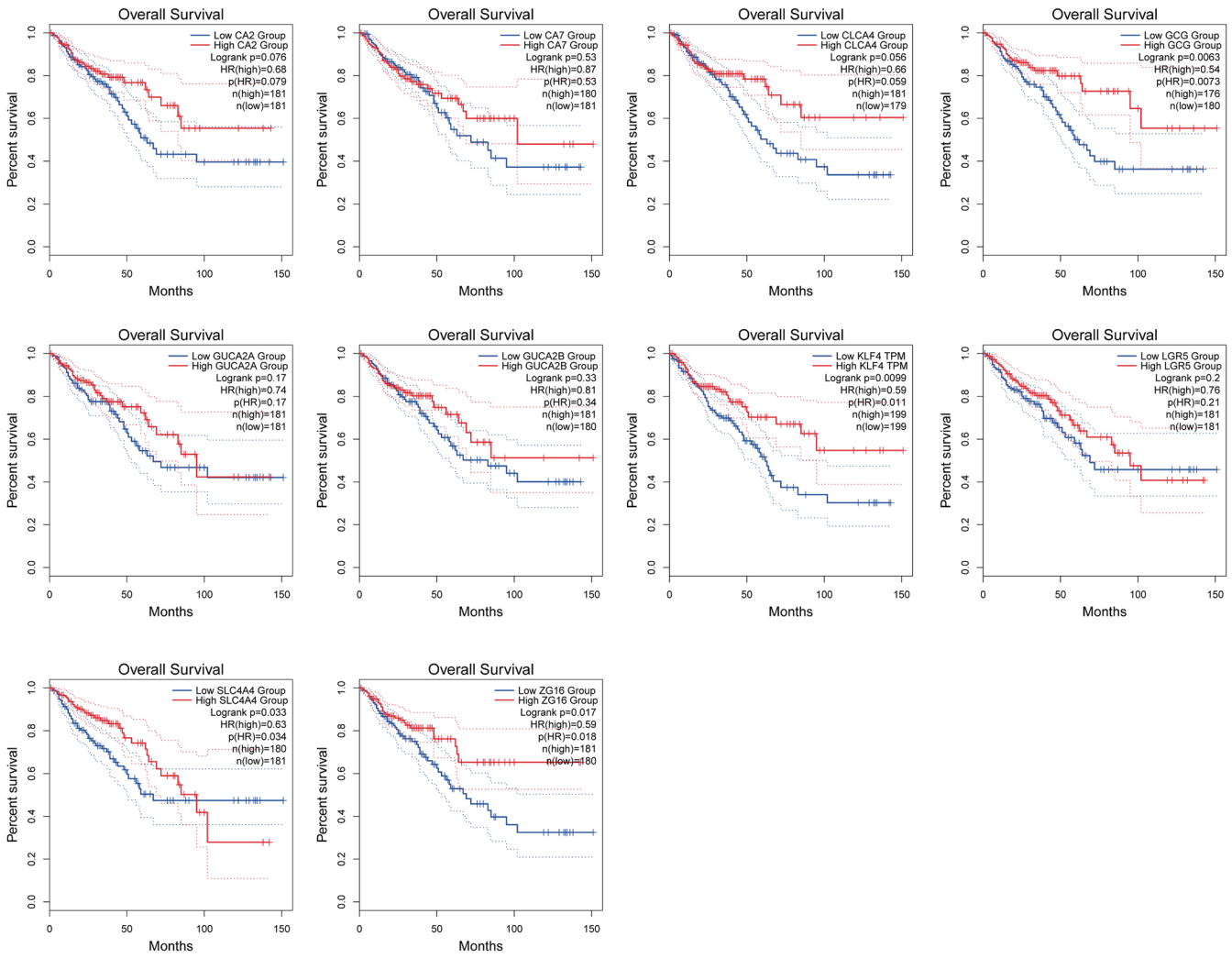

Fig. 7. Overall survival analysis of 10 hub genes in colorectal cancer patients. The red curve represents the high-expression group, and the blue curve represents the low-expression group. *p*-value <0.05. *KLF4*, Krüppel-like factor 4; HR, hazard ratio.

Data sharing statement

No additional data or information is available for this paper.

References

[1] Liu F, Wang Y, Cao Y, Wu Z, Ma D, Cai J, *et al*. Transcription factor B-MYB activates lncRNA CCAT1 and upregulates SOCS3 to promote chemoresistance in colorectal cancer. Chem Biol Interact 2023;374:110412. doi:10.1016/j.cbi.2023.110412, PMID:36812959.

[2] Mortezapour M, Tapak L, Bahreini F, Najafi R, Afshar S. Identification of key genes in colorectal cancer diagnosis by co-expression analysis weighted gene co-expression network analysis. Comput Biol Med 2023;157:106779. doi:10.1016/j.compbiomed.2023.106779, PMID:36931200.

[3] Shahnazari M, Afshar S, Emami MH, Amini R, Jalali A. Novel biomarkers for neoplastic progression from ulcerative colitis to colorectal cancer: a systems biology approach. Sci Rep 2023;13(1):3413. doi:10.1038/s41598-023-29344-y, PMID:36854781.

[4] Park YR, Kim SL, Lee MR, Seo SY, Lee JH, Kim SH, *et al*. MicroRNA-30a-5p (miR-30a) regulates cell motility and EMT by directly targeting oncogenic TM4SF1 in colorectal cancer. J Cancer Res Clin Oncol 2017;143(10):1915–1927. doi:10.1007/s00432-017-2440-4, PMID:28528497.

[5] Li M, Liu Z, Song J, Wang T, Wang H, Wang Y, *et al*. Identification of down-regulated ADH1C is associated with poor prognosis in colorectal cancer using bioinformatics analysis. Front Mol Biosci 2022;9:791249. doi:10.3389/fmolb.2022.791249, PMID:35300114.

[6] Çakmak E. A bioinformatics approach to identify potential biomarkers in non-small cell lung cancer. Cumhur Sci J Cumhur 2022;43(1):6–13. doi:10.17776/csj.976510.

[7] Zhang Z, Peng Y, Dang J, Liu X, Zhu D, Zhang Y, *et al*. Identification of key biomarkers related to epithelial-mesenchymal transition and immune infiltration in ameloblastoma using integrated bioinformatics analysis. Oral Dis 2023;29(4):1657–1667. doi:10.1111/odi.14173, PMID:35226761.

[8] Caglar HO, Duzgun Z. Identification of upregulated genes in glioblastoma and glioblastoma cancer stem cells using bioinformatics analysis. Gene 2023;848:146895. doi:10.1016/j.gene.2022.146895, PMID:36122609.

[9] Meng T, Lan Z, Zhao X, Niu L, Chen C, Zhang W. Comprehensive bioinformatics analysis of functional molecules in colorectal cancer. J Gastrointest Oncol 2022;13(1):231–245. doi:10.21037/jgo-21-921, PMID:35284121.

[10] O'Brien SJ, Bishop C, Hallion J, Fiechter C, Scheurlen K, Paas M, *et*

al. Long non-coding RNA (lncRNA) and epithelial-mesenchymal transition (EMT) in colorectal cancer: a systematic review. Cancer Biol Ther 2020;21(9):769–781. doi:10.1080/15384047.2020.1794239, PMID:32730165.
[11] Du G, Yu X, Chen Y, Cai W. MiR-1-3p Suppresses colorectal cancer cell proliferation and metastasis by inhibiting YWHAZ-mediated epithelial-mesenchymal transition. Front Oncol 2021;11:634596. doi:10.3389/fonc.2021.634596, PMID:33718221.
[12] Li X, Zhang H, Cui T, Wu Y, Wang S. MiR-143-5p inhibits proliferation, invasion, and epithelial to mesenchymal transition of colorectal cancer cells by downregulation of HMGA2. Trop J Pharm Res 2021;20(7):1337–1343. doi:10.4314/tjpr.v20i7.3.
[13] Horaira MA, Islam MA, Kibria MK, Alam MJ, Kabir SR, Mollah MNH. Bioinformatics screening of colorectal-cancer causing molecular signatures through gene expression profiles to discover therapeutic targets and candidate agents. BMC Med Genomics 2023;16(1):64. doi:10.1186/s12920-023-01488-w, PMID:36991484.
[14] Vincan E, Barker N. The upstream components of the Wnt signalling pathway in the dynamic EMT and MET associated with colorectal cancer progression. Clin Exp Metastasis 2008;25(6):657–663. doi:10.1007/s10585-008-9156-4, PMID:18350253.
[15] Ebadfardzadeh J, Kazemi M, Aghazadeh A, Rezaei M, Shirvaliloo M, Sheervalilou R. Employing bioinformatics analysis to identify hub genes and microRNAs involved in colorectal cancer. Med Oncol 2021;38(9):114. doi:10.1007/s12032-021-01543-5, PMID:34390411.
[16] Motafeghi F, Khayambashi B, Mortazavi P, Eghbali M, Salmanmahiny A, Shahsavari R, *et al*. Synergistic effect of selenium / zinc with sulfasalazine on the human colorectal cancer cell line (HT-29). Appl Vitr Toxicol 2023;9(1):3–12. doi:10.1089/aivt.2022.0007.
[17] Azbazdar Y, Karabicici M, Erdal E, Ozhan G. Regulation of wnt signaling pathways at the plasma membrane and their misregulation in cancer. Front Cell Dev Biol 2021;9:631623. doi:10.3389/fcell.2021.631623, PMID:33585487.
[18] La Vecchia S, Sebastián C. Metabolic pathways regulating colorectal cancer initiation and progression. Semin Cell Dev Biol 2020;98:63–70. doi:10.1016/j.semcdb.2019.05.018, PMID:31129171.
[19] Dai GP, Wang LP, Wen YQ, Ren XQ, Zuo SG. Identification of key genes for predicting colorectal cancer prognosis by integrated bioinformatics analysis. Oncol Lett 2020;19(1):388–398. doi:10.3892/ol.2019.11068, PMID:31897151.
[20] Han J, Zhang X, Liu Y, Jing L, Liu YB, Feng L. CLCA4 and MS4A12 as the significant gene biomarkers of primary colorectal cancer. Biosci Rep 2020;40(8):BSR20200963. doi:10.1042/BSR20200963, PMID:32797167.
[21] Zhao ZW, Fan XX, Yang LL, Song JJ, Fang SJ, Tu JF, *et al*. The identification of a common different gene expression signature in patients with colorectal cancer. Math Biosci Eng 2019;16(4):2942–2958. doi:10.3934/mbe.2019145, PMID:31137244.
[22] Wang G, Wang F, Meng Z, Wang N, Zhou C, Zhang J, *et al*. Uncovering potential genes in colorectal cancer based on integrated and DNA methylation analysis in the gene expression omnibus database. BMC Cancer 2022;22(1):138. doi:10.1186/s12885-022-09185-0, PMID:35114976.
[23] Qin L, Zeng J, Shi N, Chen L, Wang L. Application of weighted gene co-expression network analysis to explore the potential diagnostic biomarkers for colorectal cancer. Mol Med Rep 2020;21(6):2533–2543. doi:10.3892/mmr.2020.11047, PMID:32323816.
[24] Huang Y, Zhang X, PengWang, Li Y, Yao J. Identification of hub genes and pathways in colitis-associated colon cancer by integrated bioinformatic analysis. BMC Genom Data 2022;23(1):48. doi:10.1186/s12863-022-01065-7, PMID:35733095.
[25] Wang L, Zhao Y, Xu M, Zhou F, Yan J. Serum miR-1301-3p, miR-335-5p, miR-28-5p, and their target B7-H3 may serve as novel biomarkers for colorectal cancer. JBUON 2019;24(8):1120–1127.
[26] Kamdar RD, Harrington BS, Attar E, Korrapati S, Shetty J, Zhao Y, *et al*. NF-κB signaling modulates miR-452-5p and miR-335-5p expression to functionally decrease epithelial ovarian cancer progression in tumor-initiating cells. Int J Mol Sci 2023;24(9):7826. doi:10.3390/ijms24097826, PMID:37175530.
[27] Wu Q, Zhong H, Jiao L, Wen Y, Zhou Y, Zhou J, *et al*. MiR-124-3p inhibits the migration and invasion of Gastric cancer by targeting ITGB3. Pathol Res Pract 2020;216(1):152762. doi:10.1016/j.prp.2019.152762, PMID:31836324.
[28] Nalla LV, Khairnar A. Empagliflozin mediated miR-128-3p upregulation promotes differentiation of hypoxic cancer stem-like cells in breast cancer. Eur J Pharmacol 2023;943:175565. doi:10.1016/j.ejphar.2023.175565, PMID:36739077.
[29] Zamanova S, Shabana AM, Mondal UK, Ilies MA. Carbonic anhydrases as disease markers. Expert Opin Ther Pat 2019;29(7):509–533. doi:10.1080/13543776.2019.1629419, PMID:31172829.
[30] Kumar S, Rulhania S, Jaswal S, Monga V. Recent advances in the medicinal chemistry of carbonic anhydrase inhibitors. Eur J Med Chem 2021;209:112923. doi:10.1016/j.ejmech.2020.112923, PMID:33121862.
[31] Ge P, Wang W, Li L, Zhang G, Gao Z, Tang Z, *et al*. Profiles of immune cell infiltration and immune-related genes in the tumor microenvironment of colorectal cancer. Biomed Pharmacother 2019;118:109228. doi:10.1016/j.biopha.2019.109228, PMID:31351430.
[32] Lian W, Jin H, Cao J, Zhang X, Zhu T, Zhao S, *et al*. Identification of novel biomarkers affecting the metastasis of colorectal cancer through bioinformatics analysis and validation through qRT-PCR. Cancer Cell Int 2020;20:105. doi:10.1186/s12935-020-01180-4, PMID:32256214.
[33] Chen X, Chen J, Feng Y, Guan W. Prognostic value of SLC4A4 and its correlation with immune infiltration in colon adenocarcinoma. Med Sci Monit 2020;26:e925016. doi:10.12659/MSM.925016, PMID:32949121.
[34] Lee E, Cheung J, Bialkowska AB. Krüppel-like factors 4 and 5 in colorectal tumorigenesis. Cancers (Basel) 2023;15(9):2430. doi:10.3390/cancers15092430, PMID:37173904.
[35] He Z, He J, Xie K. KLF4 transcription factor in tumorigenesis. Cell Death Discov 2023;9(1):118. doi:10.1038/s41420-023-01416-y, PMID:37031197.