RESEARCH ARTICLE

Editorial Process: Submission:05/27/2025 Acceptance:01/15/2026 Published:01/22/2026

Identification and Validation of Prognostic Biomarker Signatures Associated with Overall Survival in Colorectal Cancer: Evidence from Bioinformatics Analysis and an *in vivo* Study

Mohammad Bagher Jahantab[1], Nasrin Amiri-Dashatan[2], Mostafa Rezaei-Tavirani[3], Mehdi Koushki[4,5]*

Abstract

Background: Colorectal cancer (CRC) is one of the most common gastrointestinal cancers worldwide. Thus, genes targeting is useful for the prognosis and therapy of CRC. This study aimed to identify a promising and valuable signature model of prognostic biomarkers associated with overall survival (OS) in CRC. **Methods:** Two mRNA microarray datasets (GSE18105 and GSE113513) from the Gene Expression Omnibus (GEO) database were screened to extract key genes from differentially expressed genes (DEGs) to establish a multiscale embedded gene co-expression network, protein-protein interaction network, and survival analysis. Univariate Cox analysis was conducted to construct a prognostic signature for OS using Kaplan–Meier analysis. Then, we constructed and analyzed the protein-protein interaction network using STRING and Cytoscape, respectively to establish the key genes. Finally, the selected potential prognostic genes were validated in tissue samples of CRC by quantitative real-time PCR (qRT-PCR). **Results:** In the present study, among 340 identified DEGs, four key genes (*SPP1, CHEK1, KIF18A*, and *MAD2L1*) were detected. The prognostic gene signature model demonstrated strong performance in the prognosis of CRC (AUC > 0.9). Moreover, the four key genes were also used to construct a risk-score prognostic model for OS and the findings showed that the prognostic gene signature model was highly effective in predicting the OS in CRC patients. The Gene Ontology (GO) enrichment analysis indicated the key genes were significantly associated with several CRC-related signaling pathways such as calcium-independent cell-cell adhesion. Finally, the results of qRT-PCR showed that the upregulation of *SPP1*, *CHEK1*, *KIF18A*, and *MAD2L1* was associated with poor prognosis and served as risk factors for CRC patients compared to controls. **Conclusion:** The findings of the present study provided a set of four key genes with valid clinical utility that can serve as an alternative tool for prognosis and identification of new targets in CRC treatment.

Keywords: Colorectal Cancer- Protein Interaction Network- Gene Ontology- Survival Analysis- Biomarker

Asian Pac J Cancer Prev, **27 (1)**, 163-174

Introduction

Colorectal cancer (CRC) is considered the third most common cancer worldwide, and it is predicted to have more than 1.1 million deaths by 2030 [1]. On the other hand, CRC is one of the most prevalent gastrointestinal malignancies [2]. Although significant improvements have been made in the diagnosis and treatment of CRC, global findings indicate that its high mortality remains unsatisfactory for patients with CRC due to cancer recurrence, metastasis, and resistance to radiotherapy and chemotherapy [1]. Therefore, further research is needed to decipher the molecular mechanisms and novel biomarkers associated with the development and progression of CRC. Therefore, this research can provide new perspectives and insights to identify new diagnostic and therapeutic targets and monitor disease progression.

CRC is defined as a heterogeneous disease, which is mostly caused by genetic alterations and interactions of environmental factors [3]. Based on literatures, several genes and molecular pathways such as *RACK1* and the lncRNA BCAR4 play an important role in the development of CRC [4, 5]. For instance, it has been reported that the expression of *RACK1* is significantly up-regulated in

*[1]Clinical Research Development Unit, Shahid Jalil Hospital Yasuj University of Medical Sciences, Yasuj, Iran. [2]Zanjan Metabolic Diseases Research Center, Health and Metabolic Diseases Research Institute, Zanjan University of Medical Sciences, Zanjan, Iran. [3]Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [4]Cancer Gene Therapy Research Center, Zanjan University of Medical Sciences, Zanjan, Iran. [5]Department of Clinical Biochemistry, School of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran. *For Correspondence: mehdikoushki89@gmail.com*

cancerous tissues compared to adjacent normal tissues in patients with CRC. The increased expression of *RACK1* mainly leads to increased cell proliferation, migration, and cell invasion [4]. Over the recent years, bioinformatics analysis has served as a valuable tool for several life science applications that could be applied for early diagnosis, prognosis, and treatment of CRC. In addition, advances in high throughput technologies have led to the generation of an unprecedented volume of biological data at different levels of genomics, transcriptomics and proteomics. Gene microarray profiling, a high-throughput method to detect mRNA expression in tissues, has dramatically become a promising tool in medical oncology [6]. By analyzing the differential gene expression between tumoral and control tissues, a better understanding about the molecular pathogenesis of various cancers including CRC can be achieved, which facilitates the identification of potential target genes and signaling pathways for precise therapy. Despite detailed investigations to identify novel targets for CRC management, there is limited comprehensive analysis of gene expression profiles [7, 8]. This analysis can lead to introduce key genes and signaling pathways involved in the CRC progression. Therefore, here, we analyzed GEO datasets to identify potential biomarkers and signaling pathways involved in the progression of CRC and the identified key genes via bioinformatics analysis were then validated by Real Time- PCR.

Materials and Methods

System biology approaches and bioinformatics analysis

Microarray data analysis

After searching NCBI/GEO series (http://www.ncbi.nlm.nih.gov/geo/) using keywords ["Colorectal Cancer" AND "Microarray" AND "Gene OR Transcript OR mRNA"], 1725 results were obtained. By restricting the results to "Homo sapiens" and "Tissue", 327 data series remained. From the 327 series, 6 were selected and analyzed with GEO2R platform and finally 2 datasets with the highest number of meaningful results were picked up. These included 2 gene transcript datasets, *GSE18105* and *GSE113513*. The expression datasets were all normalized with *GEO2R* and the statistically differentiating genes (*DEGs*) were obtained by the cutoff criteria of p-value<0.05 and $|\text{Fold-change}|>1.5$. The details of each dataset including the number of samples and microarray platform type are summarized in Table 1. The volcano plots showing the significant genes are presented in Figure 1.

Integration of datasets results

The differentially expressed genes were analyzed using venny online platform [11] to find the intersection between the datasets. Accordingly, 340 common genes between the gene microarrays were obtained.

Construction of protein-protein interaction network

Common genes were searched in STRING database

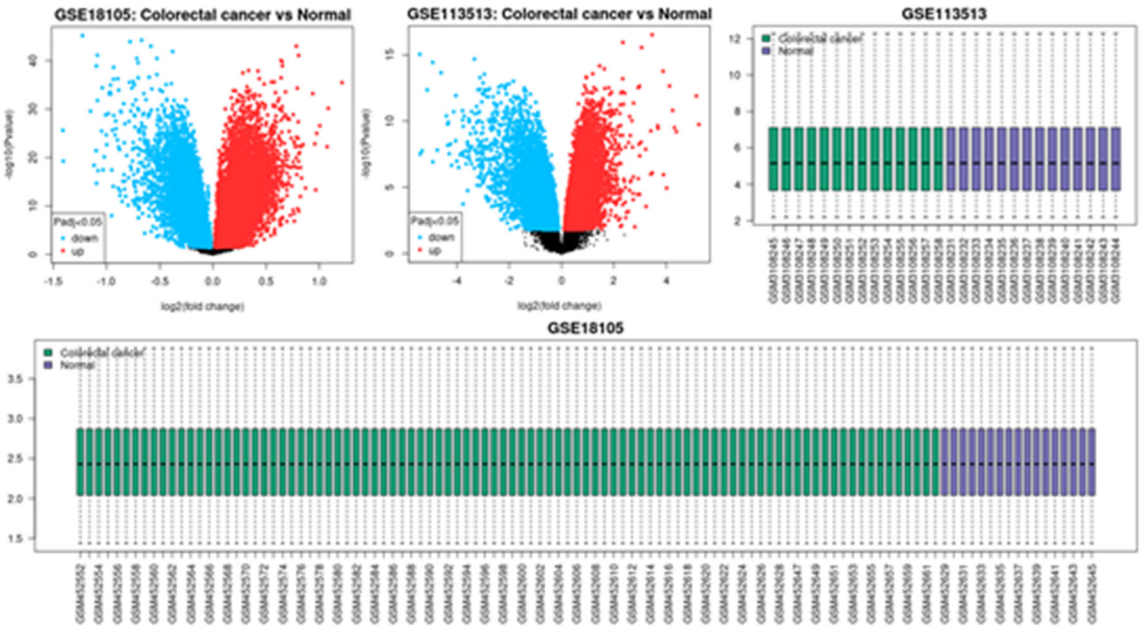

Figure 1. Normalized GEO Datasets and the Volcano Plots Showing the Significant up/down Regulated Genes in Each Dataset

Table 1. Details of the Selected Microarray Platforms

GEO series	Platform	#Tumor samples	#Control samples	Ref.
GSE18105	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	94	17	Matsuyama T. et al. [9]
GSE113513	GPL15207 [PrimeView] Affymetrix Human Gene Expression Array	14	14	Shen A. et al. [10]

to find all the possible interactions of the 340 *DEGs* and construct the protein-protein interaction network. The resulting network was further analyzed in Cytoscape software.

Gene ontology and Pathway enrichment analysis

ClueGO plug-in in Cytoscape was used to find the significant enriched biological processes (BP), molecular functions (MF), cellular components (CC), and KEGG pathways in the DEG network. The Bonferroni-corrected p-value<0.05 was set as statistically significant.

Hub or key genes selection

The constructed network was analyzed in Cytoscape and the nodes with the highest numbers of interactions were selected as the hub genes. Hub nodes usually play key roles in the regulation of the network-related pathways. These nodes might also serve as potential biomarkers of the diseases.

Survival analysis

Survival analysis was performed for the hub genes in the network. To check the possibility of the *DEGs* correlation with overall survival in colorectal cancer patients, the Human Protein Atlas (HPA) and the UALCAN databases were used. The HPA is an open-source online platform for mapping of human proteins in tissues, cells, and organs by using the data derived from omics studies and antibody-based imaging. In the "pathology" section of HPA, we sought the impact of the DEG expression levels on survival of colorectal cancer patients. The proteins with meaningful correlation with overall survival were selected according to the Kaplan-Meier survival curves. The cutoff p-value was set 0.05.

ROC analysis

After selection of the hub genes which were involved in survival of colorectal cancer patients, the ROC curves were plotted in GraphPad prism 8.0, according to the nodes normalized expression data. The ROC curves are shown in Figure 2. The p-value <0.05 and the area under the curve (AUC) > 0.90 was set as the cutoff for selection of potential biomarkers.

Experimental validation

Demographic characteristics of patients

Twelve CRC patients referred to Ayatollah Mousavi Hospital of Zanjan University of Medical Sciences between November 2022 and September 2023 provided the cancerous colon and rectal tissues and adjacent normal tissues. Patients with diabetes, autoimmune disease, cardiovascular disease, and chemotherapy were excluded from the study. This study was approved by the Ethic Committee of Zanjan University of Medical Sciences. Written informed consent was obtained from each patient prior to precipitation. (Ethical Code: IR.ZUMS.REC.1401.089). Demographic information of included patients was presented in Table 2.

Table 2. Demographic Characteristics of Patients Used for RT-qPCR Validation.

Variable	Number (n= 12)
Sex (M, F)	(6, 6)
Age	61.6 ± 16.9
Tumor size (cm)	
< 7	6
> 7	6
TNM Stage	N (%)
I	4 (33.3)
II	3 (25)
III	4 (33.3)
IV	1 (8.3)

M, male; F, female, Variable of age reported as mean ± SD

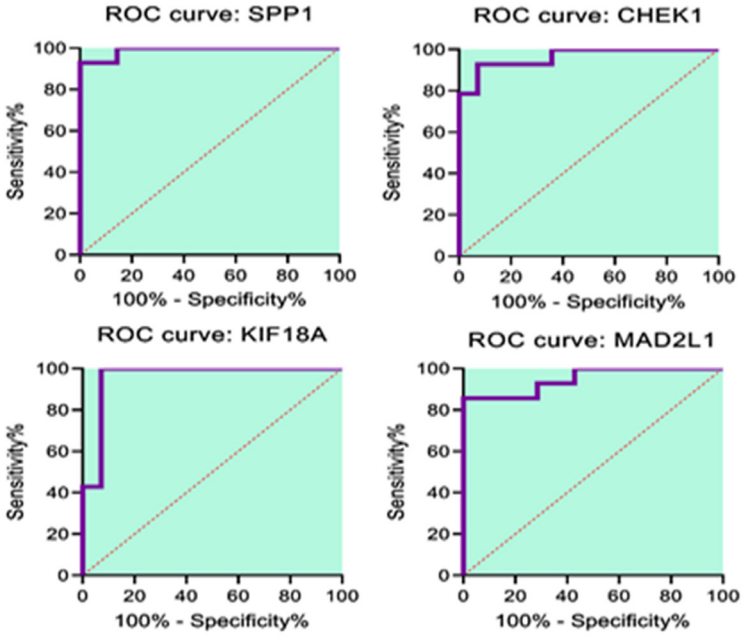

Figure 2. ROC Curve Analysis Results for Genes with Area under the Curve (AUC) ≥ 0.9 and p-vale<0.05 (including *SPP1, CHEK1, KIF18A,* and *MAD2L1*)

Validation of the selected genes by qRT-PCR

To validate the bioinformatics analysis results, we employed reverse-transcription quantitative-real-time PCR (qRT-PCR) analyses to assess expression level of *SPP1*, *CHEK1*, *KIF18A*, and *MAD2L1* on CRC tissues and adjacent normal. Total RNA was extracted using RiboEx reagent (GeneAll Biotech, Korea) as described by the manufacturer. The quality (based on the appearance of the spectra) and quantity of RNA were assessed using Nano Drop (ND-1000, Thermo Scientific Fisher, US). Three independent RNA samples were used for each real-time PCR experiment. Complementary DNA (cDNA) was synthesized from 3 µg of total RNA using RevertAid™ First Strand cDNA Synthesis Kit (Thermo Fisher Scientific Inc.) following the manufacturer's instructions. The sequence of the primers is following: I) *SPP1*: 5′- CGAGGTGATAGTGTGGTTTATGG-3′ (F); 5′-GCACCATTCAACTCCTCGCTTTC-3′ (R), II) *CHEK1*: 5′-GTGTCAGAGTCTCCCAGTGGAT-3′ (F); 5′-GTTCTGGCTGAGAACTGGAG TAC-3′ (R), III) *KIF18A*: 5′-CAGTTCAGCCTATTCCTT-3′ (F), 5′ TATCACTGTTTATGTTT GAGC-3′ (R), IV) *MAD2L1*: 5′-TTGAGTGTGACAAGACTGCAAAAG-3′ (F); 5′-CAGTGGCAGAAATGTCACCGTAG-3′ (R), and Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) gene (Biosystems, Life Technologies, USA) was used as internal control. The relative expression value of each gene was determined based on the threshold cycle (Ct) value of the target genes, normalized to that of reference genes (*GAPDH*) using the $2^{-\Delta\Delta Ct}$ method. RT-qPCR results were analyzed using SPSS, and plots were designed with GraphPad Prism Software. Data are reported as the mean ± SD and comparison between groups were analyzed by student's t-test. A p-value <0.05 was considered statistically significant.

Results

Differentially expressed genes (DEGs)

To find the differentially expressed genes, the intersection between datasets was extracted from the Venn diagram. The gene expression microarray GSE18105 resulted in 569, where GSE113513 resulted in 1946 significantly altered genes respectively according to the cutoff criteria. The intersection between these 2 datasets yielded in 340 common genes (Figure 3). The common *DEGs* were used for further analysis. Supplementary Table 1 provides the detailed names of the common *DEGs*.

Network analysis and hub genes

The protein-protein interaction (PPI) network consisted of 340 nodes and 783 edges (Figure 4). The network was analyzed in Cytoscape software and the hub nodes with the highest connectivity degrees were selected as potential biomarkers for colorectal cancer. The top 20 hub *DEGs* included *CD44, CDK1, ASPM, BMP2, CHEK1, KIF23, BUB1, MAD2L1, SPP1, WNT5A, CXCL12, TNFSF11, NUF2, SLC26A3, WNT2, CLCA1, KIF18A, LEF1, MMP3*, and *RFC3* (Table 3). Among these hub nodes, *CD44* and *SPP1* also served as bottlenecks in the network with the highest betweenness centrality.

Gene ontology

Based on ClueGO Cytoscape plugin, the meaningful enriched gene ontology (GO) terms and KEGG pathways were determined for the *DEGs* (Table 4). The most significant biological processes included calcium-independent cell-cell adhesion via plasma membrane cell-adhesion molecules, response to prostaglandin, collagen catabolic process, epithelial to mesenchymal transition, interleukin-1 beta secretion, atrial cardiac muscle tissue morphogenesis, and mesenchymal cell development, negative regulation of cell proliferation involved in contact inhibition, creatine metabolic process, and excretion. The top enriched molecular functions included beta-amyloid binding, glycolipid binding, and alcohol dehydrogenase (NAD) activity. Golgi lumen was the most significant cellular component. KEGG pathways enrichment results also showed that Nitrogen metabolism, pancreatic secretion, and bile secretion were the most meaningful pathways related to colorectal cancer.

Survival analysis

To assess the relationship between the proposed biomarkers and overall survival of colorectal cancer patients, survival data were extracted from the databases

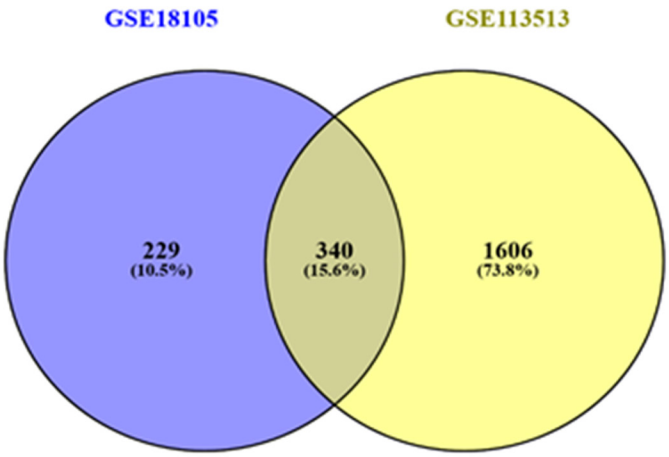

Figure 3. Venn Diagram Showing Common Genes between the Selected GEO Datasets.

Table 3. Top 20 Genes with the Highest Degrees in the Protein-Protein Interaction Network. (↑: increased in tumor group, ↓: decreased in tumor group).

No.	ID_REF	P-value	FC	Direction of gene expression (Tumor/control)	Hub Degree
1	CD44	6.91E-09	2.6	↑	30
2	CDK1	0.000000115	2.7	↑	25
3	ASPM	0.0000115	2.3	↑	22
4	BMP2	0.00000246	2.8	↓	22
5	CHEK1	0.00000391	2.0	↑	22
6	KIF23	0.00000576	2.4	↑	22
7	BUB1	0.0000164	2.0	↑	21
8	MAD2L1	0.00000129	2.4	↑	21
9	SPP1	0.017	2.6	↑	20
10	WNT5A	0.0000381	2.5	↑	19
11	CXCL12	1.2E-10	19	↓	18
12	TNFSF11	0.00246	2.1	↑	18
13	NUF2	0.000000445	2.4	↑	17
14	SLC26A3	2.97E-08	9.4	↓	17
15	WNT2	0.00000011	2.0	↑	17
16	CLCA1	0.000188	13.5	↓	16
17	KIF18A	0.000000722	2.1	↑	16
18	LEF1	0.000012	3.5	↑	16
19	MMP3	0.0000343	7.4	↑	16
20	RFC3	8.5E-09	3.0	↑	16

FC. Fold Change

Human Protein Atlas and UALCAN (Figure 5). Among the putative gene/protein markers, *CLCA1, WNT5A,* and *CHEK1* showed correlation with overall survival of colorectal cancer patients according to Human Protein Atlas. *SPP1*, LEF1, *KIF18A*, and *MAD2L1* also showed correlation with overall survival of patients with colon and/or rectal adenocarcinomas according to UALCAN database. High expression of *CLCA1, WNT5A*, and *CHEK1* are favorable to survival in colorectal cancer. This means the overall survival of patients with CRC decreases as the expression of these 3 genes goes up. Higher expression of *MAD2L1* and *KIF18A* showed favorable and *LEF1* showed unfavorable correlation with rectal adenocarcinoma according to UALCAN database. *SPP1*

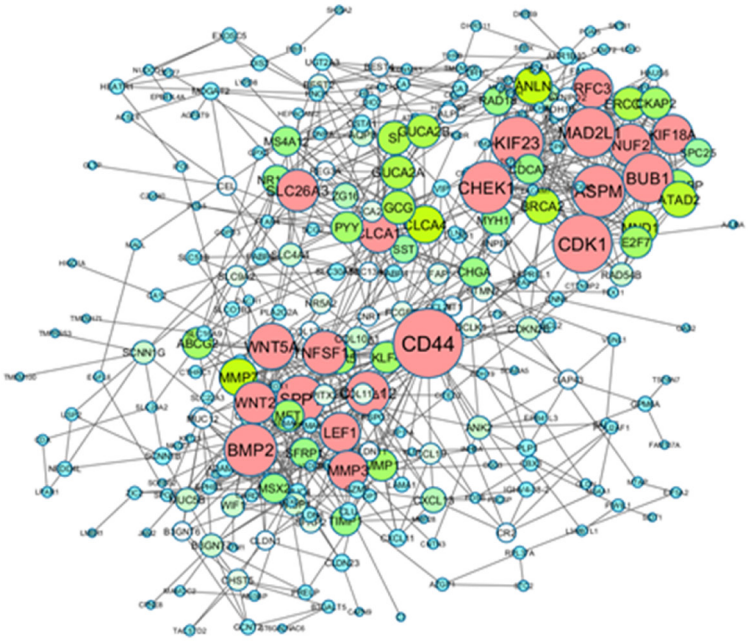

Figure 4. The Protein-Protein Interaction Network, Constructed in Cytoscape based on All the Possible Interactions of DEGs. Larger circles denote higher degree hub genes.

Table 4. Gene Ontology and KEGG Pathway Enrichment Results

GO ID	GO Term	Bonferroni-corrected P-value	Associated Genes (%)	Associated Genes Found
		Biological Process		
GO:0016338	Calcium-independent cell-cell adhesion via plasma membrane cell-adhesion molecules	260.0E-6	25.00	*BMP2, CLDN1, CLDN11, CLDN23, CLDN5, CLDN8*
GO:0034694	Response to prostaglandin	47.0E-3	8.33	*CCL19, PTGDR, SFRP1*
GO:0030574	collagen catabolic process	6.8E-3	9.88	*CEL, COL10A1, COL11A1, COL12A1, FAP, MMP1, MMP3, MMP7*
GO:0001837	epithelial to mesenchymal transition	2.4E-3	8.85	*BMP2, GCNT2, LEF1, MSX2, OLFM1, SFRP1, SFRP2, TMEM100, WNT2, WNT5A*
GO:0050702	interleukin-1 beta secretion	25.0E-3	8.11	*CCL19, NLRP2, WNT5A*
GO:0055009	atrial cardiac muscle tissue morphogenesis	8.4E-3	50.00	*PITX2, PROX1, WNT2*
GO:0014031	mesenchymal cell development	85.0E-6	8.06	*BMP2, CITED2, EDNRA, GCNT2, LEF1, MSX2, OLFM1, PITX2, SEMA6A, SEMA6D, SFRP1, SFRP2, TMEM100, WNT2, WNT5A*
GO:0060244	negative regulation of cell proliferation involved in contact inhibition	290.0E-6	57.14	*CEL, DACH1, FAP, SRPX*
GO:0006600	creatine metabolic process	45.0E-3	27.27	*CKB, CKMT2, GHR*
GO:0007588	excretion	30.0E-3	8.86	*ANPEP, GUCA2B, KCNMA1, NEDD4L, SCNN1B, SCNN1G, SLC26A3*
		Molecular Function		
GO:0001540	beta-amyloid binding	47.0E-3	8.33	*BCHE, LDLRAD3, OLFM1*
GO:0051861	glycolipid binding	49.0E-3	16.00	*CEL, DPEP1, GLTP, LAMA1*
GO:0004022	alcohol dehydrogenase (NAD) activity	20.0E-3	37.50	*ADH1B, ADH1C, DHRS9*
		Cellular Component		
GO:0005796	Golgi lumen	3.5E-3	9.47	*CHGA, MUC12, MUC2, MUC4, MUC5B, OGN, PRELP, WNT5A, ZG16*
		KEGG Pathway		
GO:0000910	Nitrogen metabolism	740.0E-6	29.41	*CA1, CA12, CA2, CA4, CA7*
GO:0004972	Pancreatic secretion	590.0E-6	10.42	*ATP2A3, CA2, CEL, CLCA1, CLCA4, KCNMA1, PLA2G2A, PRKCB, SLC26A3, SLC4A4*
GO:0004976	Bile secretion	17.0E-3	9.86	*ABCG2, AQP8, CA2, NR1H4, SLC4A4, SLC51B, SLCO1B3*

expression also had favorable correlation with both colon and rectum adenocarcinomas according to UALCAN database. *SPP1* and *CHEK1* can serve as prognostic markers for colorectal cancer where all *CHEK1*, *SPP1*, *MAD2L1*, and *KIF18A* can serve as prognostic markers for rectum adenocarcinoma.

ROC curves

ROC curves were plotted using GraphPad prism software. The results are shown in Figure 2. ROC (Receiver-Operating Characteristic) analysis is a valuable tool for evaluating the performance of a diagnostic test or the accuracy of a statistical model. It is a plot of Sensitivity (the true positive rate) in function of 100-Specificity (the false positive rate) for different cut-off points of a variable. The Area Under the curve (AUC) is a measure of how well a variable can distinguish between diagnostic groups (here, between gastric cancer and control samples). A curve closer to the upper left corner, shows higher overall accuracy of the test. We performed ROC curve analysis for top hub genes which had correlation with the overall survival of colorectal cancer patients according to human protein atlas and UALCAN databases. AUC cutoff for selection of putative biomarkers was 0.90. According to the results, *SPP1*, *CHEK1*, *KIF18A*, and *MAD2L1* showed the highest accuracy and had AUC values more than 90 percent. These nodes are proposed as prognostic markers for colorectal cancer (Table 5). They might also have important biological roles in colorectal cancer pathogenesis. The ROC curve analysis results are shown in Table 5.

Real-time qPCR Validation Results

The expression of *CHEK1*, *SPP1*, *MAD2L1*, and *KIF18A* were evaluated in CRC tissues compare to adjacent normal tissues. The results indicated that the expression levels of all genes were up-regulated significantly in tumor tissues compared to normal tissues (Figure 6).

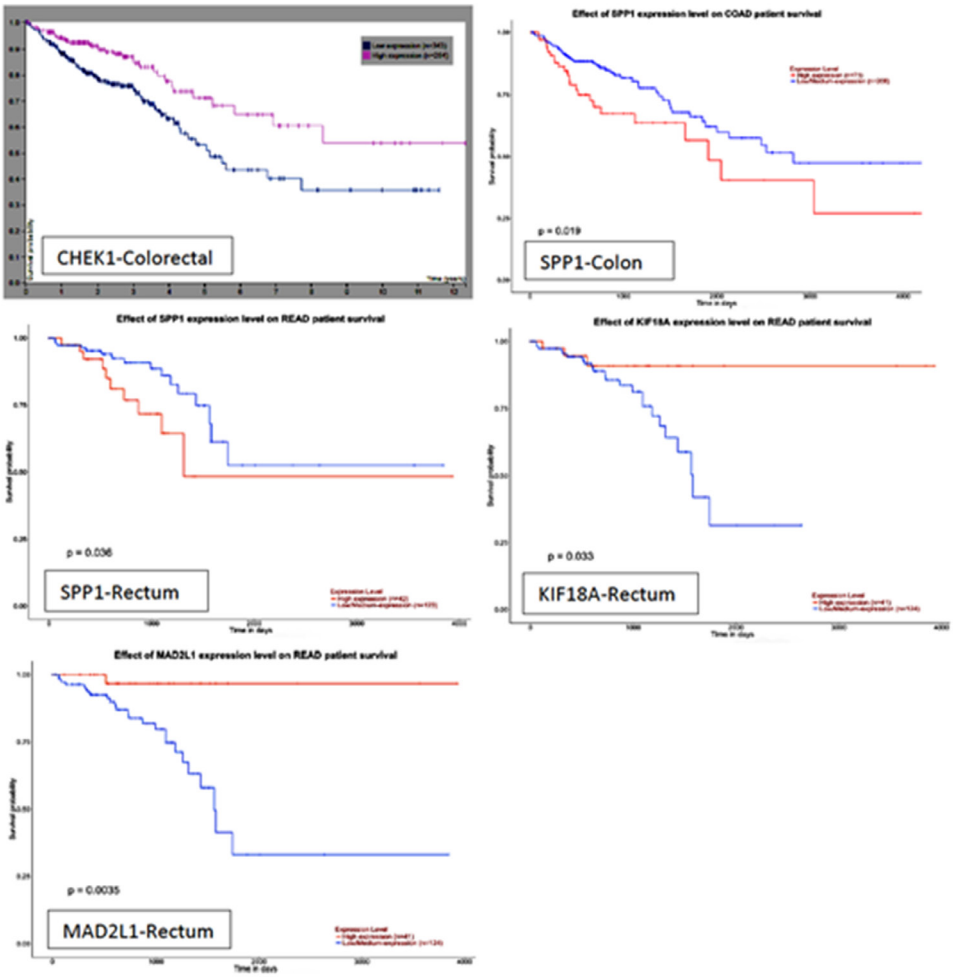

Figure 5. Kaplan-Meier Survival Curves of *CHEK1, SPP1, MAD2L1*, and *KIF18A* from the Human Protein Atlas and UALCAN database. Expression of *SPP1* shows unfavorable correlation, where the expression of *CHEK1, KIF18A,* and *MAD2L1* show favorable correlation with the survival. *SPP1* and *CHEK1* can serve as prognostic markers for colorectal cancer where all *CHEK1, SPP1, MAD2L1,* and *KIF18A* can serve as prognostic markers for rectum adenocarcinoma.

Discussion

Colorectal cancer is associated with broadly high mortality and morbidity. In addition, the failure to have early screening and diagnosis in patients with CRC leads to poor overall survival rates and prognosis. Therefore, there is an urgent requirement to identify sensitive and specific biomarkers for the management of CRC. Bioinformatics analyses enable to detect the gene expression alterations during CRC and could be an effective approach to identify novel biomarkers. An initial aim of this project was to identify a panel of prognostic

Table 5. A Panel of Potential Prognostic Biomarkers Related to Overall Survival of Colorectal Cancer, based on Kaplan-Meier Survival Curves in HPA and UALCAN databases. ROC curve analysis results are also shown. (*Markers with AUC ≥ 0.90 were selected as a prognostic panel.) *SPP1* and *CHEK1* can serve as prognostic markers for colorectal cancer where all *CHEK1, SPP1, MAD2L1*, and *KIF18A* can serve as prognostic markers for rectum adenocarcinoma.

Gene Biomarkers				
Gene symbol	KM-Plot p-value	AUC	ROC curve p-value	Survival data-related tissue
SPP1 *	0.019	0.9898 *	<0.0001	Colon, Rectum
CLCA1	0.0003	0.8622	0.0011	Colon, Rectum
WNT5A	0.00003	0.8214	0.0038	Colon, Rectum
CHEK1 *	0.00036	0.9643 *	<0.0001	Colon, Rectum
LEF1	0.024	0.8929	0.0004	Rectum
KIF18A *	0.033	0.9592 *	<0.0001	Rectum
MAD2L1 *	0.0035	0.949 *	<0.0001	Rectum

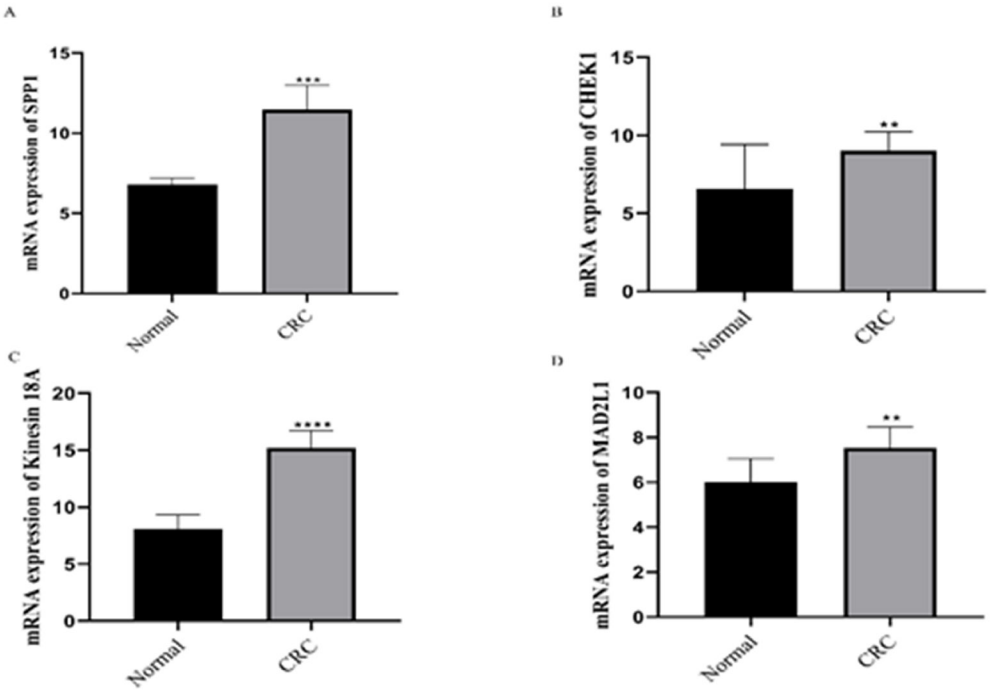

Figure 6. Relative Gene Expression pattern of A) *SPP1,* B) *CHEK1*, C) *Kinesin 18A*, and D) *MAD2L1* between CRC tissues and normal tissues by real-time PCR. The expression of *GAPDH* was used as internal control gene.

biomarkers associated with overall survival in patients with CRC by PPI network analysis. We identified 340 common *DEGs* from GSE18105 and GSE113513 datasets in CRC tissues compared to adjacent normal tissues. The GO enrichment analysis suggested shared *DEGs* were significantly enriched in calcium-independent cell-cell adhesion via plasma membrane cell-adhesion molecules, response to prostaglandin, collagen catabolic process, epithelial to mesenchymal transition, interleukin-1 beta secretion, atrial cardiac muscle tissue morphogenesis, mesenchymal cell development, and negative regulation of cell proliferation involved in contact inhibition. In accordance with the present results, previous studies have reported that the cell adhesion is an important component of malignant transformation, cancer progression, and the development of chemo-resistance [12]. Also, according to other studies, prostaglandins exert a profound influence over the adhesive, migratory, and invasive behavior of cells during the development and progression of cancer [13]. Additionally, it is demonstrated that microsomal prostaglandin E2 synthase-1 (mPGES-1) becomes upregulated in cancer [14].

Our results showed that *SPP1, CHEK1, CLCA1, WNT5A, LEF1, KIF18A,* and *MAD2L1* can serve as prognostic markers for CRC, and rectum adenocarcinoma based on HPA and UALCAN analysis. Interestingly, some nodes including *SPP1*, *CHEK1*, *KIF18A*, and *MAD2L1* were also detected as prognostic markers for CRC based on AUC values more than 90 percent in ROC curve analysis. In addition, all of these genes were presented as a hub with high degrees in the PPI network. Therefore, this genes panel can play an important role in diagnosis, and prognosis of CRC. Therefore, due to passing through several screening analyses including PPI network, survival and ROC analyses, these 4 genes have a higher value in subsequent evaluations.

Secreted phosphoprotein 1 (*SPP1*) is involved in immune regulation, cell survival, and tumor progression. Studies have demonstrated that *SPP1* plays an important role in certain individual tumors. On the other hand, the results of numerous studies indicated that SSP1 is a prognostic biomarker and also plays an oncogenic role in various types of human cancers [15-21]. In a comprehensive study, Yuan et al., discovered that *SPP1* expression was higher in the majority of the human cancers and also the highest expression of *SSP1* was remarkably associated with poor survival in various cancers based on Kaplan-Meier plotter and the PrognoScan database [16]. These data suggested that correlated genes with *SPP1* were mainly involved in positive regulation of immune cell activation and infiltration. Also, *SPP1* has shown a significant positive relationship with the immunocyte and immune marker sets infiltrating degrees. All of these findings provide strong evidence that *SPP1* may promote cancer progression via linking with carcinogenic genes and facilitating immune cells' infiltration in colon adenocarcinoma and other types of cancers [16]. However, the expression profile and oncogenic features of *SPP1* in multiple human cancers are remaining unknown. In this regard, Zeng et al., (2022) analyzed the expression of *SPP1* and its correlation with tumor prognosis, immune invasion, tumor microenvironment, and immunotherapy in 33 cancer types [17]. In conclusion, they reported that *SPP1* contributed to tumorigenesis, tumor progression, and regulated tumor immune microenvironment, revealing *SPP1* may be a new and potential target for assessment of prognosis and immunotherapy in diverse cancers [17]. Recently, another research group has found that

the *SPP1* expression level was positively associated with the infiltration level of dendritic cells, neutrophils, and macrophages in multiple cancer types. It was also significantly positively correlated with hepatitis A virus cellular receptor 2 (HAVCR2), which was observed in most tumor types.

Checkpoint kinase 1 (*CHEK1*) is an important serine/threonine kinase that blocks damaged DNA from being copied and passed on to offspring, therefore it is an area of increasing interest in cancer development and treatment [22, 23]. It has been reported that there is a correlation between *CHEK1* expression and tumor grade and also its recurrence [24, 25]. There is growing evidence that *CHEK1* is involved in tumorigenesis and plays an important role in surviving cancer cells after chemotherapy [26]. Further, studies' results suggest that *CHEK1* does not appear to be a tumor suppressor; but it assists tumor growth and might enhance resistance to anticancer therapy. According to Al-Kaabi et al., the up-regulation of *CHEK1* may be related to poor prognosis in breast cancer [27]. Furthermore, the up-regulation of *CHEK1* was observed in various human cancers such as hepatocellular carcinoma, breast and colon cancer [28-30]. In regard of CRC, Fang et al., have reported that the repression of *CHEK1* caused a significant reduction in cell proliferation and CCNB1, an important member of the cyclin family, expression in CRC cells [31]. In another study, Stawinska et al. evaluated alterations in *CHEK1* and *CHEK2* expression levels in colon cancer and reported that decreased expression of *CHEK2* may be a key mechanism involved in the development of colorectal neoplasm [32]. On the other hand, there are limited therapeutic strategy for advanced CRC. A study has shown that *CHEK1* inhibitor sensitizes resistant CRC stem cells to nortopsentin [33]. Therefore, these findings could provide a basis to develop an efficient option for CRC treatment. Recently, Tozaki et al. have reported that the combined inhibition of ataxia telangiectasia-mutated serine/threonine kinase (ATM) and *CHEK1* appeared synergistic antitumor effects and induced synergistic lethality in CRC cells at a low dose, which can be the basis for creating new treatments for colorectal cancer [34]. In addition, by integrated bioinformatics analysis, Yu et al., have presented key genes and signaling pathways in CRC that *CHEK1* were found to be promising prognostic biomarkers among CRC patients [35]. Recent treatment methods have gone towards the identification of proteins involved in mitotic regulation [36]. Drugs with mitotic inhibitor characteristics (such as taxanes and vinca alkaloids) target microtubules, and some positive results have been obtained for the treatment of multiple human cancers. Recently, kinesin motor proteins have been found to be important proteins regulating mitotic processes and also potential targets of carcinoma treatment [37, 38]. Kinesin family member 18A (*KIF18A*) is a microtubule-associated motor that contributed to cell division [39]. Some investigations have shown that RNA interference suppressed *KIF18A* expression in cells, and cells were stopped at G2/M. this means that *KIF18A* regulates the cell cycle [40]. Additionally, according to the previous studies, the low expression of *KIF18A* has been observed in normal tissues, while it has increased expression in solid tumors, including breast cancer [41], hepatocellular carcinoma [42], colorectal cancer [43], and other types of cancer. For example, Nagahara et al. evaluated the role of *KIF18A* in the CRC progression. For this purpose, they were assessed the mRNA level of *KIF18A* by qRT-PCR in patients with CRC [43]. The authors found that the *KIF18A* was upregulated significantly in CRC compared to the normal colon tissue. In addition, overexpression of *KIF18A* in CRC is significantly associated with clinicopathologic factors including tumor stage and metastasis, which indicate that *KIF18A* has a major role in CRC progression. Further in vitro and translational studies by this group demonstrated that *KIF18A* expression was correlated with metastasis and was presented as a key factor for CRC progression [43]. Additionally, Zhu et al. have conducted a bioinformatics study in order to introduce biomarkers in colon cancer [44]. The results of their analysis indicated that *KIF18A* may serve as a biomarker for the early diagnosis and progression of colon cancer [44]. Chromosomal instability (CIN) is a hallmark of cancer, and targeting of CIN-associated vulnerabilities is a newly therapeutic option in drug discovery. Recently, Tamayo et al. discovered some analogs that could be used as chemical probes to interrogate the role of *KIF18A* inhibition [45]. This was the first disclosure of *KIF18A* inhibitors with in vivo activity that could be hopeful in cancer treatment such as CRC.

In the current analysis, the mitotic arrest deficient 2-like 1 (*MAD2L1*) is identified as prognostic biomarker for CRC. *MAD2L1* as a component of spindle checkpoint has an important role in mitosis [46]. Dysregulation of *MAD2L1* could lead to CIN and aneuploidy, which may facilitate the development of human cancers. Several studies have reported the overexpression of *MAD2L1* in various human gastrointestinal cancers such as liver and gastric cancer [47-49]. Specifically, some studies have mentioned this protein as a prognostic signature in colon cancer [50]. For example, Rimkus' results demonstrate that overexpression of *MAD2L2* associated with poor prognosis in the CRC [51]. In another study by Ding et al., it has been revealed that the expression of *MAD2L1* in CRC tissues is higher than that in normal tissues. In addition, knockdown of *MAD2L1* remarkably inhibited CRC cell growth through impairing cell cycle progression and promoting apoptosis [52]. Therefore, *MAD2L1* can serve as a novel oncogenic gene, which has a role in the regulation of cancer cell growth and apoptosis and could be used as an emerging diagnostic and therapeutic biomarker for CRC. Totally, these results show the prognostic significance and expression characteristics of *SPP1*, *CHEK1*, *KIF18A*, and *MAD2L1* in the CRC that have required further verification in larger populations with CRC. To validation of our results, we finally performed mRNA level alteration analysis of four genes (including *SPP1*, *CHEK1*, *KIF18A*, and *MAD2L1*) by qRT-PCR in twelve CRC patients. The results showed the up-regulation of them in cancerous tissues compared to normal tissues. In this regard, the previous studies also reported the high expression of these genes in several types of tumor tissues compared to normal tissues that were consistent with the

current study [29, 43, 52, 53]. Several notable strengths of this study were the experimental investigation of identified prognostic genes in CRC patients, all the data analyzed in our study was extracted from GEO databases. Then, further experiments with biological tissue samples were performed to validate our findings.

In conclusion, in the present study, we found four key genes involved in prognosis of CRC. Functional analysis showed those key genes were significantly associated with CRC-related signaling pathways such as calcium-independent cell-cell adhesion. On the other hand, the up-regulation of key genes is associated with the poor prognosis in CRC patients. Taken together, this study provides useful insight on the understanding of carcinogenesis and helps in early detection and prognosis of CRC.

Author Contribution Statement

MB. J, M. K and N.AD contributed to the conception, and design and critically reviewed and approved the final manuscript as submitted. N.AD contributed to data collection, statistical analyses, and interpretation and wrote the paper draft. All authors approved the final version for submission.

Acknowledgements

We thank the Cancer Gene Therapy Research Center, Zanjan University of Medical Sciences, Zanjan, Iran that supported this study.

Funding statement

This work was financially supported by a grant (A-12-1636-2) from the Cancer Gene Therapy Research Center, Zanjan University of Medical Sciences.

Ethics issue and Consent to Participate

The second part of study involving humans was approved by the Ethic Committee of Zanjan University of Medical Sciences and was conducted in accordance with the declaration of Helsinki. Informed consent was obtained from all the participants or legal guardians of the illiterate participants for participation in the study.

Availability of data and materials

The data used to support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors have no conflicts of interest to declare.

Abbreviation list

CRC: Colorectal cancer; BP: biological processes, MF: molecular functions, CC: cellular components; HPA: Human Protein Atlas; AUC: Area under the curve; RT-qPCR: Reverse-Transcription Quantitative-Real-Time PCR; cDNA: Complementary DNA; *GAPDH*: Glyceraldehyde 3-Phosphate Dehydrogenase; mPGES-1: microsomal Prostaglandin E2 Synthase-1; *SPP1*: Secreted phosphoprotein 1; HAVCR2: Hepatitis A Virus Cellular Receptor 2; GO: Gene Ontology.

References

1. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut. 2017;66(4):683-91. https://doi.org/10.1136/gutjnl-2015-310912.
2. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. Lancet. 2019;394(10207):1467-80. https://doi.org/10.1016/s0140-6736(19)32319-0.
3. Andersen SE, Andersen IB, Jensen BV, Pfeiffer P, Ota T, Larsen JS. A systematic review of observational studies of trifluridine/tipiracil (tas-102) for metastatic colorectal cancer. Acta Oncol. 2019;58(8):1149-57. https://doi.org/10.1080/0284186x.2019.1605192.
4. Li X-Y, Hu Y, Li N-S, Wan J-H, Zhu Y, Lu N-H. *RACK1* acts as a potential tumor promoter in colorectal cancer. Gastroenterol Res Pract. 2019;2019(1):5625026. https://doi.org/10.1155/2019/5625026.
5. Ouyang S, Zhou X, Chen Z, Wang M, Zheng X, Xie M. Lncrna bcar4, targeting to mir-665/stat3 signaling, maintains cancer stem cells stemness and promotes tumorigenicity in colorectal cancer. Cancer Cell Int. 2019;19:72. https://doi.org/10.1186/s12935-019-0784-3.
6. Voduc D, Kenney C, Nielsen TO. Tissue microarrays in clinical oncology. Semin Radiat Oncol. 2008;18(2):89-97. https://doi.org/10.1016/j.semradonc.2007.10.006.
7. Chen R, Ma S, Qiao H, Su F, Wang L, Guan Q-L. Identification of target genes and prognostic evaluation for colorectal cancer using integrated bioinformatics analysis. All Life. 2022;15:160-73. https://doi.org/10.1080/26895293.2022.2026825.
8. Zhao B, Baloch Z, Ma Y, Wan Z, Huo Y, Li F, et al. Identification of potential key genes and pathways in early-onset colorectal cancer through bioinformatics analysis. Cancer control. 2019;26(1):1073274819831260. https://doi.org/10.1177/1073274819831260.
9. Matsuyama T, Ishikawa T, Mogushi K, Yoshida T, Iida S, Uetake H, et al. Muc12 mrna expression is an independent marker of prognosis in stage ii and stage iii colorectal cancer. Int J Cancer. 2010;127(10):2292-9. https://doi.org/10.1002/ijc.25256.
10. Shen A, Liu L, Huang Y, Shen Z, Wu M, Chen X, et al. Down-regulating haus6 suppresses cell proliferation by activating the p53/p21 pathway in colorectal cancer. Front Cell Dev Biol. 2021;9:772077. https://doi.org/10.3389/fcell.2021.772077.
11. Oliveros JC. Venny. An interactive tool for comparing lists with Venn's diagrams. 2007-2015. Available from: http://bioinfogp.cnb.csic.es/tools/venny/index.html.
12. Hartmann TN. Editorial: Metabolism and cell adhesion in cancer. Front cell dev biol. 2022;10. https://doi.org/10.3389/fcell.2022.871471.
13. Menter DG, Dubois RN. Prostaglandins in cancer cell adhesion, migration, and invasion. Int J Cell Biol. 2012;2012:723419. https://doi.org/10.1155/2012/723419.
14. Finetti F, Travelli C, Ercoli J, Colombo G, Buoso E, Trabalzini L. Prostaglandin e2 and cancer: Insight into tumor progression and immunity. Biology (Basel). 2020;9(12). https://doi.org/10.3390/biology9120434.
15. Gao W, Liu D, Sun H, Shao Z, Shi P, Li T, et al. *SPP1* is a prognostic related biomarker and correlated with tumor-infiltrating immune cells in ovarian cancer. BMC Cancer. 2022;22(1):1367. https://doi.org/10.1186/s12885-022-10485-8.

16. Wei T, Bi G, Bian Y, Ruan S, Yuan G, Xie H, et al. The significance of secreted phosphoprotein 1 in multiple human cancers. Front Mol Biosci. 2020;7:565383. https://doi.org/10.3389/fmolb.2020.565383.
17. Zeng P, Zhang X, Xiang T, Ling Z, Lin C, Diao H. Secreted phosphoprotein 1 as a potential prognostic and immunotherapy biomarker in multiple human cancers. Bioengineered. 2022;13(2):3221-39. https://doi.org/10.1080/21655979.2021.2020391.
18. Liu Y, Ye G, Dong B, Huang L, Zhang C, Sheng Y, et al. A pan-cancer analysis of the oncogenic role of secreted phosphoprotein 1 (*SPP1*) in human cancers. Ann Transl Med. 2022;10(6):279. https://doi.org/10.21037/atm-22-829.
19. Matsubara E, Yano H, Pan C, Komohara Y, Fujiwara Y, Zhao S, et al. The significance of *SPP1* in lung cancers and its impact as a marker for protumor tumor-associated macrophages. Cancers (Basel). 2023;15(8). https://doi.org/10.3390/cancers15082250.
20. Deepti P, Pasha A, Kumbhakar DV, Doneti R, Heena SK, Bhanoth S, et al. Overexpression of secreted phosphoprotein 1 (*SPP1*) predicts poor survival in hpv positive cervical cancer. Gene. 2022;824:146381. https://doi.org/10.1016/j.gene.2022.146381.
21. Ji X, Liu Y, Mei F, Li X, Zhang M, Yao B, et al. *SPP1* overexpression is associated with poor outcomes in alk fusion lung cancer patients without receiving targeted therapy. Sci Rep. 2021;11(1):14031. https://doi.org/10.1038/s41598-021-93484-2.
22. Zhang Y, Hunter T. Roles of chk1 in cell biology and cancer therapy. Int J Cancer. 2014;134(5):1013-23. https://doi.org/10.1002/ijc.28226.
23. Goto H, Izawa I, Li P, Inagaki M. Novel regulation of checkpoint kinase 1: Is checkpoint kinase 1 a good candidate for anti-cancer therapy? Cancer Sci. 2012;103(7):1195-200. https://doi.org/10.1111/j.1349-7006.2012.02280.x.
24. Patil M, Pabla N, Dong Z. Checkpoint kinase 1 in DNA damage response and cell cycle regulation. Cell Mol Life Sci. 2013;70(21):4009-21. https://doi.org/10.1007/s00018-013-1307-3.
25. Fadaka AO, Bakare OO, Sibuyi NRS, Klein A. Gene expression alterations and molecular analysis of *CHEK1* in solid tumors. Cancers (Basel). 2020;12(3). https://doi.org/10.3390/cancers12030662.
26. Tho LM, Libertini S, Rampling R, Sansom O, Gillespie DA. Chk1 is essential for chemical carcinogen-induced mouse skin tumorigenesis. Oncogene. 2012;31(11):1366-75. https://doi.org/10.1038/onc.2011.326.
27. Al-Kaabi MM, Alshareeda AT, Jerjees DA, Muftah AA, Green AR, Alsubhi NH, et al. Checkpoint kinase1 (chk1) is an important biomarker in breast cancer having a role in chemotherapy response. Br J Cancer. 2015;112(5):901-11. https://doi.org/10.1038/bjc.2014.576.
28. Madoz-Gúrpide J, Cañamero M, Sanchez L, Solano J, Alfonso P, Casal JI. A proteomics analysis of cell signaling alterations in colorectal cancer. Mol Cell Proteomics. 2007;6(12):2150-64. https://doi.org/10.1074/mcp.M700006-MCP200.
29. Bai E, Dong M, Lin X, Sun D, Dong L. Expressional and functional characteristics of checkpoint kinase 1 as a prognostic biomarker in hepatocellular carcinoma. Transl Cancer Res. 2022;11(12):4272-88. https://doi.org/10.21037/tcr-22-1701.
30. Wu M, Pang JS, Sun Q, Huang Y, Hou JY, Chen G, et al. The clinical significance of *CHEK1* in breast cancer: A high-throughput data analysis and immunohistochemical study. Int J Clin Exp Pathol. 2019;12(1):1-20.
31. Fang Y, Yu H, Liang X, Xu J, Cai X. Chk1-induced ccnb1 overexpression promotes cell proliferation and tumor growth in human colorectal cancer. Cancer Biol Ther. 2014;15(9):1268-79. https://doi.org/10.4161/cbt.29691.
32. Stawinska M, Cygankiewicz A, Trzcinski R, Mik M, Dziki A, Krajewska WM. Alterations of chk1 and chk2 expression in colon cancer. Int J Colorectal Dis. 2008;23(12):1243-9. https://doi.org/10.1007/s00384-008-0551-8.
33. Di Franco S, Parrino B, Gaggianesi M, Pantina VD, Bianca P, Nicotra A, et al. Chk1 inhibitor sensitizes resistant colorectal cancer stem cells to nortopsentin. iScience. 2021;24(6):102664. https://doi.org/10.1016/j.isci.2021.102664.
34. Tozaki Y, Aoki H, Kato R, Toriuchi K, Arame S, Inoue Y, et al. The combination of atm and chk1 inhibitors induces synthetic lethality in colorectal cancer cells. Cancers (Basel). 2023;15(3). https://doi.org/10.3390/cancers15030735.
35. Yu C, Chen F, Jiang J, Zhang H, Zhou M. Screening key genes and signaling pathways in colorectal cancer by integrated bioinformatics analysis. Mol Med Rep. 2019;20(2):1259-69. https://doi.org/10.3892/mmr.2019.10336.
36. Eswaran J, Soundararajan M, Knapp S. Targeting group ii paks in cancer and metastasis. Cancer Metastasis Rev. 2009;28(1-2):209-17. https://doi.org/10.1007/s10555-008-9181-4.
37. Huszar D, Theoclitou ME, Skolnik J, Herbst R. Kinesin motor proteins as targets for cancer therapy. Cancer Metastasis Rev. 2009;28(1-2):197-208. https://doi.org/10.1007/s10555-009-9185-8.
38. Marquis ca. Investigating kinesin *KIF18A* as a novel therapeutic target for aggressive cancers; 2019. Available from: Https://scholarworks.Uvm.Edu/hcoltheses/286.
39. Miki H, Okada Y, Hirokawa N. Analysis of the kinesin superfamily: Insights into structure and function. Trends Cell Biol. 2005;15(9):467-76. https://doi.org/10.1016/j.tcb.2005.07.006.
40. Mayr MI, Hümmer S, Bormann J, Grüner T, Adio S, Woehlke G, et al. The human kinesin *KIF18A* is a motile microtubule depolymerase essential for chromosome congression. Curr Biol. 2007;17(6):488-98. https://doi.org/10.1016/j.cub.2007.02.036.
41. Alfarsi LH, Elansari R, Toss MS, Diez-Rodriguez M, Nolan CC, Ellis IO, et al. Kinesin family member-18a (*KIF18A*) is a predictive biomarker of poor benefit from endocrine therapy in early er+ breast cancer. Breast Cancer Res Treat. 2019;173(1):93-102. https://doi.org/10.1007/s10549-018-4978-5.
42. Luo W, Liao M, Liao Y, Chen X, Huang C, Fan J, et al. The role of kinesin *KIF18A* in the invasion and metastasis of hepatocellular carcinoma. World J Surg Oncol. 2018;16(1):36. https://doi.org/10.1186/s12957-018-1342-5.
43. Nagahara M, Nishida N, Iwatsuki M, Ishimaru S, Mimori K, Tanaka F, et al. Kinesin 18a expression: Clinical relevance to colorectal cancer progression. Int J Cancer. 2011;129(11):2543-52. https://doi.org/10.1002/ijc.25916.
44. Zhu Y, Sun L, Yu J, Xiang Y, Shen M, Wasan HS, et al. Identification of biomarkers in colon cancer based on bioinformatic analysis. Transl Cancer Res. 2020;9(8):4879-95. https://doi.org/10.21037/tcr-20-845.
45. Tamayo NA, Bourbeau MP, Allen JR, Ashton KS, Chen JJ, Kaller MR, et al. Targeting the mitotic kinesin *KIF18A* in chromosomally unstable cancers: Hit optimization toward an in vivo chemical probe. J Med Chem. 2022;65(6):4972-90. https://doi.org/10.1021/acs.jmedchem.1c02030.
46. Guo Y, Zhang X, Yang M, Miao X, Shi Y, Yao J, et al. Functional evaluation of missense variations in the human mad1l1 and *MAD2L1* genes and their impact on susceptibility to lung cancer. J Med Genet. 2010;47(9):616-

22. https://doi.org/10.1136/jmg.2009.074252.
47. Bian Y, Kitagawa R, Bansal PK, Fujii Y, Stepanov A, Kitagawa K. Synthetic genetic array screen identifies pp2a as a therapeutic target in mad2-overexpressing tumors. Proc Natl Acad Sci U S A. 2014;111(4):1628-33. https://doi.org/10.1073/pnas.1315588111.
48. Michel LS, Liberal V, Chatterjee A, Kirchwegger R, Pasche B, Gerald W, et al. Mad2 haplo-insufficiency causes premature anaphase and chromosome instability in mammalian cells. Nature. 2001;409(6818):355-9. https://doi.org/10.1038/35053094.
49. Chen Q, Yang S, Zhang Y, Li B, Xu H, Zuo S. Identification of *MAD2L1* as a potential biomarker in hepatocellular carcinoma via comprehensive bioinformatics analysis. BioMed Research International. 2022;2022(1):9868022. https://doi.org/https://doi.org/10.1155/2022/9868022.
50. Xu S, Liu D, Cui M, Zhang Y, Zhang Y, Guo S, et al. Identification of hub genes for early diagnosis and predicting prognosis in colon adenocarcinoma. Biomed Res Int. 2022;2022:1893351. https://doi.org/10.1155/2022/1893351.
51. Rimkus C, Friederichs J, Rosenberg R, Holzmann B, Siewert JR, Janssen KP. Expression of the mitotic checkpoint gene mad2l2 has prognostic significance in colon cancer. Int J Cancer. 2007;120(1):207-11. https://doi.org/10.1002/ijc.22155.
52. Ding X, Duan H, Luo H. Identification of core gene expression signature and key pathways in colorectal cancer. Front Genet. 2020;11:45. https://doi.org/10.3389/fgene.2020.00045.
53. Xu C, Sun L, Jiang C, Zhou H, Gu L, Liu Y, et al. *SPP1*, analyzed by bioinformatics methods, promotes the metastasis in colorectal cancer by activating emt pathway. Biomed Pharmacother. 2017;91:1167-77. https://doi.org/10.1016/j.biopha.2017.05.056.

This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.