Journal of King Saud University - Science 36 (2024) 103415

HOSTED BY

Contents lists available at ScienceDirect

Journal of King Saud University - Science

journal homepage: www.sciencedirect.com

Deciphering gene expression signatures in liver metastasized colorectal cancer in stage IV colorectal cancer patients

Mahmood Rasool [a,*], Sajjad Karim [a], Absarul Haque [b], Mohammed Alharthi [c], Adeel G Chaudhary [a], Peter Natesan Pushparaj [a]

[a] Center of Excellence in Genomic Medicine Research, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[b] King Fahd Medical Research Center, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[c] Faculty of Medicine, King Abdulaziz University Hospital, King Abdulaziz University, Jeddah 21589, Saudi Arabia

ARTICLE INFO

Keywords:
Colorectal cancer liver metastasis
Gene expression signature
Gene set enrichment analysis
Differentially expressed genes
Biomarkers

ABSTRACT

Background: Colorectal cancer (CRC) liver metastasis (CRLM) is a clinical challenge, and optimizing treatment strategies is crucial for improving patient outcomes. This study aimed to identify gene expression signatures associated with CRLM for early diagnosis and improved treatment outcomes.
Methods: We obtained RNA-seq data of 34 samples (17 colorectal tumor samples from metastasized liver and 17 samples from normal surrounding colonic epithelia) from the Gene Expression Omnibus (GEO) with accession number GSE50760 and analysed them using next-generation knowledge discovery (NGKD) tools such as GEO2R and web based gene set enrichment analysis (WebGestalt).
Results: A total of 18808 genes were identified in the initial analysis which were further reduced to 2490 differentially expressed genes (DEGs) after applying different parameters using GEO2R tools. Furthermore, in the gene set enrichment analysis (GSEA), we analysed the biological processes, cellular components, and molecular functions. We analysed four pathways: KEGG, panther, reactome, and wikipathway cancer. In each analysis, we ascertained the most important top expressed and downregulated genes.
Conclusions: We identified various gene sets that could be used as important prognostic and therapeutic markers, particularly in patients with advanced CRLM. Future studies in larger cohorts may further our research findings to establish the best prognostic biomarkers for early diagnosis and overall survival.

1. Introduction

Colorectal cancer (CRC) is one of the major health issues the world is facing today (Rasool et al., 2021) and the patient's condition worsens with CRC liver metastasis (CRLM), which is present in almost one-fourth of the primary diagnosed patients and contributes to an increase in the mortality rates of CRC patients (Kim et al., 2015). The overall five-year survival rate of CRLM is less than 10 %, making it one of the deadliest diseases (Abdel-Rahman et al., 2019).

Although the outcomes of CRLM vary with the different strategies for prognosis, surgical removal of isolated CRC liver metastases showed the best curative effects. The latest results have shown many folds in survival rates after hepatic resection of CRC, implying improved prospects (House et al., 2010). Following improved targeted therapies, improvements in CRLM management have contributed significantly to improving the overall median survival rates in patients up to 30 months or more (Petrelli et al., 2024).

The main reason for better management is the selection of effective targeted therapies, such as anti-epidermal growth factor receptor (EGFR) and anti-vascular endothelial growth factor (VEGF) antibodies, which are the result of improved molecular characterization of CRC. These biomarkers not only improve overall prognosis but also enable better choices for personalized medicine approaches from palliative care to potential cures (Petrelli et al., 2024).

While observable progress has been made in recent years in understanding organ-specific cancer medicines, knowledge of metastatic cancer cases remains poor. Therefore, a comprehensive understanding of the molecular mechanisms of organotropic metastasis is important for predicting biomarker-based prognosis, utilizing it for new therapeutic inventions, and ultimately, utilizing it for overall patient survival and outcomes (Chen et al., 2018; Michl et al., 2024).

The term 'organotropic metastasis" is defined as the spread of cancer

* Corresponding author.
E-mail address: mrahmed1@kau.edu.sa (M. Rasool).

https://doi.org/10.1016/j.jksus.2024.103415
Received 24 June 2024; Received in revised form 2 August 2024; Accepted 25 August 2024
Available online 26 August 2024

to distant organs in a non-random, well-regulated manner under the control of different factors, such as molecular signatures of cancer cells, subtype of cancer, microenvironment of host immune systems, and interactions of various cells in the locality (Chen et al., 2018). Although the mechanisms that cause CRC are well studied, the molecular understanding is still very poor and evolving when it comes to CLRM therapeutic implications and management. The alterations responsible for the transformation of normal colonic epithelium to adenocarcinoma may be both acquired and inherited (Araujo et al., 2023). Consequently, it is necessary to understand the molecular alterations that lead to the development of primary CRC in metastatic disease, particularly in multidisciplinary approaches in CRLM and beyond (Araujo et al., 2023).

The molecular underpinning of CRLM is complex and involves many different factors and biological events. Hence, to develop better therapeutic strategies, it is essential to understand these molecular processes, knowledge of patients, and their selection to achieve overall survival and health outcomes (Araujo et al., 2023). The process of CRC metastasis in the liver requires special skills of a subset of CRC cells that modify themselves and acquire the ability to escape from the primary site (i.e., colon or rectal) and mount on to the liver. The steps involved in this transfer from CRC to CRLM are epithelial mesenchymal transition, relocation through the extracellular matrix, tissue invasion, and overcoming circulation and settlement in the liver parenchyma (Yu et al., 2020).

Therefore, the current study aimed to define the gene signatures in liver metastasized colorectal cancer using the tumor samples taken from the liver, analysis of the expression pattern of genes, and analysis using various next-generation knowledge discovery (NGKD) tools to dissect the underlying molecular pathways and associated genes for improved understanding and development of diagnosis and prognosis in CRC patients with liver metastasis.

2. Materials and Methods

2.1. Data acquisition

In this study, we obtained RNA-seq data of 34 samples (17 colorectal tumor samples from metastasized liver and 17 samples from normal surrounding colonic epithelia) from the GSE50760 dataset (Kim et al., 2014). Initially, the dataset contained 18 patients' samples (metastatic liver CRC) and 18 normal samples, but we removed one sample in our analysis due to a peculiar expression pattern.

2.2. GEO2R analysis

GEO2R (for comparison of two or more groups of samples in GEO series by using variety of R packages from Bioconductor project) analysis was performed using the Benjamini and Hochberg (False discovery rate) method (Benjamini and Hochberg, 1995) along with a significance level cut-off at 0.05, and log 2 fold change (log2FC) threshold at 1 to identify differentially expressed genes and generate a uniform manifold approximation and projection (UMAP) plot (UMAP is the latest technique for visualization of relatedness of samples using dimension reduction technique), volcano plot (used to reveal the statistical significance (−log10 P value) as compared to magnitude of fold change (log2 fold change) and is capable of easily understanding and visualizing differentially expressed genes), and mean difference (MD) plot (MD plot is also used to visualize expression of genes and it shows log2 fold change vs. average log2 expression values).

2.3. Web-based gene SeT AnaLysis Toolkit (WebGestalt)

The WebGestalt online tool was used for functional enrichment analysis, interpretation, and extraction of biological insights from the genes of interest (Liao et al., 2019). The basic parameters were set as follows: a) Organism of interest: Homo sapiens, b) method of interest: gene set enrichment analysis (GSEA). c- functional database: gene ontology (biological process non-redundant, cellular component redundant, and molecular function non-redundant) and pathway (kegg, panther, reactome, and wikipathway cancer). The advanced parameters were set to their respective default values. For GSEA, both the gene symbol and the long fold change (logFC) value from the gene list were uploaded with the logFC cutoff set at (1.5 to −1.5).

3. Results

The present study aimed to identify a gene expression signature predictive of CRLM development. For this research, we obtained a dataset from the NCBI Gene Expression Omnibus (GEO) database with accession number GSE50760. The primary aim of this study was to identify the gene signatures associated with liver-metastasized CRC in patients with stage IV CRC.

3.1. GEO2R

In the GEO2R analysis, we identified that the expression pattern of the genes in CRLM patients was entirely different from that in the normal tissue samples, as indicated in the UMAP plot (Fig. 1). In GEO2R, we obtained a set of 18,808 genes in total, while after applying the logFC cut-off of 1.5 and −1.5, the number of differentially expressed genes was reduced to 2490. Therefore, we employed 2490 gene sets in all analyses. The 20 most significantly 20 genes (upregulated and downregulated regulated) are listed in Table 1.

3.2. WebGestalt

In WebGestalt, we selected gene set enrichment analysis (GSEA) for all analysis parameters as a standard and modified the other options as described. Under geneontology, we further analyzed two main pathways: 1-geneontology and 2-pathway.

3.2.1. GSEA geneontology

In this study, we analysed biological processes, cellular components, and molecular functions (all noRedendant).

In geneneontology biological process (GO BP), we found "acute inflammatory response" as the most enriched underlying process (Fig. 2. accession number GO:0002526) with an Enrichment Score of 0.70394 and a Normalized Enrichment Score of 3.5471.

In geneontology cellular component (GO CC), we found "blood microparticle" as the most important (Fig. 2. accession number GO:0072562), enrichment score of 0.79645, and normalized enrichment score of 3.8727.

In geneontology molecular function (GO MF), we found "enzyme inhibitor activity" as most upregulated (Fig. 2. accession number GO:0004857), enrichment score of 0.62087, and normalized enrichment score of 3.1806. Whereas we found "passive transmembrane transporter activity" as the most downregulated (Fig. 2. accession number GO:0022803), with an enrichment score of −0.35864 and a normalized enrichment score of −2.1878. The top 10 genes involved in all these processes are listed in Table 2. The top 10 enriched genes in KEGG, Panther, and Reactome pathways based on GSEA method in WebGestalt are provided in Table 3.

3.2.2. Differentaially regulated pathways

We analysed the DEGs in WebGestalt using four different pathway databssuch as KEGG, Panther, Reactome, and Wikipathway Cancer. In KEGG, we found "complement and coagulation cascades" as the most expressed pathway (Fig. 3. accession number: hsa04610), with an enrichment score of 0.73993 and a normalized enrichment score of 3.5550.

Based on Panther, we identified "blood coagulation" as the most differntialy regulat pathway (Fig. 3. accession number P00011) with an

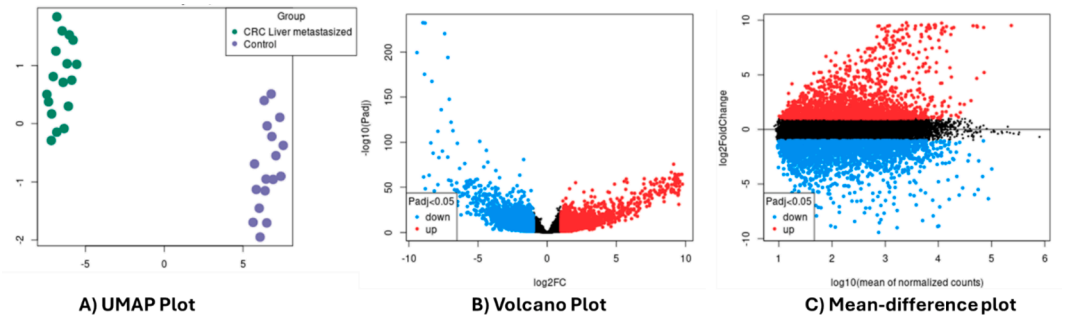

Fig. 1. Shows UMAP, Volcano and Mean-difference plots. In A, Green dots show the CRLM samples while purple denotes control samples whereas in B and C the red dots show the upregulated genes while blue dots show downregulated genes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Shows the top 20 most upregulated and downregulated genes sets.

S. No.	Genes (upregulated)	LogFC	Genes (downregulated)	LogFC
1	C8B	9.774	OTOP2	−9.421
2	HPR	9.739	TMIGD1	−8.993
3	APOA2	9.577	VSTM2A	−8.939
4	ITIH1	9.573	CLCA4	−8.877
5	F2	9.564	CA1	−8.835
6	AHSG	9.564	KRT24	−8.817
7	C8A	9.532	GCG	−8.540
8	ALB	9.516	NKX2-3	−8.420
9	CPB2	9.514	MS4A12	−8.340
10	FGG	9.493	CLDN8	−8.254
11	APCS	9.493	OTOP3	−8.167
12	FGB	9.484	CD177	−7.911
13	AFM	9.476	LINC01082	−7.910
14	APOA1-AS	9.469	PYY	−7.833
15	F9	9.468	GUCA2B	−7.670
16	PLG	9.468	HSD3B2	−7.624
17	SERPINA4	9.456	HAPLN1	−7.583
18	APOC3	9.437	BEST4	−7.418
19	AMBP	9.422	VIP	−7.195
20	FGA	9.392	CDKN2B-AS1	−7.188

enrichment score of 0.81729 and a normalized enrichment score of 3.0456. In Reactome, we found that platelet degranulation" and "response to elevated platelet cytosolic Ca2 + were the most highly expressed pathways (Fig. 3. accession numbers: R-HSA-114608 and R-HSA-76005, respectively), with an enrichment score of 0.76992 and a normalized enrichment score of 3.4019. Whereas in Reactome pathway we found "Stimuli-sensing channels" pathway as most downregulated pathway (Fig. 3. accession number: R-HSA-2672351), with an enrichment score of −0.53501 and a normalized enrichment score of −2.1980.

Based on Wikipathway Cancer, we found "One Carbon Metabolism" pathway is most expressed (Fig. 3. accession number: WP241), with an enrichment score of 0.54191 and a normalized enrichment score of 1.6907. In this wikpathway cancer we also found "Metabolic reprogramming in colon cancer" (accession number: WP4290, enrichment score 0.58271, normalized enrichment scores 1.4962), "Epithelial to mesenchymal transition in colorectal cancer" (accession number: WP4239, enrichment score 0.34684, normalized enrichment scores 1.4739) and "LncRNA involvement in canonical Wnt signalling and colorectal cancer" (accession number: WP4258, enrichment score 0.38307, normalized enrichment score 1.4619; Table 4).

4. Discussion

CRC is one of the most prevalent types of cancers worldwide. Currently, many therapeutic options are available to treat CRC, such as surgical removal of tumors, radiation therapy, interventional radiology, and drug control, including targeted therapy, immunotherapy, and chemotherapy (Morawska et al., 2023). The development of these curative measures has decreased the mortality rate of CRC (McQuade et al., 2017). In contrast, the incidence of CRC is on the rise in the general population, particularly in populations under the age of 40 years (Siegel et al., 2017).

The treatment and overall survival of patients mostly depend on the local tumor extent and the presence of distant metastasis. Therefore, knowing the CRC stage is of utmost importance for determining the treatment options. General staging assessment techniques include imaging of organs, such as abdominal pelvic and chest computed tomography (CT) scans and magnetic resonance imaging (MRI) examinations (McQuade et al., 2017).

Roman numerals were used to indicate the stages of cancer, ranging from 0 to IV. O indicates the lowest stage level and is normally limited to the colon lining. The stage IV level means that the cancer has spread to other parts of the body (metastasized) and is called the advanced stage of cancer (McQuade et al., 2017).

The most affected organ in stage IV CRC is the liver, followed by the lungs, lymph nodes, peritoneum, and ovaries (Manfredi et al., 2006; van der Geest et al., 2015). Overall survival and prognosis are based on the stage, extent of cancer spread, and CRLM (van der Geest et al., 2015). To date, the most established treatment option is to remove the tumor via surgical resection of the colon, liver, and other organs (Araujo et al. 2023). However, this option is limited to the presence of a tumor in a particular location and has not metastasized, which is the case in most stage IV patients. Therefore, it is very difficult to treat patients with stage IV CRLM, for whom surgery is not an appropriate option. In such cases, patients need to be treated based on personalized treatment strategies.

Hence, the research focus is shifting towards the development of bioinformatics models and integration of artificial intelligence (AI) tools in medical practice, which can predict cancer more precisely, especially in stage IV CRLM cases. Chakedis et al. (2017) developed a machine learning-based model using bootstrap resampling and multifactorial logistic regression analysis, showing incredible precision in predicting recurrence risk (Chakedis et al., 2017). Consequently, the idea of precision medicine based on the molecular profile of each patient to identify genotype and phenotype variations is becoming more popular. These factors play a major role in the prognosis and prediction of treatment modalities in clinical practice (Araujo et al., 2023).

In our study, where we studied all patients with stage IV CRLM, the focus was to identify the genetic signatures that are displayed in this cohort as a group that can be used in the future for prediction, diagnosis, and employment as measures for treatment success.

In GO BP, we found "acute inflammatory response (GO:0002526)" as the most enriched. This response is categorized as a brief but fast and fairly uniform response to active injury and results in the accumulation of granulocytic leukocytes and plasma proteins, which either resolve within a short span of time or stay and become a chronic inflammatory response. In this acute inflammatory response, the inter-α-trypsin

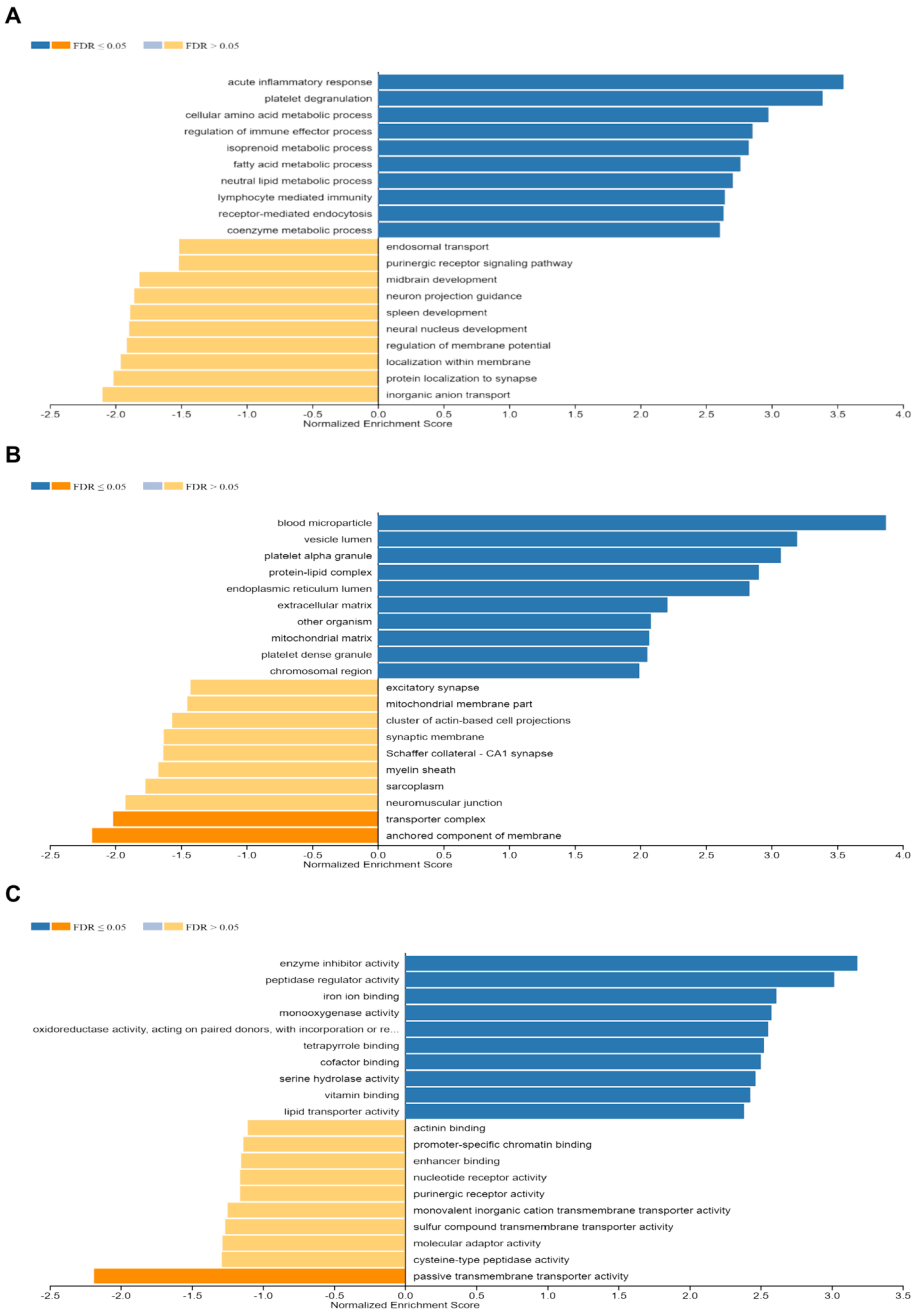

Fig. 2. Geneontology A: Biological Process B: Cellular Component C: Molecular Function.

Table 2
The top 10 enriched genes in geneontology in biological process, cellular component and molecular function.

S. No	Gene Symbol	Score (up)	Gene Symbol	Score	Gene Symbol	Score	Gene Symbol	Score (Down)
	Biological process		Cellular Component		Molecular Function		Molecular Function	
	acute inflammatory response		blood microparticle		enzyme inhibitor activity		passive transmembrane transport activity	
1	ITIH4	7.485	HPR	9.739	APOA2	9.577	OTOP2	−9.421
2	SAA1	7.328	APOA2	9.577	ITIH1	9.573	CLCA4	−8.877
3	C6	7.209	ITIH1	9.573	AHSG +	9.564	OTOP3	−8.167
4	SERPINA3	7.156	AHSG +	9.564	SERPINA4	9.456	BEST4	−7.418
5	SERPINF2	7.07	F2	9.564	APOC3	9.437	GRIK3	−6.708
6	CFHR1	7.014	C8A	9.532	AMBP	9.422	BEST2	−6.574
7	C4BPA	6.501	ALB	9.516	APOA1	9.337	CLCA1	−6.392
8	C5	6.02	APCS	9.493	SERPINC1	9.327	AQP8	−5.804
9	SERPINA1	5.217	FGG	9.493	ITIH2	9.278	GRIA4	−5.78
10	CFI	4.811	FGB	9.484	SERPIND1	8.949	TRPM6	−5.307

Table 3
Top 10 enriched genes in differentially regulated pathways based on KEGG, Panther, and Reactome.

S. No	Gene Symbol	Score	Gene Symbol	Score	Gene Symbol	Score (Up)	Gene Symbol	Score (Down)
	KEGG		Panther		Reactome		Reactome	
	Complement and coagulation cascades		Blood coagulation		Platelet degranulation		Stimuli-sensing channels	
1	C8B	9.774	F2 +	9.564	AHSG	9.564	CLCA4	−8.877
2	F2 +	9.564	FGG +	9.493	ALB	9.516	BEST4	−7.418
3	C8A	9.532	FGB +	9.484	FGG +	9.493	BEST2	−6.574
4	CPB2	9.514	F9	9.468	FGB +	9.484	CLCA1	−6.392
5	FGG +	9.493	PLG	9.468	PLG	9.468	TRPM6	−5.307
6	FGB +	9.484	FGA	9.392	SERPINA4	9.456	SCNN1B	−4.829
7	F9	9.468	SERPINC1 +	9.327	FGA	9.392	TRPA1	−3.992
8	PLG	9.468	SERPINA10	7.728	APOA1	9.337	TRPV3	−3.393
9	FGA	9.392	F7	7.488	ORM1	9.264	CASQ2	−3.28
10	SERPINC1 +	9.327	SERPINF2	7.07	ORM2	9.228		

inhibitor heavy chain 4 (ITIH4) gene was the most expressed gene (LogFC 7.845). In previous studies in dogs, this gene was found to be associated with surgical trauma and was highly expressed up to four times. Therefore, this gene may be a good biomarker for CRLM, as most patients underwent surgical removal of the tumor in grade IV patients (Soler et al., 2016). Another study in human patients using mass spectrometric wheat germ agglutinin multiple reaction monitoring found that biomarker combinations of $PF4^{54-62}$, $ITIH4^{429-438}$, and $APOE^{198-207}$ achieved a specificity of 97.5 % and sensitivity of 84.5 % in CRC diagnosis, which may serve as a highly effective diagnostic biomarker for CRC patients (Tsai et al., 2021). ITIH4 has been previously described to have higher expression in early gastric cancers using mass spectrometry and expression analysis using sera and exomes of patients with early gastric cancer (Sun et al., 2021).

In GO CC, we found "blood microparticle (GO:0072562)" as the most enriched. The blood microparticle is a phosphoslipid microvesible originating from cell types such as blood cells, platelets, and endothelial cells, and has membrane receptors with other protein characteristics of parental cells. They are heterogeneous in size (0.05 to 1.5 μm) and free of nucleic acids. They are released under certain conditions, such as shear stress, proapoptotic stimulation, or complement attack, and their release is a highly controlled process. Therefore, their release is important for the diagnosis of vascular pathologies. These microparticles are involved in various cell-to-cell interactions and signalling pathways. Their physiological roles have been demonstrated in various pathological conditions such as cancer metastasis, inflammation, and response to pathogens (Simak and Gelderman, 2006). In this biological process, we found that (HPR) gene was the most expressed (logFC 9.739). The HPR gene encodes a protein that binds to hemoglobin in a pattern similar to that of haptoglobin. HPR epitopes have been identified as predictors of breast cancer (Kuhajda et al., 1989).

In GO MF, we found "enzyme inhibitor activity (GO:0004857)" as the most upgraded and the APOA2 gene was most expressed gene (logFC 9.577). Enzyme inhibitor activity, also known as metalloenzyme inhibitor activity, is generally described as binding to, stopping, preventing, or even reducing the activity of an enzyme. APOA2 is involved in CRC. A study conducted in ninety-one patients shown that APOA2 is strongly correlated with CRC with exceptional screening efficacy, with an area under the curve (AUC) of 0.957, specificity of 93.33 %, and sensitivity of 85.71 % (Xu et al., 2024). Therefore, APOA2 may serve as a biomarker of greater interest. Whereas we found "passive transmembrane transporter activity (GO:0022803)" as the most downregulated and the OTOP2 gene was found to be most downregulated gene (logFC −9.421). This passive transmembrane transporter activity is involved in the movement of a single solute from one side of the membrane to the other side of the membrane by conformational change, using the facilitation diffusion technique, or if the solute is charged by using a membrane potential-dependent process. The OTOP2 gene has been shown to be significantly correlated with CRC. The expression of OTOP2 is significantly lower in CRC tissues than in normal individuals. Furthermore, in vitro functional assays have shown that OTOP2 gene silencing reduces caspase-3/-9 activity and promotes cell proliferation, migration, and mesenchymal transition (Guo et al., 2022).

In geneontologies, blood microparticles, and enzyme inhibitor activities, the most common genes were APOA2, ITIH1, and AHSG.

Based on KEGG, we found "complement and coagulation cascades (hsa04610)" as most differentially regulated pathway. The complement system is a proteolytic cascade in blood plasma and is a defense mechanism against pathogens and a mediator of innate immunity, whereas the coagulation cascade is the conversion of proenzymes to serine proteases, leading to the generation of thrombin, which converts soluble fibrinogen into insoluble fibrin clots. In this pathway, we found that C8B was the most expressed gene (logFC 9.774). This gene is a part of the complement system, which helps in the body's immune response. Recently, C8B has been shown to be an important predictor of survival in patients with HBV-related hepatocellular carcinoma patients (Zhang et al., 2021).

Based on Panther, we found that blood coagulation" was the most enriched pathway (P00011). In this process, multiple coagulation factors of blood interact with each other, resulting in the creation of insoluble fibrin clots to stop bleeding and blood loss. In this pathway, we found that the F2 gene was the most highly expressed (logFC 9.564), which is an essential factor in blood clotting (hemostasis). F2 dictates the formation of coagulation factor II, also called prothrombin (the inactive form of thrombin). In response to injury, fibrinogen is converted into fibrin, which constitutes the blood clot. Thrombin is believed to participate in tissue repair, cell proliferation, and angiogenesis). Thrombin is also thought to be involved in cell growth and division (proliferation), tissue repair, and angiogenesis (Lancellotti et al., 2009). Further research is needed to elucidate the role of the F2 gene in CRLM patients.

Based on Reactome, we found "Platelet degranulation (R-HAS-114608) and "Response to elevated platelet cytosolic Ca2+ (R-HAS-760050" as the highly enriched. Platelets contain specific effector molecules and release different storage granules at the time of vascular injury, mainly alpha granules, dense granules, and lysosomes that modulate cell signalling. The most highly expressed gene in these two pathways is α2-Heremans-Schmid glycoprotein (AHSG), also known as fetuin-A (logFC9.564), which is known to inhibit insulin action. A recent study in patients ASHG has been indicated that ASHG showed a modest linear association (Nimptsch et al., 2015). Therefore, more research is needed to determine the role of ASHG in patients with CRLM.

We also found "Stimuli-sensing channels (R-HAS-2672351)" pathway as most downregulated pathway. These ion channels coordinate sensations such as pain, taste, vision, warmth, and cold. These channels include acid-sensing ion channels (ASICs; Wang & Xu 2011) and transient receptor potential channels (TRPCs; Takahashi et al. 2012). The most downregulated gene in this pathway was chloride

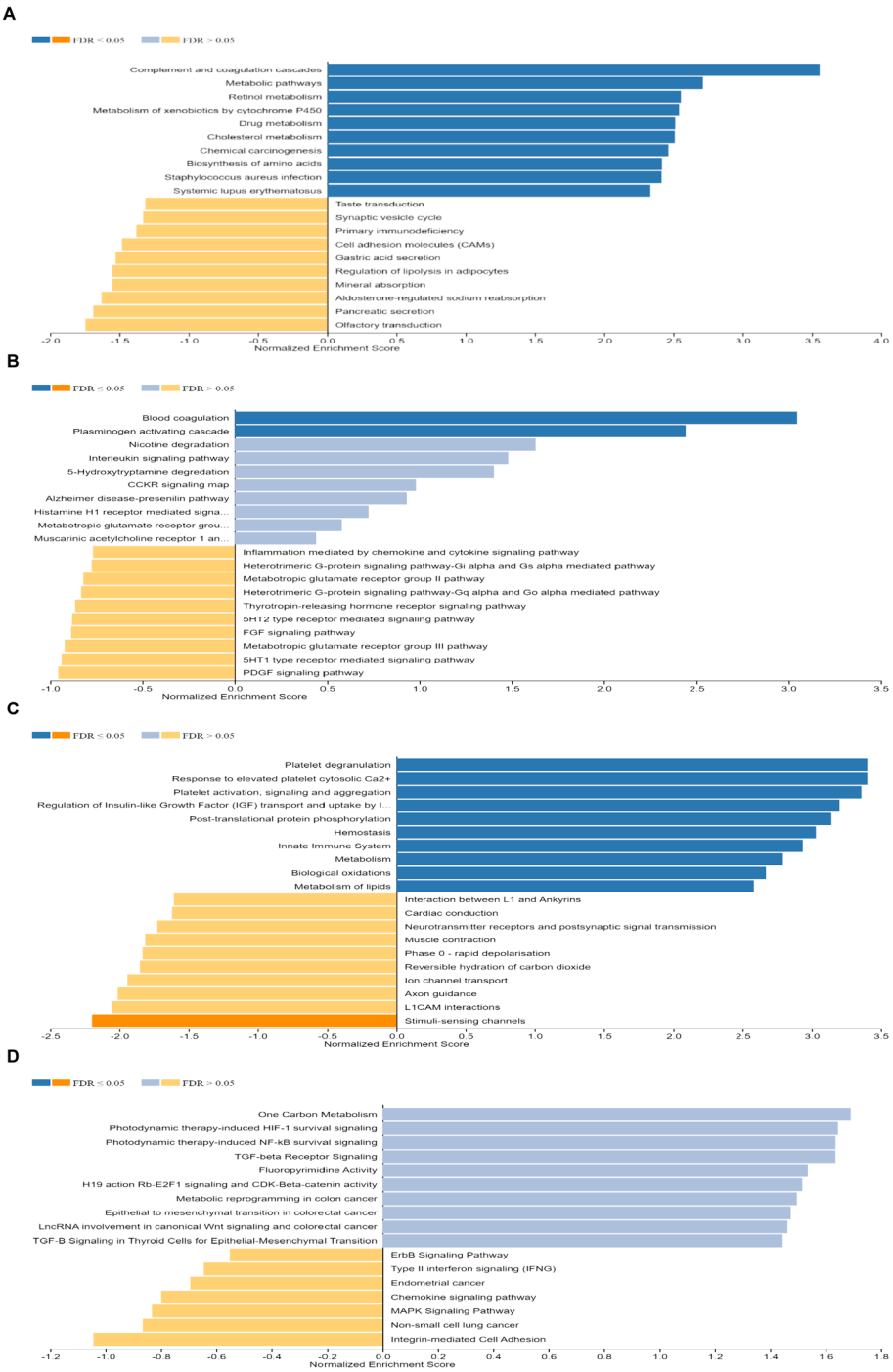

Fig. 3. Differentially regulated pathways based on GSEA method using WebGestalt: A: KEGG, B: Panther, C: Reactome, D: Wikipathway Cancer.

channel accessory 4 (CLC4A; logFC −8.877). A recent study it showed that downregulation of CLCA4 is associated with the development and progression of CRC. Its expression was shown to be lower in primary CRC cases, whereas it was even lower in liver metastases. Therefore, CLCA4 may serve as a good indicator of CRC metastasis in different organs, particularly the liver (Wei et al., 2020).

Based on Wikipathway Cancer we found "One Carbon Metabolism (WP241)" pathway as the most enriched. In this pathway, folate plays the main role, serving as the donor of single carbons in one of the three oxidation states: reduced (5-methyl-THF), intermediate (5,10 methylene-THF), and oxidized (10-formyl-THF). In this pathway, we found that the methionine adenosyltransferase 1A (MAT1A) gene was the most highly expressed (logFC 9.007). MAT1A is predominantly expressed in the normal liver and is required for the synthesis of S-adenosylmethionine, a principal biological methyl donor. Different experiments have shown that it acts as a chemopreventive agent in hepatocellular carcinoma and other forms of cancer. In experimental models, it was effective as a chemopreventive agent in HCC and perhaps other forms of cancer as well (Lu et al., 2012), but regarding CRLM, more work is needed regarding CRLM.

We also observed "metabolic reprogramming in colon cancer (WP4290). Metabolic reprogramming in cancer is known to be driven by multiple factors, such as tumor suppressor genes, oncogenes, growth factor alterations, and tumor-host cell interactions, which promote cancer cell anabolism and further tumor growth. In this pathway, we found that the (ALDOB) gene was the most enriched (logFC 5.651).

Table 4
Description of important pathways obtained in wikipathway cancer.

S. No	Gene Symbol	Score	Gene Symbol	Score	Gene Symbol	Score	Gene Symbol	Score
	One Carbon Metabolism		Metabolic reprogramming in colon cancer		Epithelial to mesenchymal transition in colorectal cancer		LncRNA involvement in canonical Wnt signalling and colorectal cancer	
1	MAT1A	9.007	ALDOB	5.651	VTN	9.301	NOTUM	6.35
2	BHMT	8.413	PSAT1	3.247	CLDN2	5.776	MIR675	6.27
3	FTCD	8.135	SLC2A1	2.787	CLDN1	5.716	H19	6.041
4	FOLH1	3.237	HK3	1.9	FOXQ1	5.585	WNT2	3.679
5	ALDH1L1	3.075	PPAT	1.602	CLDN14	5.238	FOSL1	3.582
6			SHMT2	1.589	WNT2	3.679	NKD1	3.282
7			FASN	1.571	CDH2	3.629	FZD10	3.052
8					FZD10	3.052	NKD2	2.958
9					PROX1	2.739	WNT3	2.632
10					SNAI1	2.656		

Recent studies have shown that higher expression of ALDOB promotes CRC metastasis by accelerating epithelial-mesenchymal transition (EMT) and may be used as a therapeutic agent and a potential prognostic factor (Li et al., 2017).

In wikipathway cancer, we also observed an epithelial to mesenchymal transition in colorectal cancer (WP4239)". In this EMT transition, cells lose epithelial characteristics and adopt mesenchymal properties, such as increased motility, which is a crucial part of CRC metastasis. In this pathway, we found that the vitronectin (VTN) gene was the most expressed (logFC 9.301). In a large-scale systematic proteomic quantification study, VTN was shown to be highly expressed in CRC samples. It appears that VTN expression increases with tumor stage and metastasis, and this may act as a prognostic marker in CRC (Yin et al., 2015).

Furthermore, in wikipathway cancer, we found "LncRNA involvement in canonical Wnt signalling and colorectal cancer (WP4258)." LncRNAs have been found to play important roles in CRC pathogenesis. In this pathway, we found the palmitoleoyl-protein carboxylesterase (NOTUM) gene (logFC 6.35). In a recent study, NOTUM has been shown to be involved in CRC progression. NOTUM expression was even more elevated in metastatic cells. Inhibition of NOTUM expression suppresses cancer proliferation, which proves the important role of NOTUM in metastasis and can be used as a therapeutic and diagnostic marker (Yoon et al., 2018).

In summary, we observed different pathways with gene sets that may support management based on patient- and tumor-specific factors. Our comprehensive approach, incorporating the genetic profiles of patients in stage IV, may help alleviate disease, reduce disease burden, and improve patient care. Future studies with more AI and advanced bioinformatics tools may support more refined prognostic models, address study limitations, and reach the ultimate goal of personalized medicines for CRLM patients, particularly in stage IV patients.

5. Conclusions

In conclusion, we identified various pathways and gene sets responsible for CRLM in patients with stage IV disease. The genes such as C8B, OTOP2, ITIH4, HPR, APOA2, OTOP2, F2 +, AHSG, CLCA4, MAT1A, ALDOB, VTN, and NOTUM seem to play important roles in CRLM progression. The most important pathways identified in this study were complement and coagulation cascades, blood coagulation, platelet degranulation, stimuli-sensing channels, and one carbon metabolism. These new insights will shed more light on CRC-related signalling networks and will enhance our understanding of complex tumor biology. These molecular pathways need to be validated in larger cohorts, aided by functional experiments. It appears that a better understanding of the molecular profiles of patients are needed to adopt more personalized approaches for disease management in view of the complexity of this disease by affecting multi-organ systems.

Funding

This research work was funded by Institutional Fund Projects under grant no (IFPNC-013-141-2020).

CRediT authorship contribution statement

Mahmood Rasool: Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Sajjad Karim:** Software, Methodology, Investigation, Formal analysis, Data curation. **Absarul Haque:** Writing – review & editing, Investigation. **Mohammed Alharthi:** Visualization, Investigation, Data curation. **Adeel G Chaudhary:** Visualization, Supervision, Software, Data curation. **Peter Natesan Pushparaj:** Writing – review & editing, Resources, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

All the data is available in public domain (https://www.ncbi.nlm.nih.gov/geo/). Authors will provide additional information upon reasonable request.

Acknowledgements

This research work was funded by Institutional Fund Projects under grant no (IFPNC-013-141-2020). Therefore, authors gratefully acknowledge technical and financial support from the Ministry of Education and King Abdulaziz University, Jeddah, Saudi Arabia.

References

Abdel-Rahman, O., 2019. Prognostic Value of Baseline ALBI Score Among Patients With Colorectal Liver Metastases: A Pooled Analysis of Two Randomized Trials. Clin. Colorectal Cancer 18 (1), e61–e68. https://doi.org/10.1016/j.clcc.2018.09.008.

Araujo, R.L.C., Fonseca, L.G., Silva, R.O., Linhares, M.M., Uson Junior, P.L.S., 2024. Molecular profiling and patient selection for the multimodal approaches for patients with resectable colorectal liver metastases. Hepatobiliary Surg. Nutr. 13 (2), 273–292. https://doi.org/10.21037/hbsn-22-616.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300.

Chakedis, J., Squires, M.H., Beal, E.W., Hughes, T., Lewis, H., Paredes, A., Al-Mansour, M., Sun, S., Cloyd, J.M., Pawlik, T.M., 2017. Update on current problems in colorectal liver metastasis. Curr. Probl. Surg. 54 (11), 554–602. https://doi.org/10.1067/j.cpsurg.2017.10.002.

Chen, W., Hoffmann, A.D., Liu, H., Liu, X., 2018. Organotropism: new insights into molecular mechanisms of breast cancer metastasis. NPJ Precis. Oncol. 2 (1), 4. https://doi.org/10.1038/s41698-018-0047-0.

Guo, S., Sun, Y., 2022. OTOP2, Inversely Modulated by miR-3148, Inhibits CRC Cell Migration, Proliferation and Epithelial-Mesenchymal Transition: Evidence from Bioinformatics Data Mining and Experimental Verification. Cancer Manag. Res. 14, 1371–1384. https://doi.org/10.2147/CMAR.S345299.

House, M.G., Ito, H., Gönen, M., Fong, Y., Allen, P.J., DeMatteo, R.P., Brennan, M.F., Blumgart, L.H., Jarnagin, W.R., D'Angelica, M.I., 2010. Survival after hepatic resection for metastatic colorectal cancer: trends in outcomes for 1,600 patients during two decades at a single institution. J. Am. College Surgeons. 210 (5), 744–755. https://doi.org/10.1016/j.jamcollsurg.2009.12.040.

Kim, T.M., Jung, S.H., An, C.H., Lee, S.H., Baek, I.P., Kim, M.S., Park, S.W., Rhee, J.K., Lee, S.H., Chung, Y.J., 2015. Subclonal Genomic Architectures of Primary and Metastatic Colorectal Cancer Based on Intratumoral Genetic Heterogeneity. Clinical Cancer Res.: Off. J. Am. Assoc. Cancer Res. 21 (19), 4461–4472. https://doi.org/10.1158/1078-0432.CCR-14-2413.

Kim, S.K., Kim, S.Y., Kim, J.H., Roh, S.A., Cho, D.H., Kim, Y.S., Kim, J.C., 2014. A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. Mol. Oncol. 8 (8), 1653–1666. https://doi.org/10.1016/j.molonc.2014.06.016.

Kuhajda, F.P., Piantadosi, S., Pasternack, G.R., 1989. Haptoglobin-related protein (Hpr) epitopes in breast cancer as a predictor of recurrence of the disease. N. Engl. J. Med. 321 (10), 636–641. https://doi.org/10.1056/NEJM198909073211003.

Lancellotti, S., De Cristofaro, R., 2009. Congenital prothrombin deficiency. Semin. Thromb. Hemost. 35 (4), 367–381. https://doi.org/10.1055/s-0029-1225759.

Li, Q., Li, Y., Xu, J., Wang, S., Xu, Y., Li, X., Cai, S., 2017. Aldolase B Overexpression is Associated with Poor Prognosis and Promotes Tumor Progression by Epithelial-Mesenchymal Transition in Colorectal Adenocarcinoma. Cell. Physiol. Biochem. 42 (1), 397–406. https://doi.org/10.1159/000477484.

Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., Zhang, B., 2019. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 47 (W1), W199–W205. https://doi.org/10.1093/nar/gkz401.

Lu, S.C., Mato, J.M., 2012. S-adenosylmethionine in liver health, injury, and cancer. Physiol. Rev. 92 (4), 1515–1542. https://doi.org/10.1152/physrev.00047.2011.

Manfredi, S., Lepage, C., Hatem, C., Coatmeur, O., Faivre, J., Bouvier, A.M., 2006. Epidemiology and management of liver metastases from colorectal cancer. Ann. Surg. 244 (2), 254–259. https://doi.org/10.1097/01.sla.0000217629.94941.cf.

McQuade, R.M., Stojanovska, V., Bornstein, J.C., Nurgali, K., 2017. Colorectal Cancer Chemotherapy: The Evolution of Treatment and New Approaches. Curr. Med. Chem. 24 (15), 1537–1557. https://doi.org/10.2174/0929867324666170111152436.

Michl, M., Taverna, F., Woischke, C., Li, P., Klauschen, F., Kirchner, T., Heinemann, V., von Bergwelt-Baildon, M., Stahler, A., Herold, T. M., et al., 2024. Identification of a gene expression signature associated with brain metastasis in colorectal cancer. *Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico*. Advance online publication. https://doi.org/10.1007/s12094-024-03408-5.

Morawska, I., Cieszanowski, A., 2023. Assessment of the response to systemic treatment of colorectal liver metastases on cross-sectional imaging - a systematic review. Pol. J. Radiol. 88, e512.

Nimptsch, K., Aleksandrova, K., Boeing, H., Janke, J., Lee, Y.A., Jenab, M., Kong, S.Y., Tsilidis, K.K., Weiderpass, E., Bueno-De-Mesquita, H.B., et al., 2015. Plasma fetuin-A concentration, genetic variation in the AHSG gene and risk of colorectal cancer. Int. J. Cancer 137 (4), 911–920. https://doi.org/10.1002/ijc.29448.

Petrelli, F., Arru, M., Colombo, S., Cavallone, M., Cribiu', F.M., Villardita, V., Floris, P., Digiesi, L., Severgnini, G., Moraes, M.T., et al., 2024. BRAF mutations and survival with surgery for colorectal liver metastases: A systematic review and meta-analysis. Eur. J. Surg. Oncol.: J. Eur. Soc. Surg. Oncol. Br. Assoc. Surg. Oncol. 50 (6), 108306 https://doi.org/10.1016/j.ejso.2024.108306.

Rasool, M., Carracedo, A., Sibiany, A., Al-Sayes, F., Karim, S., Haque, A., Natesan Pushparaj, P., Asif, M., Achakzai, N.M., 2021. Discovery of a novel and a rare Kristen rat sarcoma viral oncogene homolog (KRAS) gene mutation in colorectal cancer patients. Bioengineered 12 (1), 5099–5109. https://doi.org/10.1080/21655979.2021.1960715.

Siegel, R. L., Fedewa, S. A., Anderson, W. F., Miller, K. D., Ma, J., Rosenberg, P. S., Jemal, A. 2017. Colorectal Cancer Incidence Patterns in the United States, 1974-2013. *Journal of the National Cancer Institute*. *109*(8), djw322. https://doi.org/10.1093/jnci/djw322.

Simak, J., Gelderman, M.P., 2006. Cell membrane microparticles in blood and blood products: potentially pathogenic agents and diagnostic markers. Transfus. Med. Rev. 20 (1), 1–26. https://doi.org/10.1016/j.tmrv.2005.08.001.

Soler, L., García, N., Unzueta, A., Piñeiro, M., Álava, M.A., Lampreave, F., 2016. Purification and determination of C-reactive protein and inter-α-trypsin inhibitor heavy chain 4 in dogs after major surgery through generation of specific antibodies. Vet. Immunol. Immunopathol. 179, 26–31. https://doi.org/10.1016/j.vetimm.2016.07.010.

Sun, Y., Jin, J., Jing, H., Lu, Y., Zhu, Q., Shu, C., Zhang, Q., Jing, D., 2021. ITIH4 is a novel serum biomarker for early gastric cancer diagnosis. Clin. Chim. Acta; Int. J. Clin. Chem. 523, 365–373. https://doi.org/10.1016/j.cca.2021.10.022.

Takahashi, N., Kozai, D., Mori, Y., 2012. TRP channels: sensors and transducers of gasotransmitter signals. Front. Physiol. 3, 324. https://doi.org/10.3389/fphys.2012.00324.

Tsai, I.J., Su, E.C., Tsai, I.L., Lin, C.Y., 2021. Clinical Assay for the Early Detection of Colorectal Cancer Using Mass Spectrometric Wheat Germ Agglutinin Multiple Reaction Monitoring. Cancers 13 (9), 2190. https://doi.org/10.3390/cancers13092190.

van der Geest, L.G., Lam-Boer, J., Koopman, M., Verhoef, C., Elferink, M.A., de Wilt, J.H., 2015. Nationwide trends in incidence, treatment and survival of colorectal cancer patients with synchronous metastases. Clin. Exp. Metastasis 32 (5), 457–465. https://doi.org/10.1007/s10585-015-9719-0.

Wang, Y.Z., Xu, T.L., 2011. Acidosis, acid-sensing ion channels, and neuronal cell death. Mol. Neurobiol. 44 (3), 350–358. https://doi.org/10.1007/s12035-011-8204-2.

Wei, L., Chen, W., Zhao, J., Fang, Y., Lin, J., 2020. Downregulation of CLCA4 expression is associated with the development and progression of colorectal cancer. Oncol. Lett. 20 (1), 631–638. https://doi.org/10.3892/ol.2020.11640.

Xu, R., Shen, J., Song, Y., Lu, J., Liu, Y., Cao, Y., Wang, Z., Zhang, J., 2024. Exploration of the application potential of serum multi-biomarker model in colorectal cancer screening. Sci. Rep. 14 (1), 10127. https://doi.org/10.1038/s41598-024-60867-0.

Yin, X., Zhang, Y., Guo, S., Jin, H., Wang, W., Yang, P., 2015. Large scale systematic proteomic quantification from non-metastatic to metastatic colorectal cancer. Sci. Rep. 5, 12120. https://doi.org/10.1038/srep12120.

Yoon, J.H., Kim, D., Kim, J., Lee, H., Ghim, J., Kang, B.J., Song, P., Suh, P.G., Ryu, S.H., Lee, T.G., 2018. NOTUM Is Involved in the Progression of Colorectal Cancer. Cancer Genom. Proteom. 15 (6), 485–497. https://doi.org/10.21873/cgp.20107.

Yu, X., Zhu, L., Liu, J., Xie, M., Chen, J., Li, J., 2020. Emerging Role of Immunotherapy for Colorectal Cancer with Liver Metastasis. OncoTargets Ther. 13, 11645–11658. https://doi.org/10.2147/OTT.S271955.

Zhang, Y., Chen, X., Cao, Y., Yang, Z., 2021. C8B in Complement and Coagulation Cascades Signaling Pathway is a predictor for Survival in HBV-Related Hepatocellular Carcinoma Patients. Cancer Manag. Res. 13, 3503–3515. https://doi.org/10.2147/CMAR.S302917.