Research Journal of Computer and Information Technology Sciences _____ISSN 2320 – 6527
Vol. **13(1),** 1-11, June **(2025)** *Res. J. Computer and IT Sci.*

Differential Gene Expression Analysis and Molecular Docking to Identify Potential Biomarkers of Colorectal Cancer

Pujaa B.* and Madhumitha G.
Department of Bioinformatics, Stella Maris College, Chennai, Tamilnadu, India
pujaa@stellamariscollege.edu.in

Available online at: www.isca.in, www.isca.me
Received 22nd November 2024, revised 6th April 2025, accepted 11th May 2025

Abstract

Colorectal cancer is one of the leading cancer types worldwide in terms of incidence and mortality. This type of cancer originates from the large intestine, near the gastrointestinal tract which includes the colon and rectum. Colon familial colorectal cancer refers to a situation where the first degree relatives have a higher chance of developing the disease with a high mortality rate. Considering this, the current study integrated a Bioinformatics approach to elucidate the gene expression patterns and molecular interactions associated with colorectal cancer (CRC) progression. Gene expression profiling revealed differential expression of key genes (CXCL1, CXCL2, CXCL3, COL1A1, COL1A2) across different stages of colon adenocarcinoma. Notably, early-stage samples exhibited elevated expression of CXCL genes, while MMP1 and MMP3 were upregulated in advanced stages. Gene Ontology analysis highlighted enrichment in processes related to extracellular matrix organization and proteolysis. Prognostic biomarkers, neurotransmitters, neuropeptides, and intracellular signaling pathways were identified for the CRC survival prediction. Protein-protein interaction network analysis identified key modules, unveiling protein clusters for further investigation. Molecular docking analysis targeting CRC-associated proteins CXCL2 and CXCL3 identified potential therapeutic compounds with high affinity, including Isorhamnetin, Kaempferol, Jaranol, Curcumin, EGCG, Hesperidin, Resveratrol, Liquiritin, and Baicalin. This comprehensive analysis provides valuable insights into the molecular landscape of CRC, identifying potential therapeutic targets and biomarkers for precision medicine strategies in CRC management.

Keywords: Colorectal cancer, Bioinformatics, Biomarkers, Differential gene expression, Molecular Docking.

Introduction

Colorectal cancer is a multifaceted disease with complex molecular mechanisms influencing its onset and progression. CRC has been observed to exhibit familial clustering patterns since the early 1900s. Like the majority of cancers, colorectal cancer can develop as a result of mutations in specific genes. Oncogenes, tumor suppressor genes, and genes involved in DNA repair pathways might all possess these modifications. Colorectal carcinomas can be classified as sporadic, hereditary, or familial based on where the mutation originated. Sporadic tumors, which are about 70% of all colorectal cancers, are malignancies caused by point mutations. Because multiple genes might be targeted by mutations, the molecular pathophysiology of sporadic cancer is different[1].

Less than five percent of cases of colorectal cancers belong to the hereditary type. These malignancies are brought on by hereditary mutations that inevitably lead to the growth of cancer cells. Familial colorectal cancer is a condition where there is a higher chance of the disease developing within a family, especially among first-degree relatives[2]. Patients with ulcerative colitis have a 3.7% increased risk of acquiring colorectal cancer, and those with Crohn's disease have a 2.5% increased risk[3]. Colorectal cancer claims the lives of over a million people worldwide, annihilating over half a million. In Europe, CRC ranked third among aggressive cancer types in both men and women[4].

With thousands of new cases and fatalities expected in the upcoming years, the current prospects are not that promising[5]. It is therefore extremely important to understand the progressive mechanism of this disease and to find new solutions to treat the patients.

Tests leveraging gene expression profiling and biomarker identifications are growing progressively more common these days. These techniques help in examining the expression of genes in samples of tumor and normal tissue, as well as samples from various disease stages. Various findings from this field of research helped in developing novel therapeutics in oncology[6]. These comparative studies are considered to be a valuable source of information concerning the disease prognosis and the most effective course of action for individual patients[7]. The identification of critical genetic or epigenetic changes in carcinogenesis, tumor growth, metastasis, and recurrence, as well as promising cancer biomarkers for diagnosis, prognosis, and treatment prediction, have been made easier by significant advances in microarray and high-throughput sequencing technologies. Integrated bioinformatics analysis, a comprehensive strategy to increase sample size, unify cross-platform standardization of expression profiles, and eliminate invalid raw data, has been widely adopted to identify differentially expressed genes (DEGs) at mRNA and non-coding RNA level in colorectal cancer (CRC) to overcome the limitations and obtain convincing results[8].

This study endeavors to dive into the intricate molecular landscape of CRC, employing a multifaceted approach merging bioinformatics analyses and computational tools. Through this comprehensive investigation, the study aims to achieve several key objectives: Firstly, it seeks to elucidate the gene expression patterns and molecular interactions pivotal in CRC development, to identify novel therapeutic targets and prognostic biomarkers. Furthermore, it endeavors to unravel the enriched Gene Ontology categories related to upregulated and down regulated differentially expressed genes (DEGs) in CRC, thereby shedding light on the biological processes, molecular functions, and cellular components implicated in CRC pathogenesis. The study focuses on the identification of prognostic biomarkers crucial for predicting CRC survival outcomes. Through the exploration of protein-protein interaction networks, it aims to unveil key modules and signaling pathways intricately involved in CRC progression. By combining computational analyses, molecular docking and bioinformatics approaches, this study aspires to offer valuable insights into the molecular landscape of CRC. Ultimately, these insights are intended to contribute to the advancement of precision medicine strategies for more effective CRC management to improve patient outcomes and quality of life.

Methodology

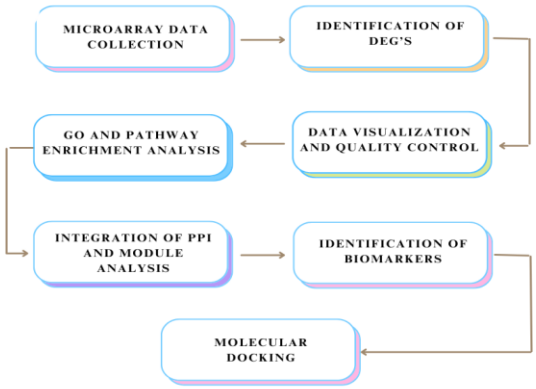


Figure-1: Graphical abstract of methodology

Data Collection: The microarray datasets GSE20916 and GSE14526 were retrieved from GEO database. Both datasets are based on the Affymetrix GPL570 platform (Affymetrix Human Genome U133 Plus2.0 Array). On the other hand, the GSE184093 dataset are from the Agilent GPL20115 platform. Collectively, these datasets encompass a total of 55 normal samples and 131 primary colorectal cancer samples, providing a diverse set of data for studying gene expression patterns and potential biomarkers associated with CRC[9].

Identification of DEGs: Differential gene expression analysis was done using Limma package in R software. Log2 transformation threshold of > 2 for upregulated genes and < 2 for down regulated genes was applied with a P-value cut off <0.05 for the identification of statistically significant genes. A table of top significant genes, based on the defined statistical criteria highlighting the most relevant differentially expressed genes in the dataset was generated. The R software generated volcano plots highlighting all significant DEGs. Using the Morpheus online tool (https://software.broadinstitute.org/morpheus/), heat maps were generated highlighting the expressions of top 10 upregulated and down regulated DEGs.

GO and pathway enrichment analysis for DEG's: The Database for Annotation, Visualization, and Integrated Discovery (DAVID) provides comprehensive functional annotation tools that help analyze the biological meaning behind large lists of genes. The list of DEGs, along with the gene identifiers such as Entrez Gene IDs or gene symbols and the appropriate organism were uploaded for GO (functional annotation) and KEGG (pathway enrichment) analysis[10]. Terms/pathways with P-values <0.05 were considered statistically significant. For interpretation, DAVID provided graphs and tables summarizing the enriched terms and pathways[11].

Integration of protein-protein interaction (PPI) network: STRING is a database of known and predicted protein-protein interactions, including direct (physical) and indirect (functional) associations derived from computational prediction, knowledge transfer between organisms, and aggregated interactions from primary databases. The significant upregulated and all down regulated DEGs were uploaded along with recognized gene identifiers. Interaction parameters included only those with experimentally validated combined scores > 0.4. A network was constructed from these interactions, representing the associations among the encoded proteins. This network was then imported to Cytoscape, an open access tool which helps in visualizing molecular interaction networks and integration with gene expression profiles[12]. The densely connected regions (primary modules) were identified using the Molecular Complex Detection (MCODE) plug-in with default parameters. Functional annotation (GO) and pathway analysis (KEGG) of DEGs within these modules were performed using DAVID with P-value <0.05.

Identification of prognostic biomarkers: The Cancer Genome Atlas (TCGA) was the primary data source for this analysis. By comparing gene expression levels between tumor and normal samples, differentially expressed genes in colorectal cancer were identified. The identified DEGs were then uploaded to UALCAN, an interactive web source for analyzing cancer OMICS data, providing access to publicly available data, identifying biomarkers, performing in silico validation, and offering various analyses including gene expression profiles and patient survival information[13]. Statistical analysis was done on UALCAN, with a threshold of P< 0.05 to identify potential differences in gene expression and prognosis. In addition, factors such as the sex of the individual that might play a crucial role in the disease prognosis were considered and those genes that lacked sufficient data were excluded[14].

Molecular Docking: BIOVIA Discovery Studio provides in silico techniques like molecular mechanics and free energy calculations, aiding in the discovery of small and large-molecule therapeutics. It offers tools for computational chemists and structural biologists to engineer novel biotherapeutics and small-molecule drugs. Totally, 23 bioactive compounds were identified from the traditional Chinese medicine SJZD for molecular docking[15]. The 3D structures of these ligands were retrieved from the PubChem database. The protein targets 6WZJ and 6WZK selected from literature review were retrieved from the Protein Data Bank (PDB) and were subjected to molecular docking using Discovery studio. Docking was performed for each ligand-protein pair using the LibDock algorithm and the best pose for the interactions were assessed. The binding modes and potential interactions between ligands and target proteins were visualized. The ligands with high docking scores and favorable binding interactions were shortlisted.

Results and Discussion

Identification of DEGs: A total of 131 primary CRC samples and 55 normal samples were utilized for the analysis. The analysis revealed 309 upregulated genes and 234 down regulated genes in tumor samples as compared to normal samples. A volcano plot representing all DEGs was created using R software, providing a comprehensive overview of the gene expression changes between CRC and normal samples. Additionally, heatmaps were generated for the top 10 upregulated and down regulated DEGs, respectively, using the Morpheus online tool (Figure-2 and 3). Mean-Difference (MD) plot was generated in R for the DEGs where, the points above the line indicate genes that are upregulated in the condition plotted on the y-axis compared to the condition on the x-axis, while points below the line represent genes that are down regulated (Figure-4 and 5). Box plot displays distribution of data and its central tendency. Longer boxes indicate more out spread data, while shorter boxes indicate more clustered data (Figure-6 and 7).

Expression density plot shows the expression level of a gene in two groups (control, test). This plot is plotted against density and intensity of the gene. The spread of the curve indicates how dispersed the data points are. A wider curve suggests greater variability, while a narrower curve suggests less variability. Concentration of the curve in a specific region indicates where most of the data points lie (Figure-8 and 9). Histogram plot represents the density, which indicates how many genes have a p-adj value within a certain range.

The plot tails off towards both 0.0 and 1.0, indicating that there are fewer genes with very low or very high p-adj values (Figure-10 and 11). t-statistic quantile plot shows the percentile of t-statistic distribution. The points in this plot align closely with a diagonal line which indicates t-statistic distribution closely matches the theoretical distribution (Figure-12 and 13).

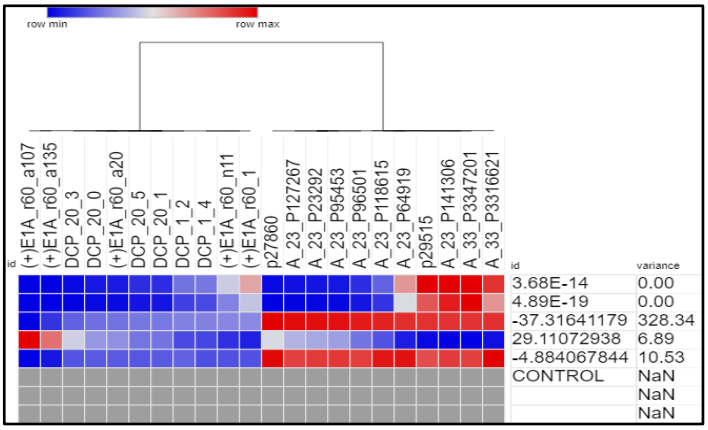


Figure-2: Heatmap of top 10 upregulated DEGs and top 10 down regulated DEGs for GSE20916.

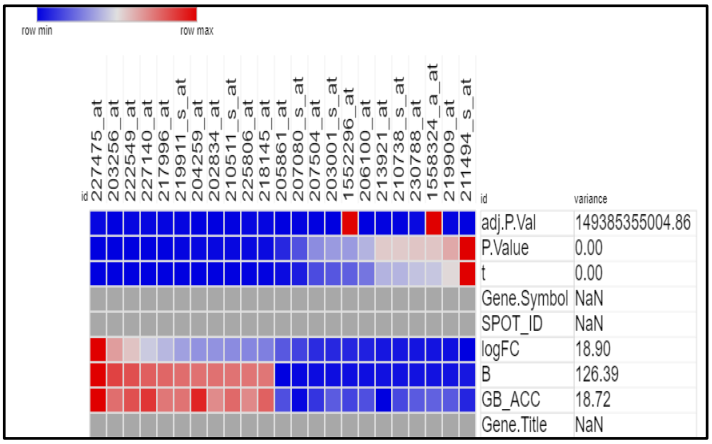


Figure-3: Heatmap of top 10 upregulated DEGs and top 10 down regulated DEGs for GSE184093.

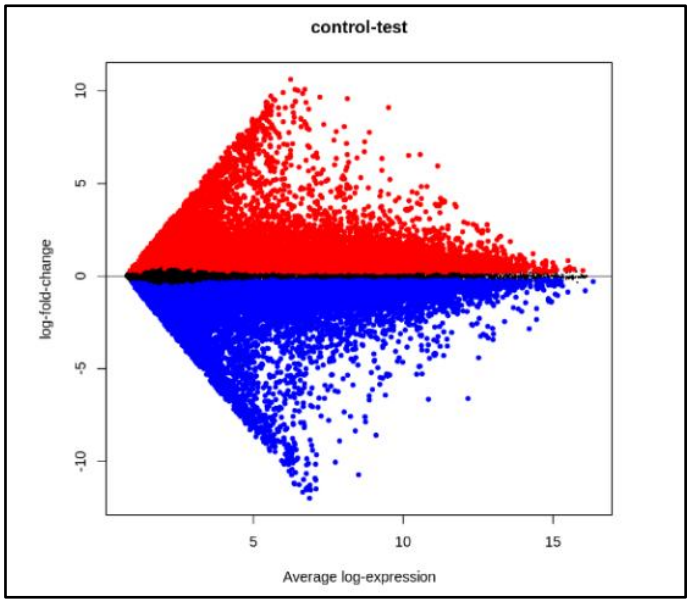


Figure-4: Mean-Difference (MD) for the DEGs present in GSE20916.

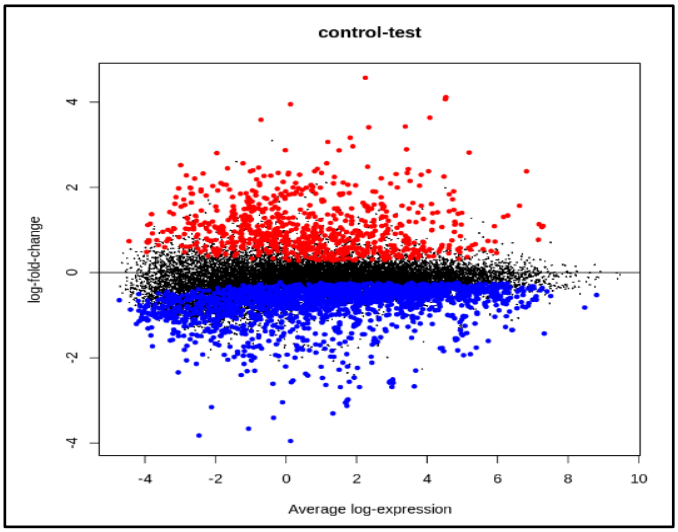


Figure-5: Mean-Difference (MD) for the DEGs present in GSE184093.

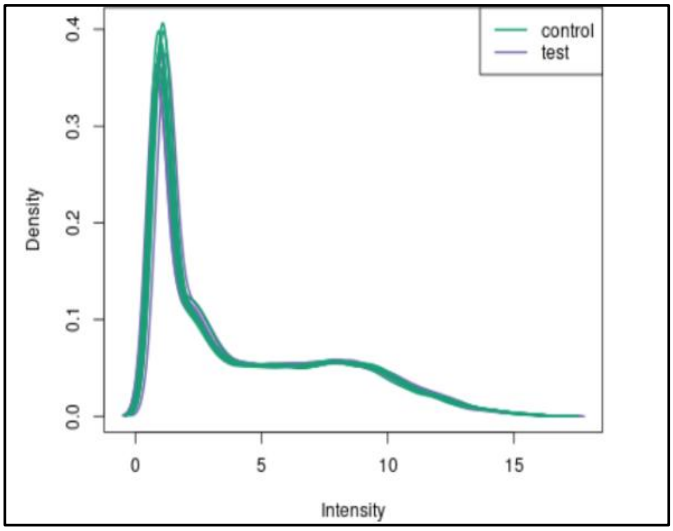


Figure-8: Expression density plot for GSE20916.

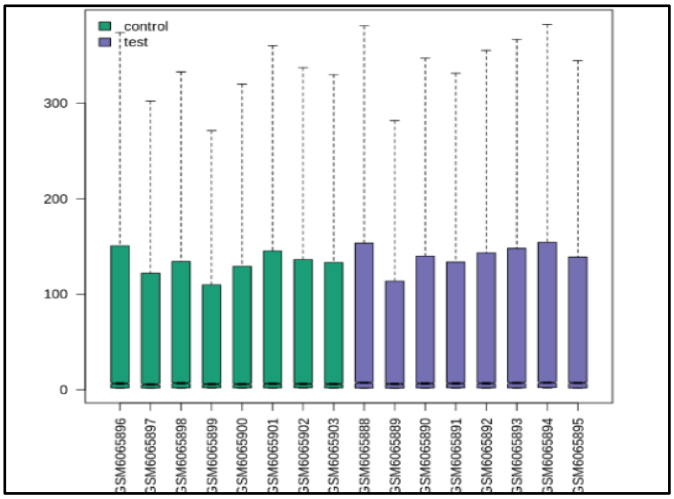


Figure-6: Box plot displaying distribution of data and its central tendency for GSE20916.

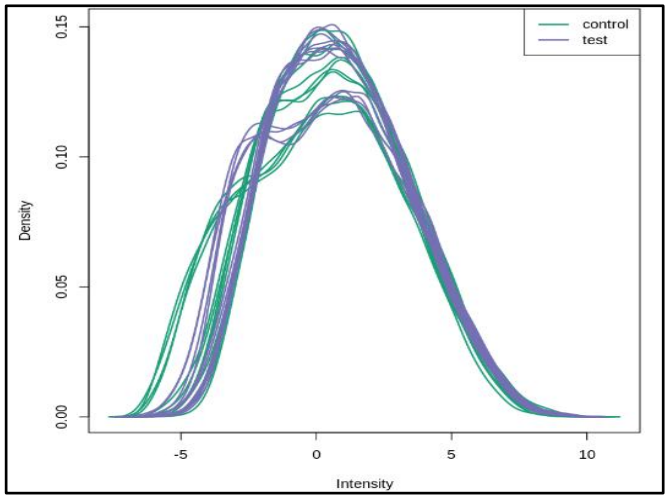


Figure-9: Expression density plot for GSE184093.

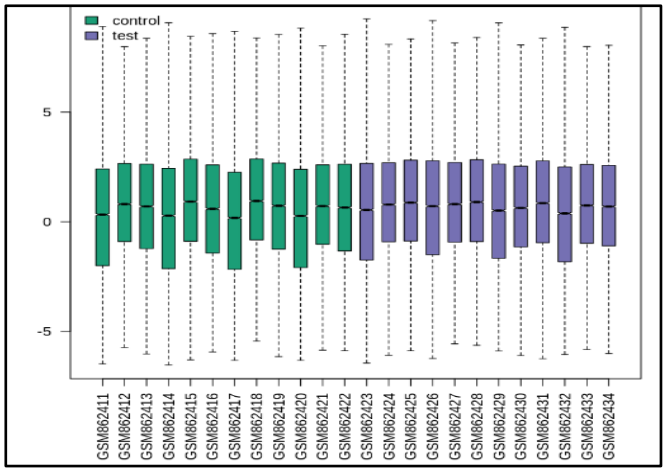


Figure-7: Box plot displaying distribution of data and its central tendency for GSE184093.

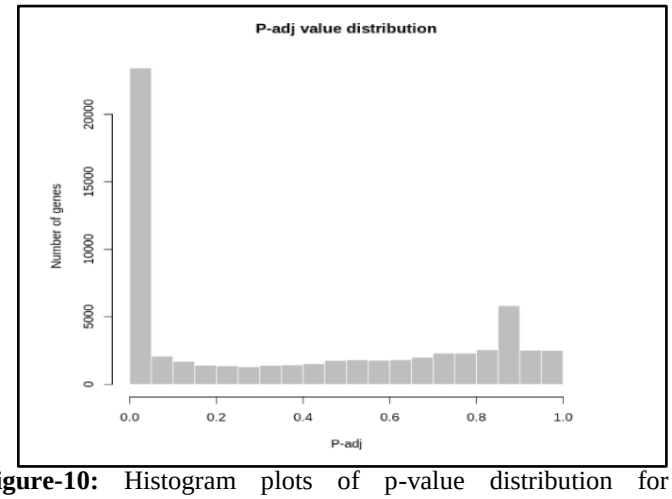


Figure-10: Histogram plots of p-value distribution for GSE20916.

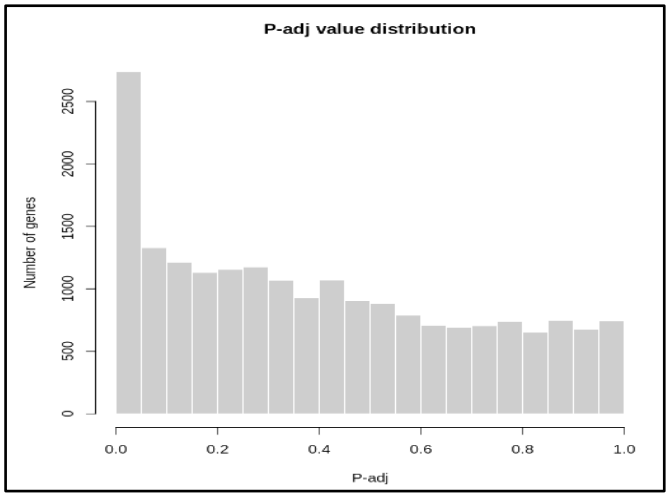


Figure-11: Histogram plots of p-value distribution for GSE184093.

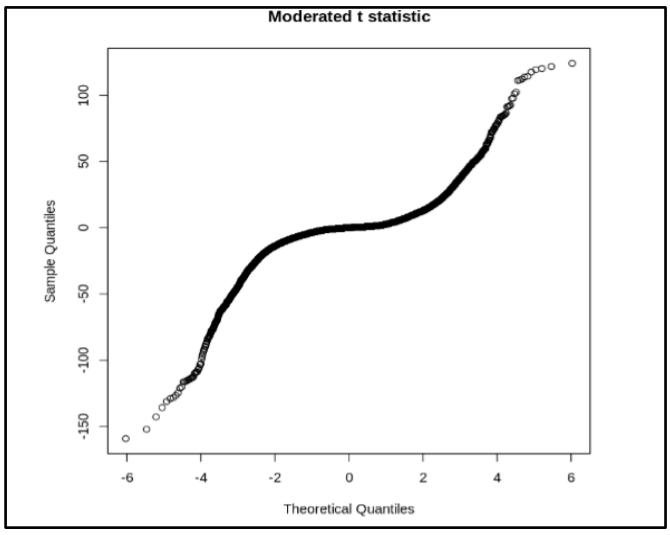


Figure-12: t- statistic quantile plots for GSE20916.

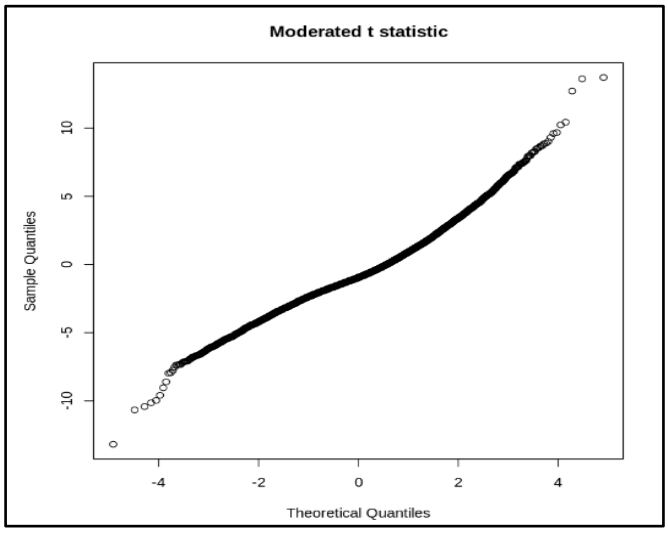


Figure-13: t- statistic quantile plots for GSE184093.

GO term enrichment analysis: The analysis identified significant enrichment in specific Gene Ontology (GO) categories for both upregulated and down regulated differentially expressed genes (DEGs). Upregulated and down regulated DEGs significantly enriched in Biological processes (BP) including Collagen breakdown (catabolism), extracellular matrix organization and assembly, response to various stimuli (lipopolysaccharide, UV-A radiation, beta-amyloid), promoting type I interferon signaling, and protein complex formation. As for Molecular functions (MF), the DEGs were enriched in Serine and metalloprotease activity (protein breakdown), extracellular matrix structural components, and zinc ion binding. With regards to Cellular components (CC), the DEGs were enriched in Extracellular matrix, exosomes (membrane-bound vesicles), cell surface, collagen trimmers, and Golgi lumen (protein processing compartment).

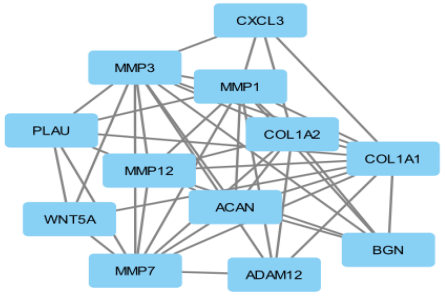
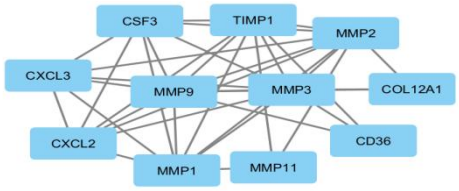
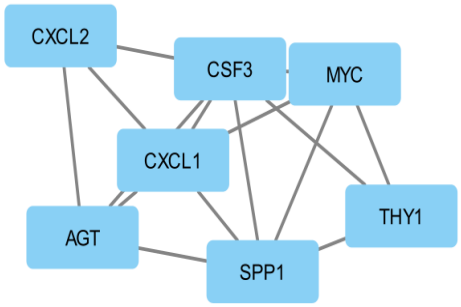
KEGG pathway analysis: Upregulated and down regulated DEGs Enriched Pathways: Cell adhesion molecules, Salivary secretion, Calcium signaling pathway, Platelet activation, IL-17 signaling pathway, Neuroactive ligand-receptor interaction, Lipid metabolism: Atherosclerosis, Biosynthesis of cofactors, Viral protein interaction with cytokine and cytokine receptor, Amoebiasis, NF-kappa B signaling pathway, Insulin resistance in colorectal cancer. Upregulated DEGs were found to be associated with pathways involved in cell signaling, immune response, lipid metabolism, and cellular adhesion, which could indicate alterations in these processes in the analyzed dataset.

Module screening from the PPI network: The protein-protein interaction (PPI) network was constructed using STRING database and Cytoscape software was used in visualizing and analyzing complex networks. A total of 137 nodes and 500 edges were analyzed, was then utilized to identify modules or clusters within the PPI network. The Molecular Complex Detection (MCODE) plugin in Cytoscape was used for this purpose. Top Sub-Networks: The PPI sub-networks, representing modules of interacting proteins, were analyzed. In this case, the top three sub-networks were selected for further investigation (Table-1). There were 12 genes involved in sub-network-1, 11 genes involved in sub-network-2 and 7 genes involved in sub-network-3. GO and KEGG pathway analyses were conducted separately for genes within the top-scored sub-networks.

Hub Gene Identification: Hub genes are highly connected genes within a network and often play crucial roles in cellular processes. In this study, the top 9 hub genes were identified using STRING software, which integrates information from various sources to predict protein-protein interactions and prioritize hub genes based on their connectivity within the network, the top hub genes were CXCL3, CXCL2, CXCL1, COL1A2, COL1A1, PLAU, MMP1, MMP3 and AGT. All the hub genes were found to be upregulated.

Identification of prognostic biomarkers: CXCL2 and CXCL3 were the top two genes identified as prognostic biomarkers. These genes were related to colorectal cancer survival (Figure-14 and 15). Additionally, CXCL1, COL1A1, COL1A2 and PLAU genes were identified to be affected by the sex.

Table-1: Top 3 primary modules of PPI sub-network by plug-in MCODE in Cytoscape software.

Sub-network	Nodes	Edges	Score
	12	45	8.182
	11	36	7.200
	7	15	5.000

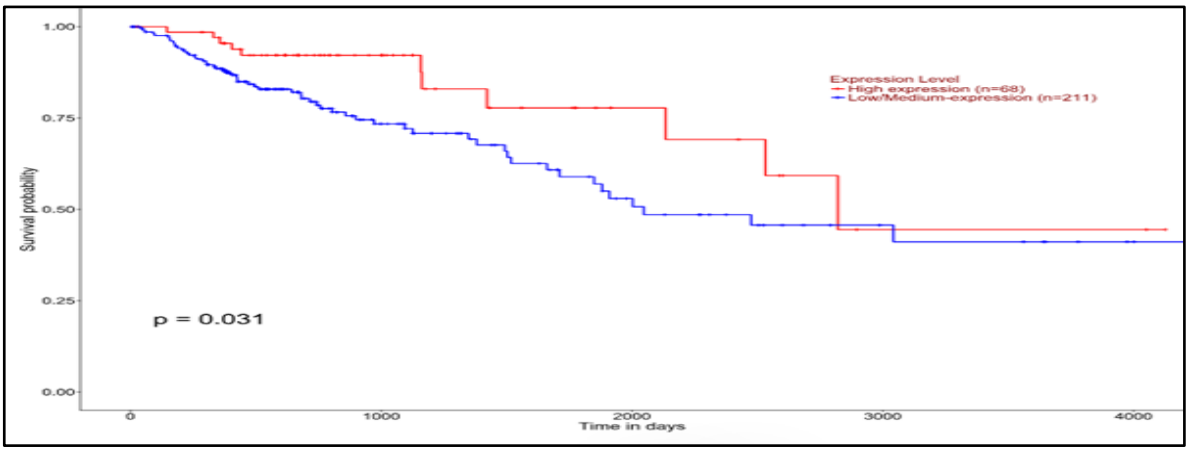


Figure-14: Effect of CXCL2 expression level on CRC survival.

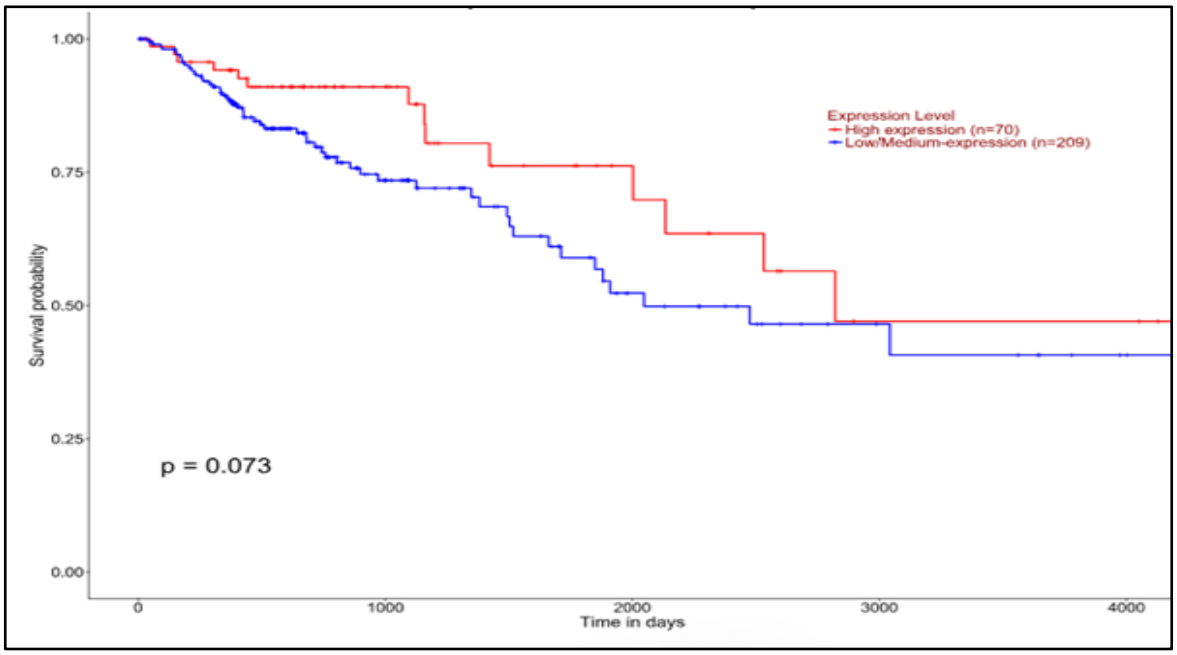


Figure-15: Effect of CXCL3 expression level on CRC survival.

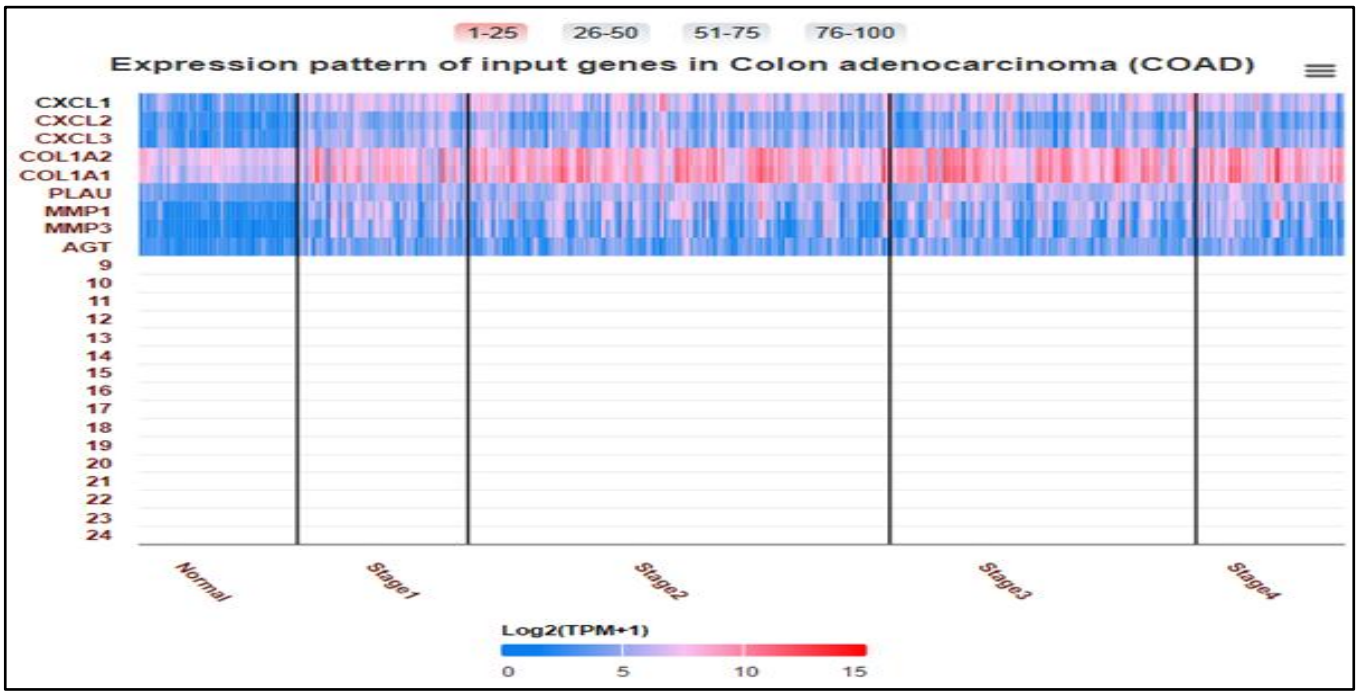
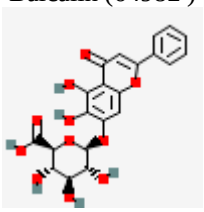
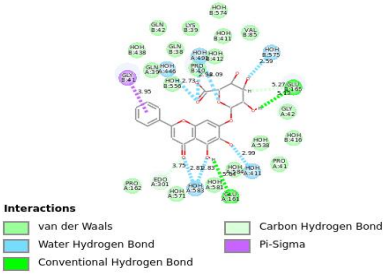
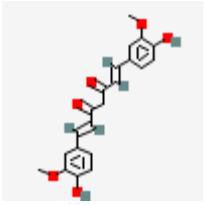
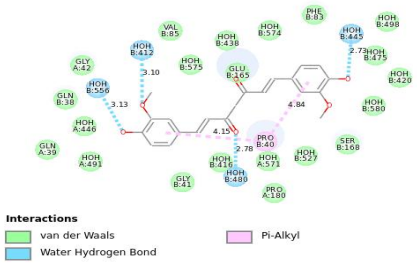
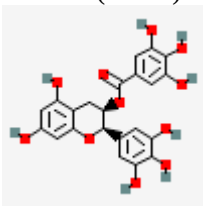
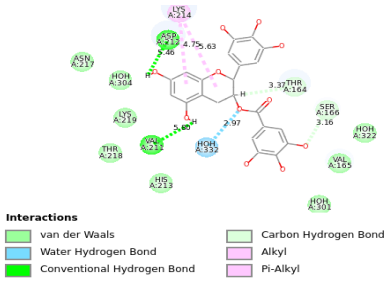
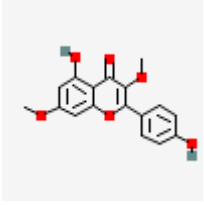
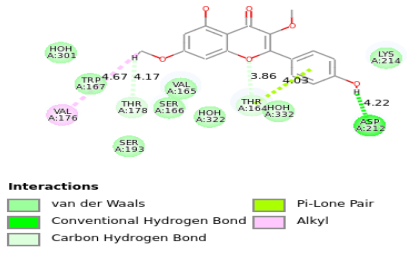
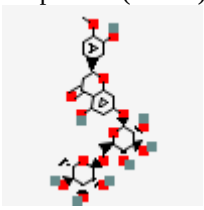
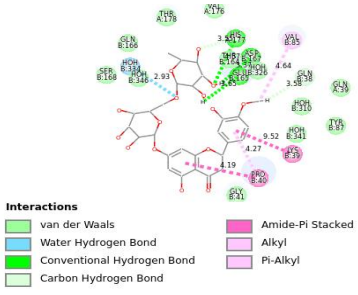
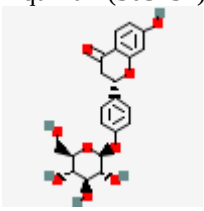
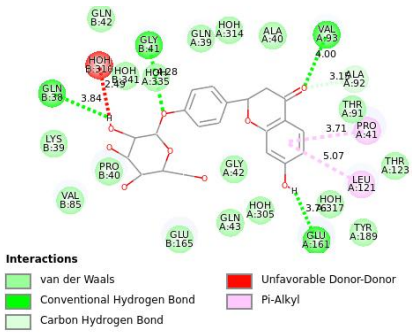
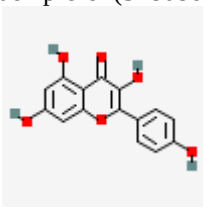
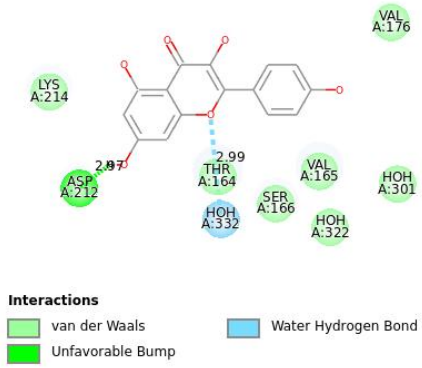
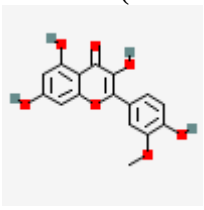
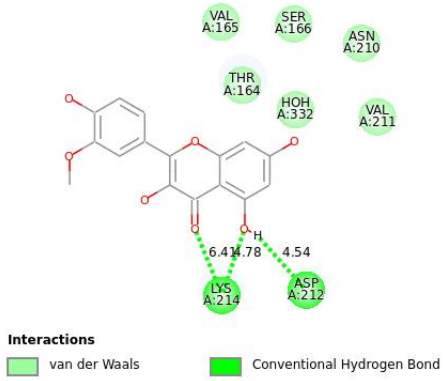
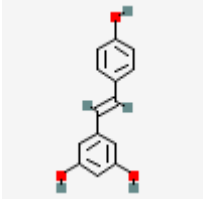
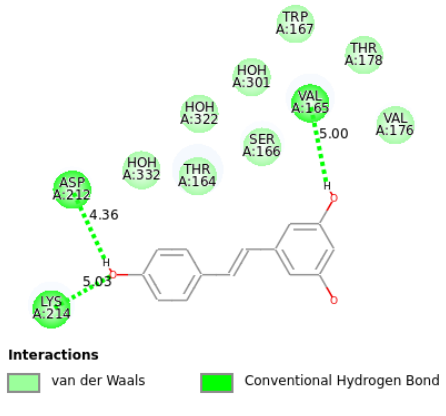


Figure-16: The expression pattern of identified genes in colon adenocarcinoma.

Molecular Docking: In accordance with this study, the CRC target genes CXCL2 (PDB id: 6wzj) AND CXCL3 (PDB id: 6wzk) were included. The PDB database contained the protein crystal structures that corresponded to the targets. The Grid box size and location of the protein binding pockets were predicted using LibDock. Following the loading of target proteins and bioactive compounds into Discovery Studio, the LibDock score was used to determine the affinity value of the optimal binding postures (Table-2). Target protein affinity showed an average value of 106.820kcal/mol. Therefore, the therapeutic benefit of SJZD for CRC treatment has been confirmed at the molecular docking level. The top 5 affinity absolute value complexes were Baicalin, Curcumin, (-)-Epigallocatechingallate, Jaranol, Hesperidin, Liquiritin, Kaempferol, Isorhamnetin, and Resveratrol.

Table-2: Docking scores.

Compounds having protein affinity (Kcal/mol)	Target protein	2D structure of the binding affinity of compounds with the protein complex	LibDockscore of target
Baicalin (64982) 	C-X-C motif chemokine ligand 2(CXCL2)		104.745
Curcumin (969516) 	C-X-C motif chemokine ligand 2(CXCL2)		88.2583
(-)-Epigallocatechin gallate; EGCG(65064) 	C-X-C motif chemokine ligand 2(CXCL2)		131.228
Kumatakenin;(5318869) Jaranol 	C-X-C motif chemokine ligand 2(CXCL2)		122.068
Hesperidin (10621) 	C-X-C motif chemokine ligand 2(CXCL2)		106.068

Liquiritin (503737) 	C-X-C motif chemokine ligand 3(CXCL3)		99.3404
Kaempferol (5280863) 	C-X-C motif chemokine ligand 3(CXCL3)		111.722
Isorhamnetin (5281654) 	C-X-C motif chemokine ligand 3(CXCL3)		81.8974
Resveratrol (445154) 	C-X-C motif chemokine ligand 3(CXCL3)		116.063

EGCG has shown better interaction affinity for the CXCL2 analog and formed the hydrogen bond with valine165 (distance= 2.97) and Van der Waals interaction with valine165, serine166 and threonine164 amino acids. EGCG is a catechin found in green tea known for its antioxidant and anticancer properties. It has been studied for its potential benefits in cancer prevention and treatment due to its ability to modulate various signaling pathways.

Kaempferol has shown the better interaction affinity for the CXCL3 analog and formed the hydrogen bond with threonine164 (distance=2.99) and Van der Waals interaction with valine165, serine166 and lysine214 amino acids. Kaempferol is a flavonoid present in various plants and fruits with antioxidant, anti-inflammatory, and anticancer properties. It has been investigated for its potential role in cancer prevention and treatment due to its effects on signaling pathways.

Conclusion

The comprehensive analysis conducted in this study aimed to explore the gene expression profiles associated with colorectal cancer (CRC) and identify potential prognostic biomarkers. Through differential gene expression analysis, a total of 309 upregulated genes and 234 down regulated genes were identified in CRC samples compared to normal samples. The GO term enrichment analysis revealed significant biological processes, molecular functions, and cellular components that are dysregulated in CRC. Specifically, genes involved in collagen breakdown, extracellular matrix organization, and response to various stimuli were found to be significantly enriched among the differentially expressed genes. The top Hub genes CXCL3, CXCL2, CXCL1, COL1A2, COL1A1, PLAU, MMP1, MMP3 and AGT were identified. These findings suggest alterations in key cellular processes and signaling pathways in CRC pathogenesis.

Moreover, the KEGG pathway analysis highlighted important pathways associated with CRC development, including cell adhesion molecules, calcium signaling pathways, and lipid metabolism. The upregulated DEGs were primarily associated with cell signaling, immune response, and cellular adhesion pathways, indicating potential dysregulation in essential biological mechanisms contributing to CRC progression and the prognostic biomarkers CXCL2 and CXCL3 were identified. The molecular docking analysis identified potential therapeutic compounds for CRC treatment, with compounds such as EGCG, Jaranol and Kaempferol showing promising affinity for target proteins associated with CRC. These findings suggest the potential utility of these compounds as candidates for further investigation in CRC treatment strategies.

Overall, this study contributes to the understanding of CRC pathogenesis and offers potential avenues for personalized medicine and targeted therapy in CRC management.

Future research should focus on experimentally validating the functional roles of identified prognostic biomarkers in CRC pathogenesis and disease progression. Functional studies can provide crucial mechanistic insights into how these biomarkers influence tumor biology and impact patient outcomes. Developing biomarker panels that encompass a range of molecular signatures, including gene expression profiles, protein levels, and genetic mutations, can improve the accuracy of prognostic assessments and enable personalized treatment strategies for CRC patients[16].

References

1. Lynch, H. T., & de la Chapelle, A. (2003). Hereditary colorectal cancer. *The New England journal of medicine*, 348(10), 919–932.

2. Johns, L. E., & Houlston, R. S. (2001). A systematic review and meta-analysis of familial colorectal cancer risk. *The American journal of gastroenterology*, 96(10), 2992–3003.

3. Eaden, J. A., Abrams, K. R., & Mayberry, J. F. (2001). The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut*, 48(4), 526–535.

4. Siegel, R., Naishadham, D., & Jemal, A. (2013). Cancer statistics. *CA: a cancer journal for clinicians*, 63(1), 11–30.

5. Mármol, I., Sánchez-de-Diego, C., Pradilla Dieste, A., Cerrada, E., & Rodriguez Yoldi, M. J. (2017). Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. *International journal of molecular sciences*, 18(1), 197.

6. Boland, C. R., & Lynch, H. T. (2013). The history of Lynch syndrome. *Familial cancer*. 12(2), 145–157. https://doi.org/10.1007/s10689-013-9637-8

7. Bertucci, F., Salas, S., Eysteries, S., Nasser, V., Finetti, P., Ginestier, C., Charafe-Jauffret, E., Loriod, B., Bachelart, L., Montfort, J., Victorero, G., Viret, F., Ollendorff, V., Fert, V., Giovannini, M., Delpero, J. R., Nguyen, C., Viens, P., Monges, G., Birnbaum, D., … Houlgatte, R. (2004). Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. *Oncogene*, *23*(7), 1377–1391. https://doi.org/10.1038/sj.onc.1207262

8. Lech, G., Slotwinski, R., & Krasnodebski, I. W. (2014). The role of tumor markers and biomarkers in colorectal cancer. *Neoplasma*, 61(1), 1–8.

9. Diboun, I., Wernisch, L., Orengo, C. A., & Koltzenburg, M. (2006). Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. *BMC genomics*, 7, 252. https://doi.org/10.1186/1471-2164-7-252

10. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C.,

Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*, 25(1), 25–29.

11. Gene Ontology Consortium (2006). The Gene Ontology (GO) project in 2006. *Nucleic acids research*, 34(Database issue), D322–D326. https://doi.org/10.1093/nar/gkj021

12. Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., & Jensen, L. J. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic acids research*, 41(Database issue), D808–D815. https://doi.org/10.1093/nar/gks1094

13. Chandrashekar, D. S., Bashel, B., Balasubramanya, S. A. H., Creighton, C. J., Ponce-Rodriguez, I., Chakravarthi, B. V. S. K., & Varambally, S. (2017). UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia (New York, N.Y.)*, 19(8), 649–658. https://doi.org/10.1016/j.neo.2017.05.002

14. Liu, X., Bing, Z., Wu, J., Zhang, J., Zhou, W., Ni, M., Meng, Z., Liu, S., Tian, J., Zhang, X., Li, Y., Jia, S., & Guo, S. (2020). Integrative Gene Expression Profiling Analysis to Investigate Potential Prognostic Biomarkers for Colorectal Cancer. *Medical science monitor: international medical journal of experimental and clinical research*, 26, e918906. https://doi.org/10.12659/MSM.918906

15. Shang, L., Wang, Y., Li, J., Zhou, F., Xiao, K., Liu, Y., Zhang, M., Wang, S., & Yang, S. (2023). Mechanism of Sijunzi Decoction in the treatment of colorectal cancer based on network pharmacology and experimental validation. *Journal of ethnopharmacology*, 302(Pt A), 115876. https://doi.org/10.1016/j.jep.2022.115876

16. Van Cutsem, E., Cervantes, A., Nordlinger, B., Arnold, D., & ESMO Guidelines Working Group (2014). Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of oncology: official journal of the European Society for Medical Oncology*, 25 Suppl 3, iii1–iii9. https://doi.org/10.1093/annonc/mdu260