REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 19 June 2024 Revised: 03 July 2024 Accepted: 17 July 2024

Bioinformatics-Driven Identification of Hub Genes as Potential Therapeutic Targets in Colorectal Cancer

Manivannan Rangasamy[1], SureshKumar Gopal*, Delipkumar Kaliyappan [2], Dhamotharaprasath R[3], Moulika G[4], Kalaiyarasan G[5], Jayaprasanth J[6]

[1]Professor, Department of Pharmaceutical Biotechnology, Excel College of Pharmacy, Affiliated with Tamil Nadu, Dr. M.G.R. Medical University, Tamil Nadu
[2] Student, Excel College of Pharmacy Affiliated with Tamil Nadu, Dr. M.G.R. Medical University, Tamil Nadu
[3]Student, Excel College of Pharmacy Affiliated with Tamil Nadu, Dr. M.G.R. Medical University, Tamil Nadu
[4]Student, Excel College of Pharmacy Affiliated with Tamil Nadu, Dr. M.G.R. Medical University, Tamil Nadu
[5]Student, Excel College of Pharmacy Affiliated with Tamil Nadu, Dr. M.G.R. Medical University, Tamil Nadu
[6]Student, Excel College of Pharmacy Affiliated with Tamil Nadu, Dr. M.G.R. Medical University, Tamil Nadu

***Corresponding Author:** Mr.G.Suresh Kumar
*Professor, Department of Pharmaceutical Biotechnology, Excel College of Pharmacy, Affiliated with Tamil Nadu, Dr. M.G.R. Medical University, Tamil Nadu. Email: sureshbiotech1983@gmail.com

Abstract
This study used bioinformatics to identify potential therapeutic targets for colorectal cancer (CRC). Gene expression data from GEO2R were analyzed to identify differentially expressed genes in CRC. FunRich software was used to create the Venn diagrams. The potential "hub genes" were selected from the STRING database. Functional analyses, including Gene Ontology and pathway enrichment, were performed using Database for Annotation, Visualization, and Integrated Discovery (DAVID). Patient survival data from GEPIA, gene expression related to disease stage and metastatic progression, and genetic changes were examined using the cBioPortal and Human Protein Atlas. Among differentially expressed genes, CXCL8, FOXC1, ICOS, and MCF2 were identified as potential hub genes from 89 upregulated genes which plays a significant role in CRC development. The identified genes may serve as valuable biomarkers for diagnosing CRC and predicting patient outcomes, potentially informing the development of targeted treatments to improve patient survival rates.

Keywords: Bioinformatics, Colorectal cancer (CRC), Differentially expressed genes, Hub gene, Therapeutic targets

Introduction

Colorectal cancer (CRC) is a significant global health concern, ranking as the third most prevalent malignancy and second leading cause of cancer-related mortality, with an estimated 1.9 million new cases and 0.9 million deaths reported in 2020. Studies have shown a relatively high incidence of signet ring carcinoma (13%) and advanced-stage disease in the liver and lung metastases(1,2). Several studies have shown that dietary habits, such as high consumption of red meat and eggs, increase the incidence of colorectal cancer (CRC)(3). Other lifestyle factors, including low physical activity, smoking, alcohol consumption, and metabolic disorders such as high body mass index, type 2 diabetes, and hypertension, have also been linked to a greater risk of CRC(4,5). The primary pathophysiology of colorectal cancer (CRC) involves chromosomal instability, microsatellite instability, and epigenetic modifications, including DNA methylation, histone modifications, and non-coding RNA alterations(6). Mutations in APC, KRAS, TP53, and MYC are primarily associated with tumor development in both sporadic and inflammatory bowel disease-associated CRC (7).

Microarray technology coupled with bioinformatics enables analysis of large datasets generated from genomic, transcriptomic, and proteomic analyses (8). Advanced platforms, such as Affymetrix GeneChip and Illumina BeadArray, are widely used in microarray studies(9). Complementing microarray analysis, the integration of deep learning and neural graph networks has further enhanced the prediction of protein function, aiding drug discovery and improving cancer prognosis and diagnosis (10). The NCBI Gene Expression Omnibus (GEO) is a public repository of gene expression and epigenomic data containing over 200,000 studies and 6.5 million samples. It provides web-based tools for the analysis and visualization of differential expressions, including new interactive graphical plots available in GEO2R(11).

The Cancer Genome Atlas (TCGA) is a comprehensive public repository of genomic data covering 33 cancer types, including rare cancers (12). It contains information on DNA sequences, transcriptional data, and epigenetic modifications, enabling the identification of genes and pathways, and accurate cancer classification(13). TCGA dataset has supported the development of several databases, such as UALCAN (14), cBioPortal(15), STRING(16), and the Human Protein Atlas (17) for data collection and analysis. mRNA expression, networking, and DNA methylation were visualized using software tools such as Cytoscape(18) and Fun Rich(19) software tools to identify significant prognostic markers in various cancers. This comprehensive analysis aids in identifying tumor-related genes, and the utilization of microarray technology facilitates the comprehensive examination of these crucial genes with the objective of identifying potential molecular targets and diagnostic markers.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 19 June 2024 Revised: 03 July 2024 Accepted: 17 July 2024

Our study aimed to use bioinformatics tools to identify the key genes involved in the development of colorectal cancer (CRC) and assess their potential as treatment targets. We began by examining microarray data and identified significantly upregulated genes associated with cancer-related pathways, particularly those associated with CRC. To obtain a clearer picture of CRC progression, we analysed these genes through protein—protein interaction (PPI) networks, Gene Ontology (GO) functional classifications, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway mapping.
Furthermore, we conducted survival and expression analyses of these critical genes and examined the variation in their mRNA levels across cancer stages, nodal metastasis, and patient survival rates. By correlating survival data with gene expression patterns, we identified specific hub genes that exhibited significant potential as therapeutic targets or biomarkers for CRC diagnosis.

Materials and methods

Data collection

The Gene Expression Omnibus (GEO) accession number GSE164191 dataset includes blood specimens from a cohort of 59 individuals with colorectal cancer and 62 healthy individuals, all of which were analyzed using the Affymetrix Human Genome U133 Plus 2.0 Array platform. This gene expression dataset, submitted by Sun et al.(20) and available in the GEO database, allows researchers to analyze gene expression in the blood to identify immune-associated genes specifically linked to colorectal cancer. These findings could aid in developing diagnostic tools to differentiate colorectal cancer patients from healthy individuals based on blood gene expression profiles.

Data processing of DEGs

Differentially expressed genes (DEGs) between colorectal cancer patients and control subjects were identified using GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r), applying a statistical significance threshold of $p < 0.05$. Genes with a log-fold change (log2FC) > 1 were classified as upregulated, whereas those with log2FC < −1 were considered downregulated(21). A volcano plot was used to visualize the DEGs, demonstrating the relationship between the fold change and the p-value for gene expression levels.

Gene Ontology (GO) Enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses.

Gene Ontology (GO) analysis is a commonly used method for categorizing genes and their RNA or protein products into specific GO categories, enabling the detection of distinct biological characteristics from high-throughput transcriptomic or genomic data. The Kyoto Encyclopedia of Genes and Genomes (KEGG) provides a collection of databases that contain detailed insights into genomes, diseases, molecular pathways, drugs, and chemical compounds. The Database for Annotation, Visualization, and Integrated Discovery (DAVID; http://david.ncifcrf.gov, version 6.8) is a web-based database that integrates biological knowledge with data analysis tools, and is frequently used to analyze GO and KEGG pathways. A p-value <0.05 was considered the threshold for statistically significant enrichment results(22).

Protein-protein interaction network and Cytohubba analysis

STRING version 12.0 (https://string-db.org/), also known as Search Tool for the Retrieval of Interacting Genes/Proteins, is an online database developed to identify both physical and functional associations between genes and proteins. In this study, we used STRING to construct the protein–protein interaction (PPI) network of differentially expressed genes (DEGs), setting a combined interaction score greater than 0.4 as the threshold(23). Further analysis of the PPI network was conducted using Cytoscape, in which key hub genes were identified using the CytoHubba plugin, which applies 11 distinct algorithms. From this analysis, the top ten dysregulated genes were identified and designated as hub genes.

mRNA expression and survival analysis of the hub genes

Survival rates and the relationships between key genes in colorectal cancer (CRC) subjects were explored using various databases, including UALCAN (http://ualcan.path.uab.edu/), GEPIA (http://gepia.cancerpku.cn/), and KM plotter (https://kmplot.com/analysis/). Survival analysis was conducted using the Kaplan-Meier method in combination with the log-rank test, with statistical significance set at $p < 0.05$, to highlight significant correlations between gene expression levels and patient survival outcomes. To confirm gene expression levels, CRC patient data from The Cancer Genome Atlas (TCGA) were analyzed. Based on the transcript per million (TPM) values, patients were categorized into two groups for visualization in GEPIA: the low/medium expression group (TPM below the upper quartile) and the high expression group (TPM above the upper quartile).

Prognostic and metastatic potential of hub genes

To evaluate the prognostic role of hub genes, mRNA expression data from the initial and late stages of colorectal cancer (CRC) were examined using the GEPIA database (http://gepia.cancer-pku.cn/). The metastatic potential of these hub genes was assessed using the UALCAN database (http://ualcan.path.uab.edu/). The cBio Cancer Genomics Portal (http://cbioportal.org) presents the mutation and expression profiles of hub genes from colorectal tissue specimens in TCGA dataset. This interface provides an in-depth mutation analysis and identification of gene mutations, duplications, and deletions across 20 CRC studies. The protein levels of the hub genes in cancerous tissues were further analyzed using the Human Protein Atlas (HPA) database

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 19 June 2024 Revised: 03 July 2024 Accepted: 17 July 2024

(https://www.proteinatlas.org/), which provides immunohistochemistry-based expression data, enabling the assessment of protein expression profiles across multiple human tissues.

Results

Identification of DEGs

The GSE164191 dataset, containing gene expression data of blood specimen collected from 59 colorectal cancer and 62 normal individuals, was used to identify significantly dysregulated genes. Differential expression was determined using a p-value < 0.05 and |logFC| > 1. A total of 145 genes were identified, including 89 up-regulated and 56 down-regulated genes. The volcano plot (Figure 1) illustrates these results, with red dots indicating upregulated genes and green dots indicating downregulated genes. A detailed list of the identified genes is provided in Table 1

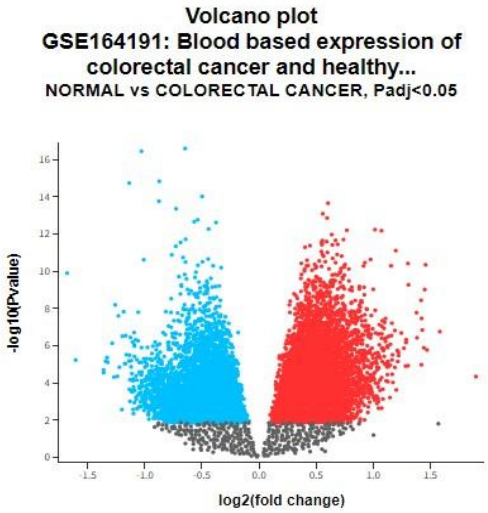


Figure 1. Volcano Plot of Differential Gene Expression in Colorectal Cancer vs. Normal Tissues (GSE164191 Dataset)

Table 1: List of Upregulated and Downregulated Differentially Expressed Genes (DEGs) in CRC

DEGs	Gene
Upregulated (89)	GOLGA8A, MYBL1, TNFRSF25, SUPT16H, RBM8A, STK26, PTPN4, ZNF677, ELK4, ZBTB10, ENPP4, PIGL, GRAMD1C, SENP7, ZBTB1, CXCL8, CBLB, TMEM267, MBLAC2, ZNF540, KLRC4, TSEN15, SCML1, ENPP5, KCTD7, RBAK, TTN, MAT2B, ZNF91, ENPP4, MDFIC, TMEM263, LMAN1, ELP2, ERCC6L2, AGBL3, RABEP1, MIR146A, CCDC65, ICOS, TAF9B, LMAN1, TBC1D32, CENPK, ZNF397, E2F5, FAM188A, RDX, B4GALT6, WDR60, SH2D1A, ZEB1, ZC2HC1A, MKLN1, CCDC82, MIER3, ABCD2, ZNF770, TRIM23, S1PR5, PGBD1, HLA-DRB4, VPS13B, ARMCX4, TMEM117, ZNF883, GEN1, ARAF, UTP23, PFDN4, BRMS1L, B3GALT2, CD22, ZNF404, CXorf57, ZNF91, THAP9, FSD1L, C9orf3, N4BP2L1, RGS1, HLA-DRB4.
Downregulated (56)	PPP3R1, C1QC, TP53I11, SERPINA3, MMP9, GRPR, C1QB, TBL1X, IRS1, GPER1, PNPLA1, MYO7B, USF1, MAGI1, NFIC, ERFE, FOXC1, LYPD8, UNC5D, PGLYRP1, OLAH, MYL9, PCYT1B, LMO2, ZNF704, NLRP3, CD177, SCGB2A2, RNF130, GRB10, CPNE4, PNLIPRP2, PRKAA2, ERC2, NDST4, SIGLEC11, EFCAB6, NR2F2, DCT, ARHGEF12, ACSM2A, ZNF536, MCF2, DSC3, ONECUT2, TMEM150C, PCDHB6, CACNA1E, ALB, NOVA1, NLGN1, CAGE1, CDH7, RAP1GAP, SLC17A1, SIM1.

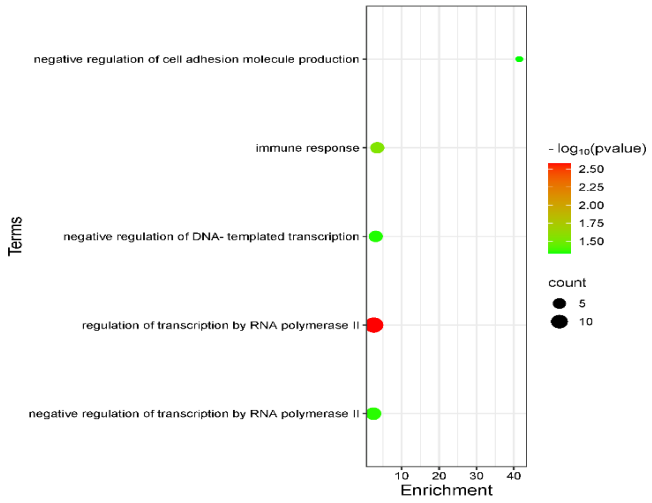
GO and KEGG pathway analysis

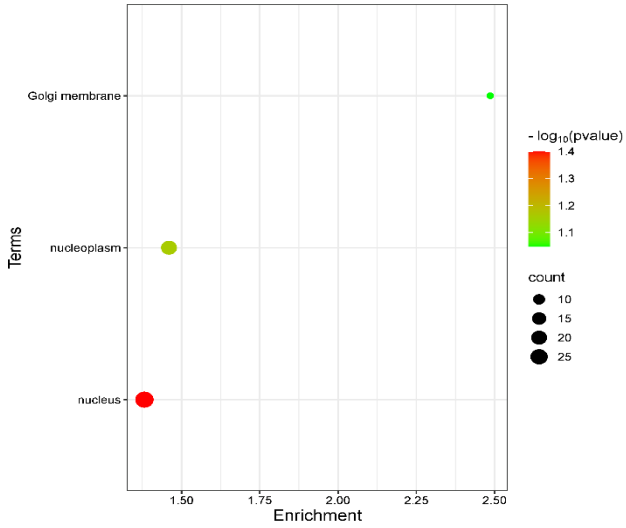
A total of 89 upregulated and 56 downregulated genes were examined using the DAVID software. The top five significant pathways were identified based on the different categories.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 19 June 2024 Revised: 03 July 2024 Accepted: 17 July 2024

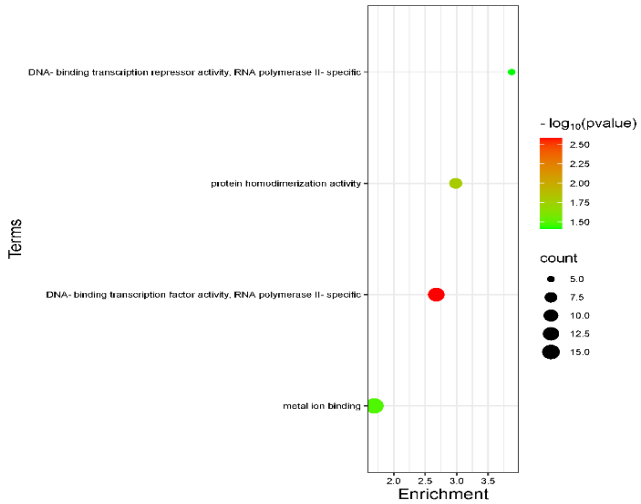
In the Biological Process category, the upregulated DEGs were associated with negative regulation of transcription by RNA polymerase II, regulation of transcription by RNA polymerase II, negative regulation of DNA-templated transcription, immune response, and negative regulation of cell adhesion molecule production (Figure 2(a)), whereas the downregulated genes were attributed to calcium-dependent cell–cell adhesion via plasma membrane cell adhesion molecules, glucose homeostasis, positive regulation of transcription by RNA polymerase II, fatty acid biosynthesis, and the innate immune response.
In the Cellular Component category, the upregulated DEGs were related to the nucleus, nucleoplasm, and Golgi membrane (Figure 2(b)), whereas the downregulated DEGs were associated with the postsynapse, complement component C1q complex, complement component C1 complex, membrane, and blood microparticles.
In the Molecular Function (MF) category, upregulated DEGs were enriched in RNA polymerase II cis-regulatory region sequence-specific DNA binding, DNA-binding transcription factor activity, RNA polymerase II-specific protein homodimerization activity, and metal ion binding (Figure 2(c)), whereas downregulated DEGs were associated with calcium ion binding, identical protein binding, sequence-specific DNA binding, peptidoglycan binding, and triglyceride lipase activity.
In terms of KEGG pathway enrichment, the downregulated genes are associated with the Herpes simplex virus 1 infection pathway, while the upregulated genes are linked to the axon guidance, pertussis, C-type lectin receptor signaling pathway, alcoholic liver disease, and oxytocin signaling pathways.

(a) Biological Processes





(b) Cellular Components



(c) Molecular Function

Figure 2. Bubble map illustrating GO and KEGG pathway analyses for upregulated DEGs. The top 5 terms from GO and KEGG pathway enrichment analyses. Statistical significance was set at a p-value < 0.05.

PPI network construction and module analysis

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 19 June 2024 Revised: 03 July 2024 Accepted: 17 July 2024

A PPI network of DEGs in CRC was constructed using the STRING database. Upon uploading 145 DEGs, a total of 168 genes were mapped into the PPI network, comprising 131 nodes and 52 edges, with a PPI enrichment p-value of 0.21 (Figure 3). Furthermore, an interaction network of 145 DEGs and their neighbouring genes was generated using FunRich. Significant modules within the network were identified using the MCODE plugin (Figure 4), with the two most prominent functional clusters selected (Module 1, MCODE score = 4.50; Module 2, MCODE score = 3.00) (Figure 4(a) & 4(b)),

To further refine our analysis and identify the key genes responsible for CRC pathology, we compiled the cytoHubba plugin, ten genes exhibiting the highest degree scores (MMP9, IRS1, FOXC1, CXCL8, TTN, RGS1, ICOS, ALB, MCF2, and ZEB1) were identified as hub genes for colorectal cancer (Figure 5(a)). Subsequently, the STRING database was used to construct the PPI network for these hub genes (Figure 5(b)), and FunRich software was employed to generate the interaction network of the hub genes along with their related genes (Figure 5(c)). In Figure 4(a), the PPI network of the hub genes consists of 10 nodes and 11 edges, with an average local clustering coefficient of 0.78 and a PPI enrichment p-value of less than 1.0e−16.

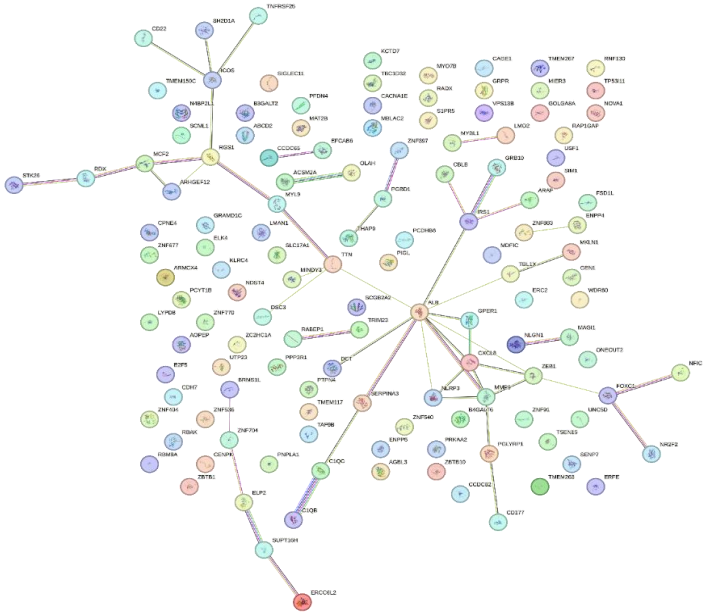
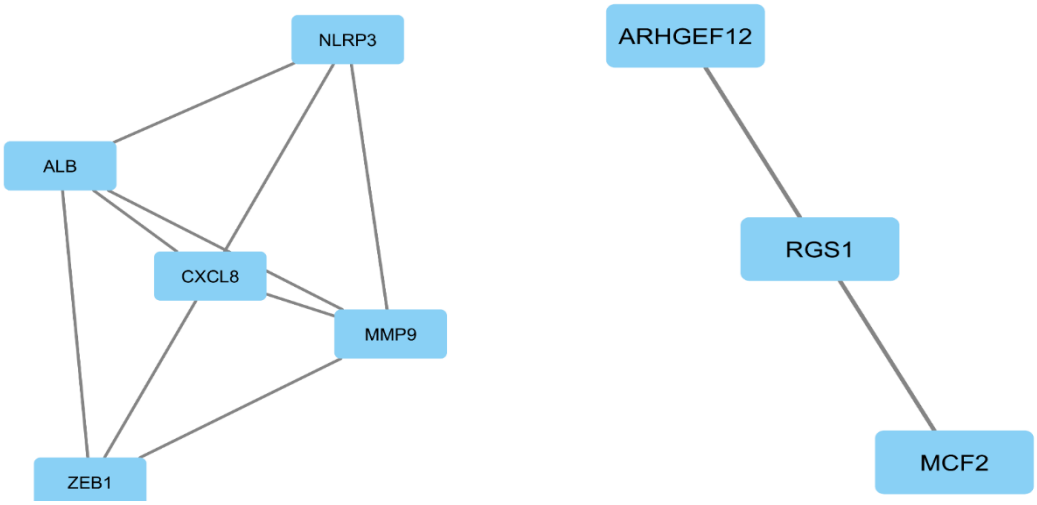


Figure 3. Protein-Protein Interaction (PPI) Network of Differentially Expressed Genes (DEGs). This figure shows the overall PPI network constructed from the 145 identified DEGs, comprising 131 nodes and 52 edges, as generated by the STRING database



(a) Module 1 with a score of 4.50. (b) Module 2 with a score of 3.00

Figure 4. Significant Modules Identified from the PPI Network via MCODE.This figure displays the two most significant functional clusters (modules) identified from the PPI network using the MCODE plugin

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 19 June 2024 Revised: 03 July 2024 Accepted: 17 July 2024

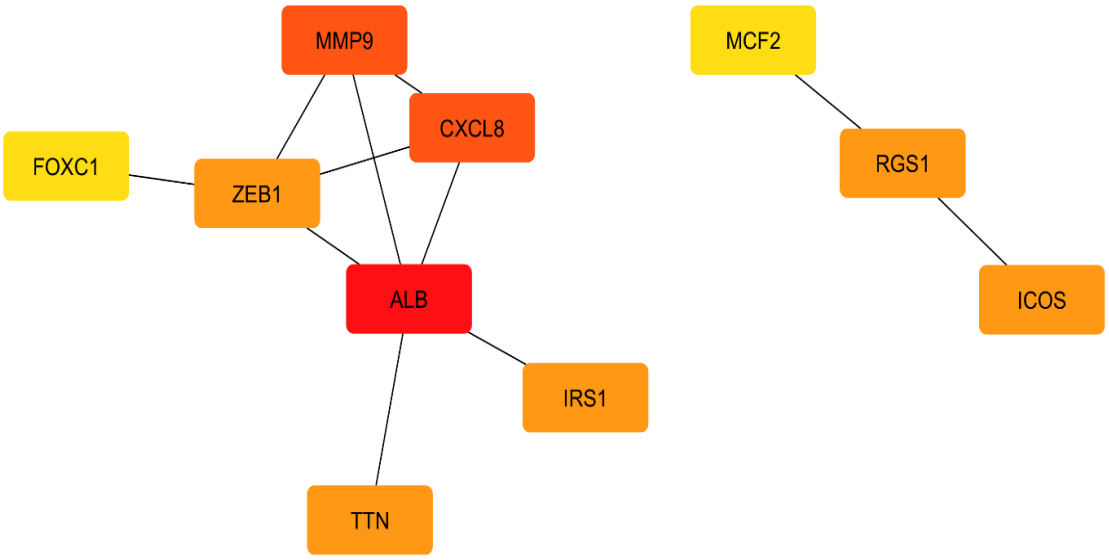


Figure 5. Identification and Interaction Networks of Hub Genes. (a) The top 10 hub genes identified using the CytoHubba

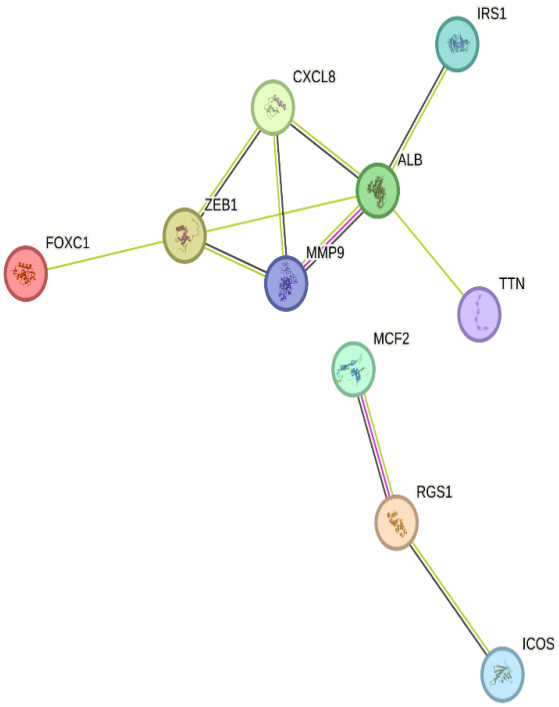


Figure 5(b). The PPI network of these 10 hub genes, generated using the STRING database.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 19 June 2024 Revised: 03 July 2024 Accepted: 17 July 2024

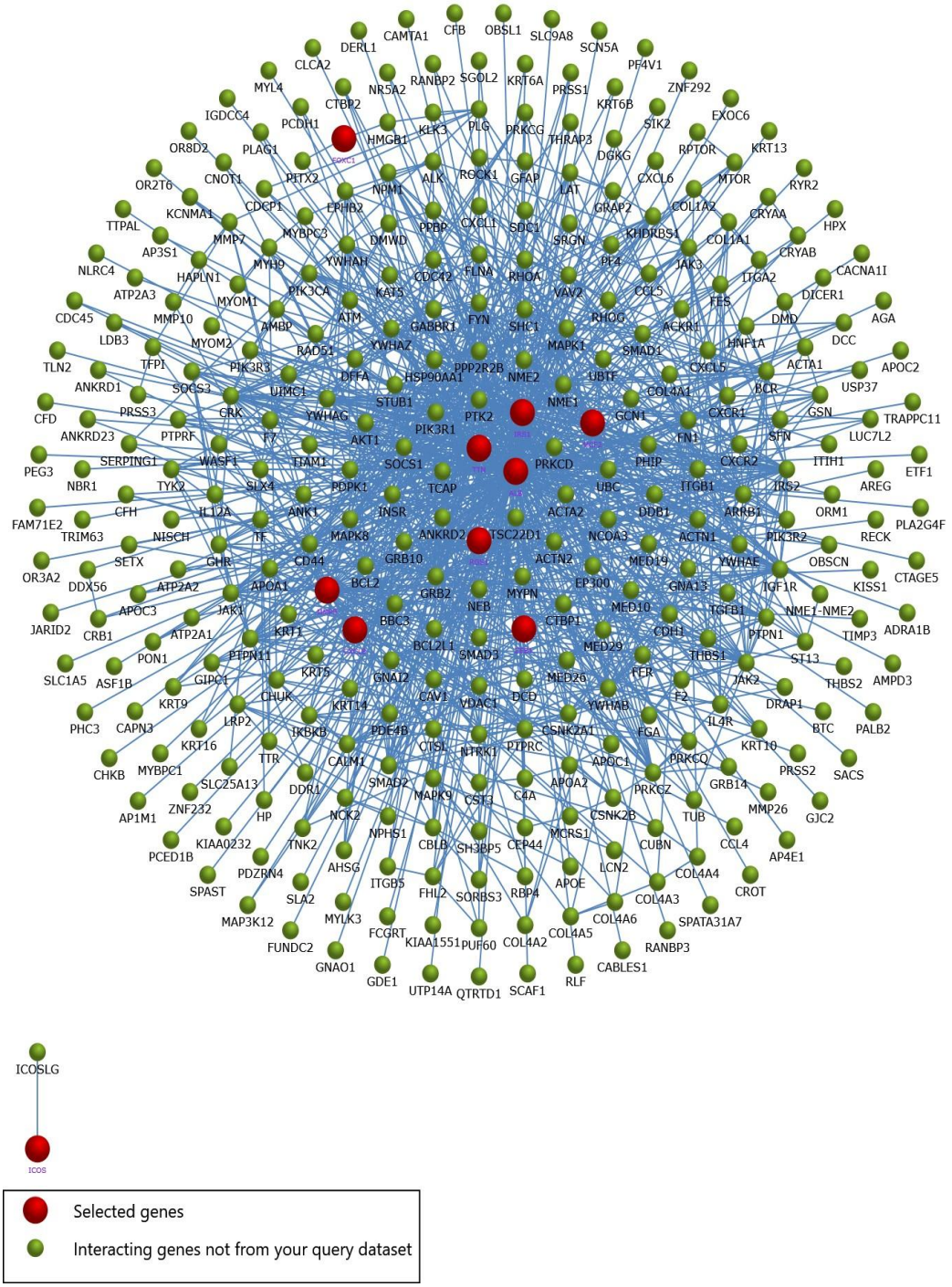


Figure 5 (c). The interaction network of the hub genes and their neighbours, created with FunRich software

Expression validation and survival analysis of the hub genes in CRC

To validate the potential hub genes identified for colorectal cancer and corroborate previous findings, survival analysis was conducted using multiple databases including the KM plotter, GEPIA, and UALCAN. These analyses revealed the prognostic significance of the selected hub genes. Four genes exhibited statistically significant differential expression ($P < 0.05$) in patients with CRC. Elevated expression levels of CXCL8, FOXC1, ICOS, and MCF2 were associated with diminished survival rates in colorectal cancer patients (Figure 6(a)). Moreover, GEPIA analysis showed that the expression levels of four hub genes (CXCL8, FOXC1, ICOS, and MCF2) were significantly elevated in tumor samples (Figure 6(b)). These findings further corroborated that the increased expression levels of the hub genes, in conjunction with the survival data from patient samples analyzed using the GEPIA database, serve as significant prognostic indicators of colorectal cancer tumorigenesis.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 19 June 2024 Revised: 03 July 2024 Accepted: 17 July 2024

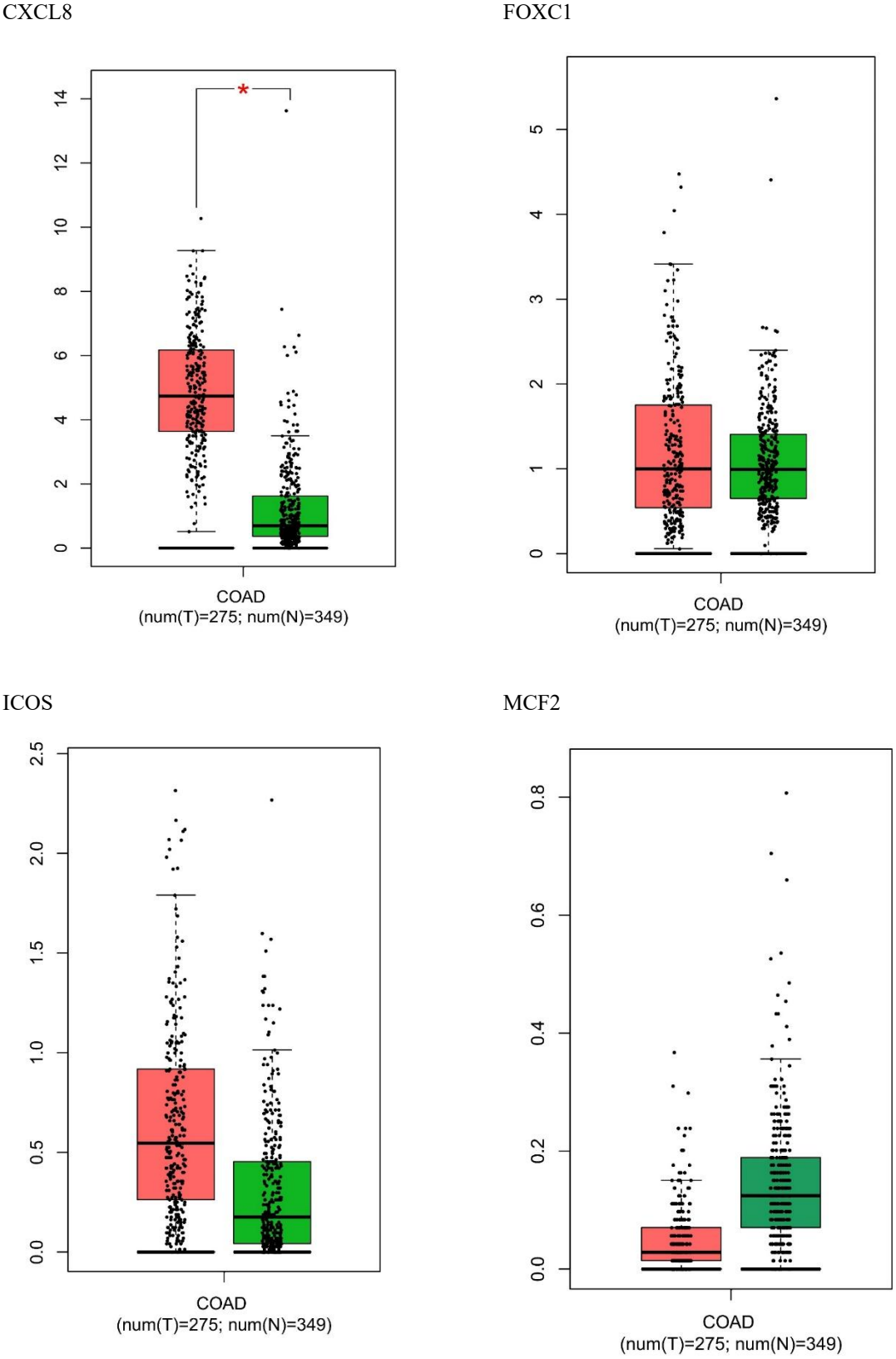
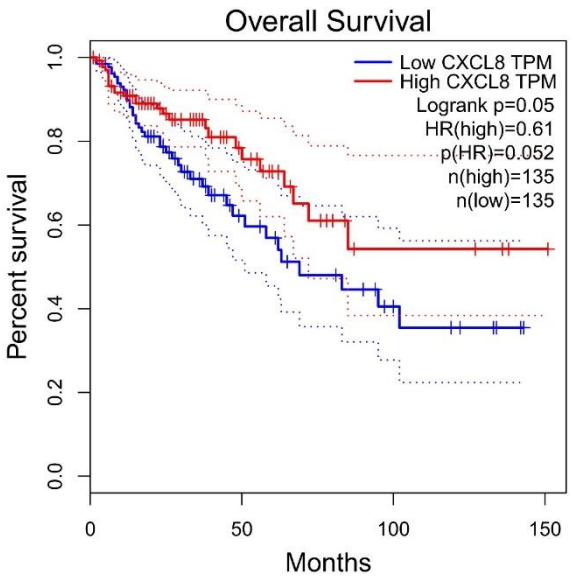


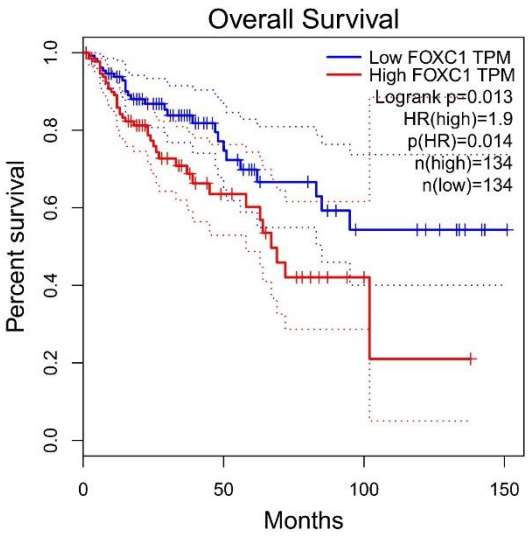
Figure 6. Expression and Survival Analysis of Hub Genes (CXCL8, FOXC1, ICOS, and MCF2). (a) Kaplan-Meier survival plots showing that elevated expression of the hub genes is associated with lower survival rates

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 19 June 2024 Revised: 03 July 2024 Accepted: 17 July 2024

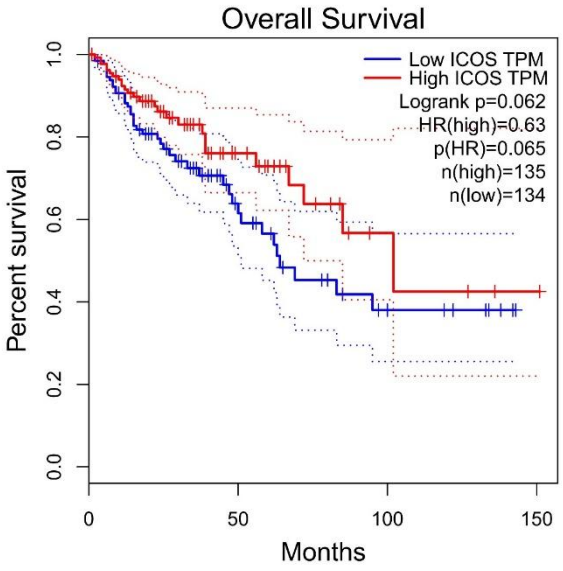
CXCL8



FOXC1



ICOS



MCF2

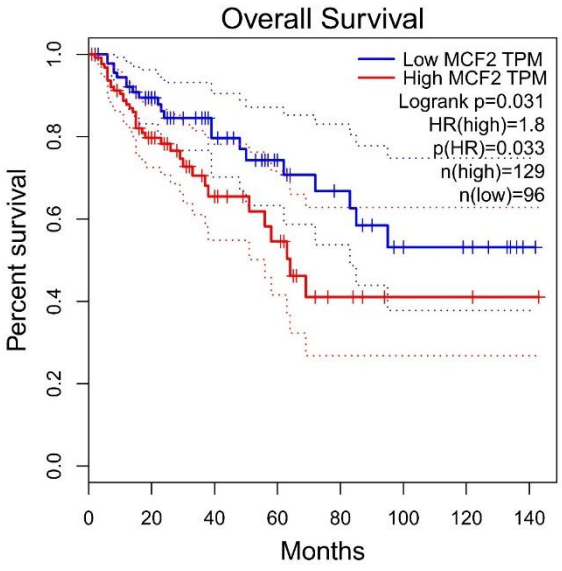


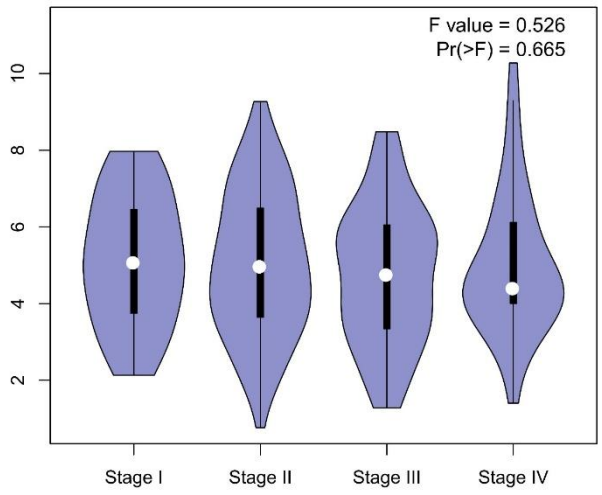
Figure 6 (b). Box plots from the GEPIA database showing significantly higher mRNA expression levels of the hub genes in tumor samples (orange) compared to normal samples (blue)

Analysis of the prognostic value of the hub genes

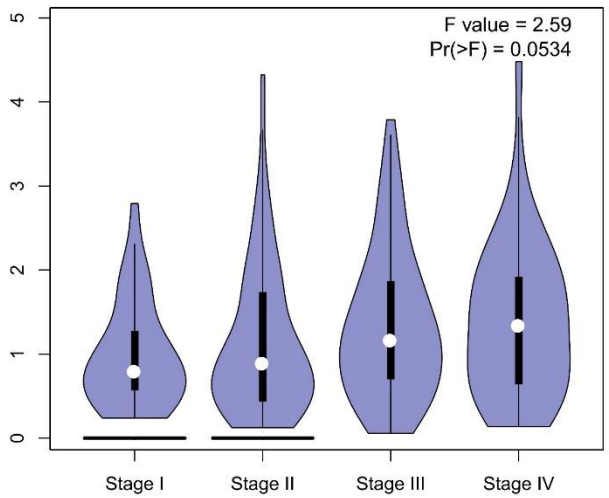
Investigating the prognostic importance of oncogenes is crucial for understanding their role in cancer progression. To validate the prognostic significance of hub genes, such as CXCL8, FOXC1, ICOS, and MCF2, the GEPIA database was employed, and the results are presented in (Figure 7(a)). Examination of mRNA expression in CRC datasets revealed markedly increased levels of these hub genes during advanced stages. Additionally, the UALCAN database was used to assess the nodal metastasis status, consistently demonstrating enhanced metastatic potential across all hub genes, as illustrated in (Figure 7(b)). These observations collectively indicate that hub genes display oncogenic characteristics and contribute to the development of colorectal cancer (CRC).

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 19 June 2024 Revised: 03 July 2024 Accepted: 17 July 2024

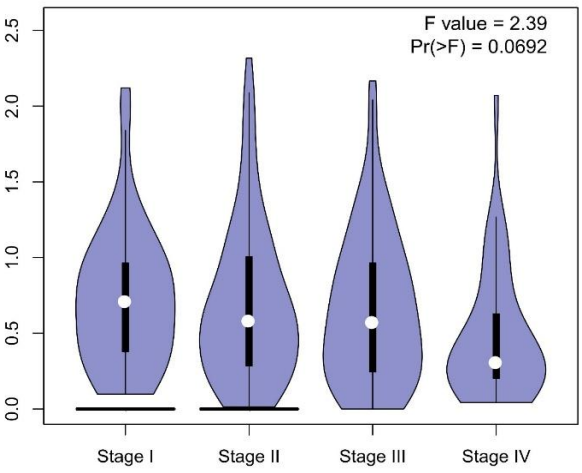
CXCL8



FOXC1



ICOS



MCF2

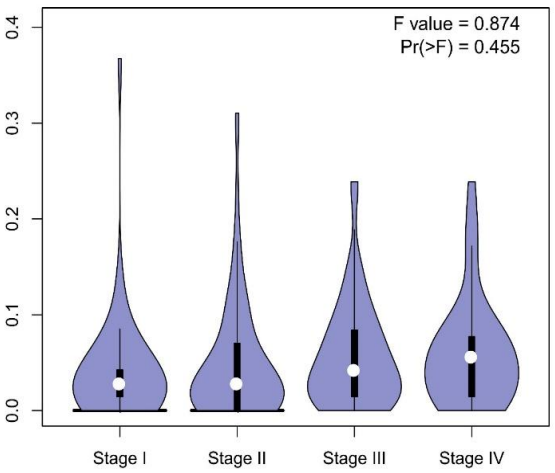
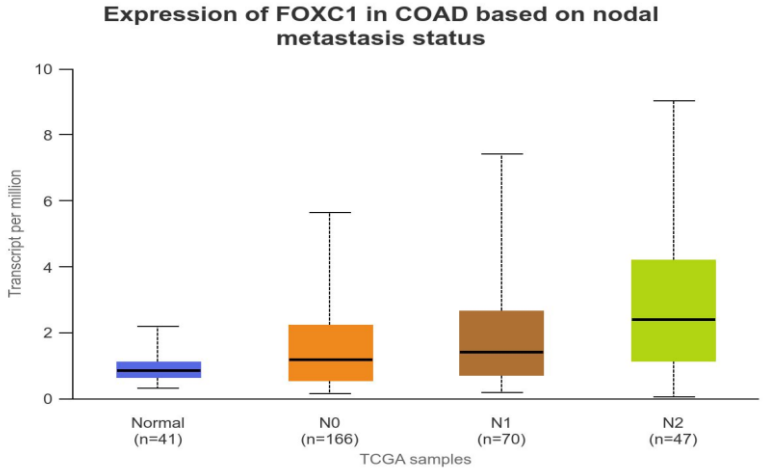


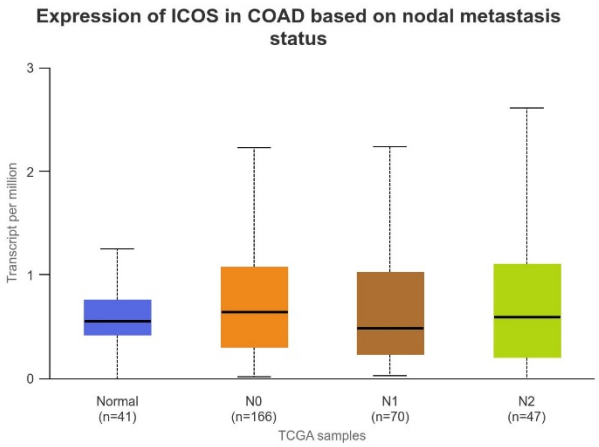
Figure 7 (a) mRNA expression levels of hub genes across different cancer stages (I-IV)

FOXC1



REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 19 June 2024 Revised: 03 July 2024 Accepted: 17 July 2024

ICOS



MCF2

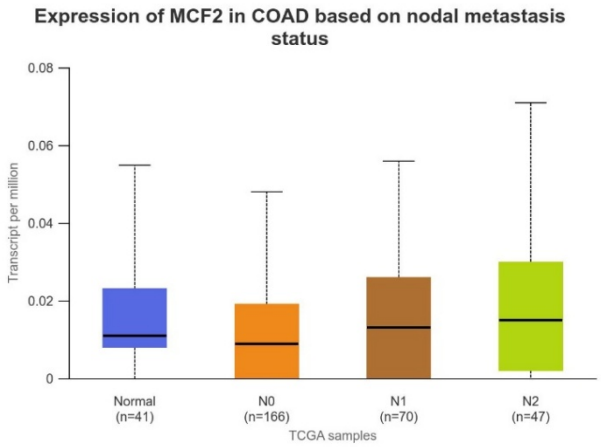
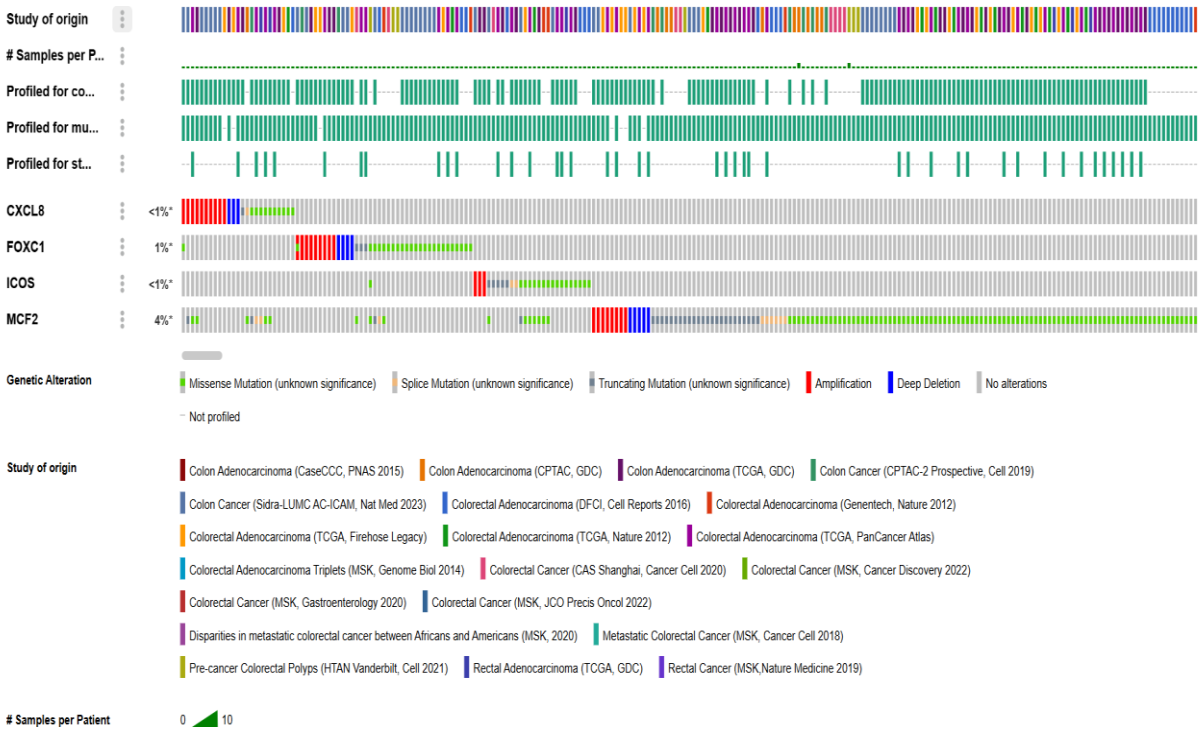


Figure 7 (b). Nodal metastasis status (N0–N3) for the hub genes, showing increased expression is linked to greater metastatic potential

Figure 7. Prognostic Value of Hub Genes Across Cancer Stages and Nodal Metastasis

Identification of mutations in hub genes

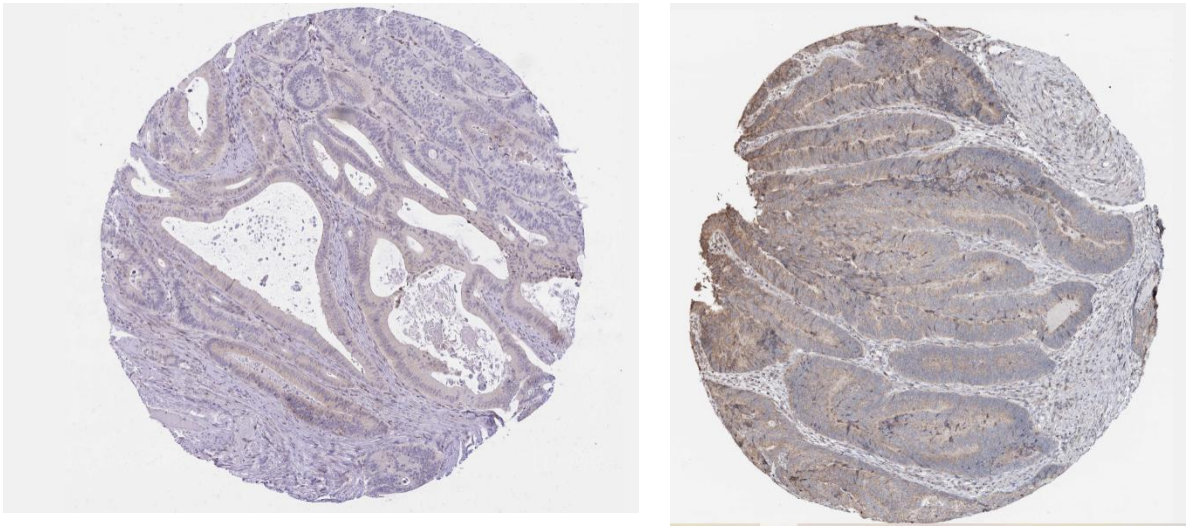
We utilized the cBioPortal database to examine genetic variations in hub genes associated with colorectal cancer (CRC). Our analysis focused on the expression levels and mutation status of four hub genes (CXCL8, FOXC1, ICOS, and MCF2) across 20 CRC studies, encompassing 5,855 samples. Figure 8(a) displays the frequency and types of mutations observed in these hub genes in the CRC samples. The analysis showed that Mutations were detected in approximately 4% of the 5,855 samples. Mutation plots for CXCL8, FOXC1, ICOS, and MCF2 in CRC indicated a mutation rate surpassing 1.0% in patient samples from The Cancer Genome Atlas (TCGA) datasets. Furthermore, we validated the expression levels of these two hub genes CXCL8 and ICOS in CRC using immunostaining data from the Human Protein Atlas (HPA) database, as shown in (Figure 8(b)). Immunostaining results corroborated that CXCL8 and ICOS were notably upregulated in CRC tissues. These observations add to the mounting evidence that alterations in these hub genes correlate with the clinical characteristics of patient samples, implying their potential role in CRC initiation and progression.



(a) Mutation frequencies and types (amplifications, deletions, etc.) for CXCL8, FOXC1, ICOS, and MCF2 in CRC samples from the cBioPortal database

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 19 June 2024 Revised: 03 July 2024 Accepted: 17 July 2024

CXCL8 FOXC1



(b) Immunohistochemistry images from the Human Protein Atlas, validating the upregulation of CXCL8 and FOXC1 protein levels in CRC tissues.

Figure 8. Genetic Alterations and Protein Expression Validation of Hub Genes.

Discussion

Colorectal cancer (CRC) is the second leading cause of cancer-related mortality in the United States and is responsible for more than 50,000 deaths by 2024, a number surpassed only by lung and bronchus cancer. Annually, an estimated 150,000 new cases are diagnosed, with a slightly higher incidence in males than in females. In the field of colorectal cancer research, microarray technology has emerged as a potent analytical tool for examining gene expression patterns, offering valuable insights into the processes of carcinogenesis and enhancing diagnostic and therapeutic approaches(24). It also reveals the complex biology of CRC, contributing to improved patient care through personalized medicine (25).

Examination of the GSE164191 dataset revealed 145 genes exhibiting differential expression in colorectal cancer (CRC) tissues compared to normal tissues, comprising 89 upregulated and 56 downregulated genes. This notable variation in gene expression, characterized by a p-value < 0.05 and $|\log FC| > 1$, emphasizes the intricate nature of gene regulation in CRC. Elucidating these differentially expressed genes (DEGs) is of paramount importance given their potential as diagnostic biomarkers and therapeutic targets.

Key upregulated genes, including CXCL8, FOXC1, ICOS, and MCF2, are known to play important roles in tumor progression and metastasis. Conversely, downregulated genes may represent loss of function in pathways crucial for cellular stability, possibly contributing to tumorigenesis. The identification of DEGs underscores the need to explore their specific roles in CRC. Using the DAVID tool for functional enrichment analysis, we gained insight into the biological processes, cellular components, and molecular functions associated with these differentially expressed genes (DEG). The upregulated genes were primarily involved in transcriptional regulation, immune responses, and cell adhesion processes, which are likely to be disrupted in CRC. For example, the involvement of DEGs in RNA polymerase II-mediated transcription suggests that abnormal gene expression may drive oncogenesis.

In contrast, the downregulated genes were linked to essential functions, such as calcium-dependent –cell adhesion and glucose homeostasis, both of which are critical for tissue integrity and metabolic stability. Disruption of these processes may promote tumor invasion and metastasis, emphasizing the functional importance of these differentially expressed genes (DEG)s

Differentially expressed genes (DEG) were analyzed using a protein-protein interaction (PPI) network that revealed strong connections, forming a network with 131 nodes and 52 edges. Hub genes, such as CXCL8, FOXC1, ICOS, and MCF2, which are highly connected within this network, play pivotal roles in CRC biology. This analysis of the PPI network highlights the interconnected nature of these genes, suggesting possible pathways for therapeutic interventions.

Tools such as STRING and FunRich were used to analyze protein-protein interactions, offering a broader view of the molecular landscape of CRC. This approach helps to pinpoint regulatory nodes that may influence tumor behavior, which is key to developing therapies that disrupt these networks and halt tumor growth.

Survival analysis using databases such as KM Plotter and GEPIA demonstrated that elevated expression levels of specific hub genes, notably CXCL8, FOXC1, ICOS, and MCF2, correlated with decreased survival rates in patients with CRC. These observations indicate that these genes may function as prognostic indicators associated not only with tumor development but also with disease progression.

Validation across multiple databases has strengthened the utility of these genes as prognostic indicators. Additionally, analysis revealed that these hub genes had mutation rates above 1% in patient samples, implying that genetic changes may enhance

their oncogenic effects. This highlights the importance of integrating genetic and expression data to fully understand the role of these genes in CRC.

The identified hub genes play critical roles in CRC pathophysiology. CXCL8 is a pro-inflammatory cytokine known for promoting angiogenesis, tumor growth, and immune cell recruitment within the tumor microenvironment. Elevated CXCL8 expression in CRC may contribute to immune evasion and increased metastatic potential(26). FOXC1, a transcription factor, is associated with epithelial-mesenchymal transition (EMT), facilitating tumor invasion and metastasis(27). Similarly, ICOS, an immune checkpoint regulator, is implicated in immune modulation, potentially influencing CRC progression by enhancing immune suppression(28). MCF2, a guanine nucleotide exchange factor, is involved in cellular signaling pathways, including cytoskeletal rearrangements and cell migration(29). These roles suggest that these genes are pivotal in tumor development and metastasis, warranting further investigation into their molecular functions in CRC.

Our findings correlate with earlier studies on the oncogenic roles of CXCL8 and FOXC1 in several cancers, including colorectal cancer. Nonetheless, the identification of ICOS and MCF2 as crucial factors in CRC is relatively unprecedented, emphasizing their potential as novel biomarkers. While previous studies investigated these genes in different experimental frameworks, our multi-omics analysis uniquely combines survival analysis with expression data, providing clinically relevant insights.

The identified hub genes provide potential opportunities for clinical applications. CXCL8 and FOXC1, as regulators of angiogenesis and metastasis, demonstrate feasibility as therapeutic targets. Developing inhibitors or monoclonal antibodies against these proteins could restrict tumor progression. ICOS could serve as a target for immune checkpoint inhibitors to enhance anti-tumor immunity. Furthermore, MCF2's role in cell signaling underscores its potential as a target by interfering with metastatic pathway.

Future research should authenticate these outcomes through experimental studies, including quantitative RT-PCR and immunohistochemistry, to validate differential gene expression in CRC tissues. Analyzing pathways involving these hub genes (e.g., KEGG pathways) in greater detail could reveal applicable molecular targets for drug development. Combining multi-omics data, such as proteomics and metabolomics, could provide a more comprehensive understanding of CRC biology.

The translational potential of these results is significant. High expression levels of CXCL8, FOXC1, ICOS, and MCF2 were correlated with poor survival outcomes, validating their potential as prognostic biomarkers. These genes could also be analyzed as predictive markers for personalized medicine, assisting in the identification of CRC patients based on risk profiles or treatment responses. Their role in immune modulation and metastasis underscores their potential to enhance diagnostic precision and therapeutic strategies.

This study provides valuable insights into the molecular landscape of CRC by investigating and elucidating key hub genes. These findings underscore the importance of integrating bioinformatics with clinical data to identify novel biomarkers and therapeutic targets. Further studies should aim to experimentally validate these genes and explore their potential for clinical translation, ultimately advancing personalized medicine in CRC.

Conclusion

This study provides a comprehensive examination of aberrant gene expression in colorectal and sheds light on its potential implications for tumor biology. The elucidation of differentially expressed genes, in conjunction with functional enrichment and protein-protein interaction network analyses, provides crucial insights into the molecular underpinnings of CRC. Moreover, the prognostic relevance of the identified hub genes underscores the need for additional research to explore their prospective utility as therapeutic targets and biomarkers in CRC management. Subsequent studies should aim to delineate the precise roles of these genes in CRC progression and elucidate their interactions within the broader landscape of tumor biology.

Funding Statement: This study did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sector.

Ethical Compliance: This study did not involve human participants or animal studies; therefore, ethical approval was not required.

Data Access Statement: Research data supporting this publication are available from the corresponding author upon reasonable request.

Conflict of Interest Declaration: The authors declare that they have no affiliations with or involvement in any organization or entity with any financial interests in the subject matter or materials discussed in this manuscript.

Author Contributions: Suresh Kumar Gopal contributed to the design and implementation of the research, Kamalakannan Dhanabalan contributed to the analysis of the results and the writing of the manuscript, and Manivannan Rangasamy conceived the original idea and supervised the project.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: 19 June 2024 Revised: 03 July 2024 Accepted: 17 July 2024

References

1. Patil PS, Saklani A, Gambhire P, Mehta S, Engineer R, De'Souza A, et al. Colorectal cancer in India: an audit from a tertiary center in a low prevalence area. Indian Journal of Surgical Oncology. 2017;8(4):484-490.
2. Pitchumoni CS, Broder A. Epidemiology of colorectal cancer. In: Colorectal neoplasia and the colorectal microbiome. Academic Press; 2020. p. 5-33.
3. Mahajan D. Unmasking the Silent Threat: Colorectal Cancer's Alarming Surge in India's Young Adults. Journal of Surgical Specialties and Rural Practice. 2023;4(3):111-113.
4. Alaryani FS, Alrdahe SST. A review of treatment, risk factors, and incidence of colorectal cancer. International Journal of Applied Pharmaceutics. 2022;14:1-6.
5. Ferreira AM, Chodankar SU, Vaz FS, D'souza DB, Kulkarni MS. Risk factors for colorectal cancer in Goa, India: A hospital-based case–control study. Indian Journal of Community Medicine. 2021;46(3):474-478.
6. Jung G, Hernandez-Illan E, Moreira L, Balaguer F, Goel A. Epigenetics of colorectal cancer: biomarker and therapeutic potential. Nature Reviews Gastroenterology & Hepatology. 2020;17(2):111-130.
7. Lucafò M, Curci D, Franzin M, Decorti G, Stocco G. Inflammatory bowel disease and risk of colorectal cancer: An overview from pathophysiology to pharmacological prevention. Frontiers in Pharmacology. 2021;12:772101.
8. Vanden Heuvel JP. Analysis of gene expression. In: PCR Protocols in Molecular Toxicology. CRC Press; 2019. p. 41-98.
9. Fajriyah R. Paper review: An overview on microarray technologies. Bulletin of Applied Mathematics and Mathematics Education. 2021;1(1):21.
10. Gawehn E, Schneider G, Hiss JA. Deep Learning in Drug Discovery. Molecular Informatics. 2016;35(1):3-14. doi:10.1002/minf.201501008..
11. Clough E, Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. Nucleic Acids Research. 2024;52(D1):D138-D144. doi:10.1093/nar/gkad965.
12. Kadali SSK, Gowlikar R, Fatima SN. The Cancer Genomic Atlas–"TO CONQUER CANCER". International Journal of Molecular and Immuno Oncology. 2021;6(2):76-81.
13. Li X, Li MW, Zhang YN, Xu HM. Common cancer genetic analysis methods and application study based on TCGA database. Hereditas. 2019;41(3):234-242.
14. Chandrashekar DS, Karthikeyan SK, Korla PK, Patel H, Shovon AR, Athar M, et al. UALCAN: An update to the integrated cancer data analysis platform. Neoplasia. 2022;25:18-27.
15. Yichao Y, Yongbai L, Hailiang L, Dongbo C, Bo L, Murshed A. COL1A1, COL1A2, CHN1, and FN1 Promote Tumorogenesis and Act as Markers of Diagnosis and Survival in Gastric Cancer Patients. Current Pharmaceutical Biotechnology. 2025. doi:10.2174/0113892010332329241119104430.
16. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Research. 2021;49(D1):D605-D612. doi:10.1093/nar/gkaa1074.
17. Digre A, Lindskog C. The Human Protein Atlas—Spatial localization of the human proteome in health and disease. Protein Science. 2021;30(1):218-233. doi:10.1002/pro.3987.
18. Almokhtar BJ, Elengoe A. Determination of the EGFR Gene Role in Lung Cancer Pathway using STRING and Cytoscape Software. Letters in Applied NanoBioScience. 2024;13(3):131. doi:10.33263/LIANBS133.131.
19. Fu L, Zhao L, Li F, Wen F, Zhang P, Yang X, et al. Pharmacological mechanism of quercetin in the treatment of colorectal cancer by network pharmacology and molecular simulation. Journal of Biomolecular Structure and Dynamics. 2024;42(13):7065-7076. doi:10.1080/07391102.2023.2235589.
20. Sun Z, Xia W, Lyu Y, Song Y, Wang M, Zhang R, et al. Immune-related gene expression signatures in colorectal cancer. Oncology Letters. 2021;22(1):543.
21. Wang C, Zhang L. Bioinformatics-based identification of key genes and pathways associated with colorectal cancer diagnosis, treatment, and prognosis. Medicine. 2022;101(37):e30619.
22. Atasever S. Identification of potential hub genes as biomarkers for breast, ovarian, and endometrial cancers. Frontiers in Life Sciences and Related Technologies. 2024;5(1):74-82. doi:10.51753/flsrt.1405816.
23. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Research. 2023;51(D1):D638-D646.
24. Lan Y, Yang X, Wei Y, Tian Z, Zhang L, Zhou J. Explore key genes and mechanisms involved in colon cancer progression based on bioinformatics analysis. Applied Biochemistry and Biotechnology. 2024;196(9):6253-6268.
25. 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 Medical Genomics. 2023;16(1):64.
26. Asokan S, Bandapalli OR. CXCL8 signaling in the tumor microenvironment. In: Tumor Microenvironment: The Role of Chemokines–Part B. Academic Press; 2021. p. 25-39.

27. Zhang H, Nie J, Bao Z, Shi Y, Gong J, Li H. FOXC1 promotes EMT and colorectal cancer progression by attracting M2 macrophages via the TGF-β/Smad2/3/snail pathway. Cellular Signalling. 2025;130:111680.
28. Yang J, Liu J, Chen Y, Tang W, Bo K, Sun Y, et al. Investigation of ICOS, CD28 and CD80 polymorphisms with the risk of hepatocellular carcinoma: a case–control study in eastern Chinese population. Bioscience Reports. 2019;39(7):BSR20181824.
29. Zaritsky A, Tseng YY, Rabadán MA, Krishna S, Overholtzer M, Danuser G, et al. Diverse roles of guanine nucleotide exchange factors in regulating collective cell migration. Journal of Cell Biology. 2017;216(6):1543-1556.