American Journal of Biology and Natural Sciences
https://biojournals.us/index.php/AJBNS
ISSN: 2997-7185

Exploratory Analysis of Gene Expression Patterns in Chemotherapy-Treated Colorectal Cancer Patients

Mustafa Ayad Farooq
Department of Biology, College of Science, University of Baghdad, Baghdad, Iraq

Hind Mahmood Jumaah, Israa Ayoub Alwan
Department of Biotechnology, College of Science, University of Baghdad, Baghdad, Iraq

Received: 2025, 15, Sep
Accepted: 2025, 21, Oct
Published: 2025, 30, Nov

Annotation: CRC remains to be one of the major causes of death by cancer all over the world and metastatic CRC poses a challenge in both treatment and prognosis because of heterogeneity of the tumors and resistance to drugs. The authors investigated the profile of gene expression differences in 21 patients who received chemotherapy with 5-fluorouracil, oxaliplatin, and irinotecan, with CRC. By means of principal component analysis and hierarchical clustering, it was possible to find out specific molecular subgroups, which unveiled the most important genes that exhibited high expression variations. These results give an understanding of the molecular dynamics of the development of tumors and response to treatment in CRC. The identified gene signatures can be considered as prospective prognosis biomarkers and targets of personal treatment. To continue this line of study, future-oriented research ought to concentrate on functional confirmation of the individual genes and their harmonization with the multi-omics data to

improve precision medicine in CRC.

Keywords: colorectal cancer (CRC), gene expression, chemotherapy, 5-fluorouracil, oxaliplatin, irinotecan.

INTRODUCTION

Certain categories of cancer are in second place in causing death all over the world after cardiovascular diseases, colon cancer is in the third place with respect to occurrence and second with respect to mortality [1]. whereas, rectal cancer constitutes more than 35 percent of the burden of colorectal cancer, which in turn is a unique category of cancers and has stood in second place as relationships with respect to the origin of the cancers in the colon positively [2]. The colorectal cancer stem cells (CCSCs) are not only known to cause tumor growth, recurrence and treatment resistance but as time goes on they are proving unique to colorectal cancer [3]. In April 2024, the World Health Organization/International Agency for Research on Cancer (IARC) issued their Global Cancer Statistics 2022 as CA: Cancer Journal for Clinical. This is a report about the incidence rates and mortality rates of 36 categories of cancers in 185 countries and a territory in the world, and it is analyzed by gender, through geographic regions and by level of Human Development Index (HDI). In the year 2022, around the globe, 9.74 million people are estimated to have died and 19.96 million people are estimated to have a new death case of cancer and of these 9.3 percent were caused by colorectal cancer (903,859 cases) [4]. Notwithstanding all research and the increase in knowledge regarding colorectal cancer (CRC), its metastatic variant has been one of the most significant problems because of the lack of curing measures and improper prognosis [5]. Metastatic colorectal cancer is essentially treated by means of chemotherapy therapy. The 5-fluorouracil, oxaliplatin and irinotecan 3 are reported to be the three major drugs known to treat metastatic colorectal cancer with irinotecan further being metabolized into active compound SN-38[6]. The analysis of gene expression is important to learn how genes regulate the cell activities and condition the health and sickness. It facilitates new developments in personalized medicine, diagnosis of disease, and bio-technology [7]. This work tries to apply the analysis of differential gene expression in efforts to know some of the genetic changes that occur in the cases of colorectal cancer, especially among the patients who receive the most common types of chemotherapy which include 5-FU, oxaliplatin, and irinotecan. Through the potential identification of the genes that exhibit a significant expression variation between diseased and control samples, the research will reveal the actual molecular causes of the disease and enhance the insights relating to how patients respond to treatment and contribute to an additional development of more specific and effective treatment methods. Hence, this paper attempts to perform profiling of the differentially expressed genes in the colorectal cancer patients receiving standard chemotherapies to infer any molecular processes that drive response to chemotherapeutic treatments and interpatient tumor heterogeneity, which can eventually led to progress in the development of precision medicine.

Literature Review

1. Epidemiology and Clinical Challenges of Colorectal Cancer

Colon and/or cancer of the rectum, collectively termed as colorectal cancer (CRC) is a serious health concern given that it has been listed as the third most reported and second most causes of cancer deaths in all parts of the world [8]. Approximately, 9.4 percent of cancer-related deaths resulted because of CRC in 2020 [9]. Nevertheless, taking into consideration the vast rise in cases detected in the elderly generation, it is believed that the international prevalence of CRC will be greater than twofold by 2035, and the highest rise will happen in the less developed countries [10]. High incidence rates are reported in countries with high values on human development index

(HDI), especially due to to dietary habits and lifestyle issues red meat and processed meat and sodas consumption and obesity [11]. Due to the arrival of the age of personalized medicine in the last two decades, precision medicine attempts to detect the molecular changes that might be responsive to precision treatment [12]. Although remarkable progress is observed in the terms of early detection, surgical treatment, and systemic therapy, CRC represents a prominent clinical issue given its heterogeneity and the complexity of the molecular pathways, the survival rates of CRC have been increasing in the last several decades, yet the disease is still a crucial health issue [13]. But recent developments in the fields of molecular biology and precision medicine start to change the reality of treatment of colorectal cancer and provide more hope to achieve better possibilities of more personalized treatment and more effective therapy [14].

Figure 1: Global impact of colorectal cancer. This figure was created by the author based on the reviewed literature.

2. Chemotherapy Treatments for Metastatic Colorectal Cancer

Chemotherapy means using chemicals to inhibit malignant cells or the infectious agents of a disease, such as microorganisms, without much affecting the host cells. Therefore, the treatment can be broadly divided into cancer chemotherapy and antimicrobial chemotherapy. The drugs belong to these categories, and are different from the others as they are generally intended to kill or inhibit the cell [15]. The first-line therapy of mCRC usually involves chemotherapy regimen using fluoropyrimidine and either oxaliplatin and/or irinotecan or without vascular endothelial growth factor receptors (VEGFR)- or epidermal growth factor receptors (EGFR)-targeting therapies [16]. Nevertheless, the great majority of patients will progress after initial treatment. The existence of certain biological determinants in a particular patient, such as the mutational status of rat sarcoma proto-oncogene (RAS) or defective mismatch repair (deficient mismatch repair/microsatellite instability mutational status) could also affect treatment plan [17]. The US FDA just approved fruquintinib (8 November 2023) in adult patients with mCRC and who have received at least one prior line of fluoropyrimidine-, oxaliplatin-, and irinotecan- based chemotherapy, an anti-VEGF therapy, and, where RAS wild-type and medically appropriate, an anti-EGFR therapy. The single most frequent grade 3/4 AE was hypertension, whereas most AEs (greater than 20%) were hypertension, palmar-plantar erythrodysesthesia, proteinuria, dysphonia, abdominal pain, diarrhea, and asthenia [18].

3. The Role of Gene Expression Profiling in Colorectal Cancer

Evidence is emerging that genes do not work alone but are part of huge and complex information-processing networks of interacting genes, proteins and small molecules. They are self-organizing networks in humans into functional specialized modules which associate with each other to learn to turn each other on and off to meet the changing external and internal circumstances. These interactions of reciprocal feedbacks correlate information conclude and hence facilitate healthy growth and longevity in a person [19]. Human genome contains approximately 63,000 genes all made of DNA, and of them 20,000 are transcribed into RNAs that encode proteins, the other 43,000 genes are transcribed into non-protein-coding RNAs (ncRNAs) which have many functions that are regulatory [20].

Data Sours and Result

Gene expression data were obtained from the publicly available GEO database (accession number GSE251958), comprising RNA sequencing profiles from 21 colorectal cancer patients treated with chemotherapy [21]. Principal component analysis (PCA) was performed to reduce dimensionality and visualize sample clustering. Hierarchical clustering based on the top 500 most variable genes was used to identify molecular subgroups. Differential gene expression analysis was conducted using t-tests between identified clusters, with p-values adjusted for multiple testing using the False Discovery Rate (FDR) method to ensure statistical robustness.

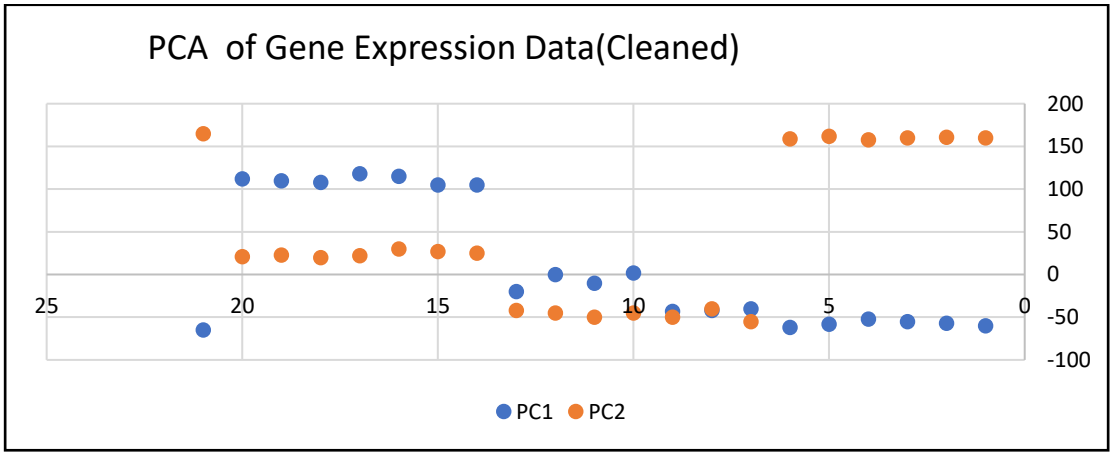

Figure 2: PCA Plot of Cleaned Gene Expression Data.

Principal Component Analysis (PCA) was performed on the gene expression data from 21 colorectal cancer samples to reduce the high-dimensional data into two main components for visualization and interpretation. The first principal component (PC1) explains 25.10% of the total variance in the dataset. The second principal component (PC2) explains 21.70% of the variance. Each point in the plot represents an individual sample labeled accordingly (Sample 1, Sample 2, ..., Sample 21). The PCA plot reveals the underlying structure of the data, showing how samples cluster or separate based on their gene expression profiles. Samples clustering closely together indicate similar gene expression patterns. Samples that are separated relating to perceived difference in the expression of genes indicate the existence of a substantial change and this could be either biological or clinic varied. The results of this exploratory analysis set a foundation to further research on group difference and identification of other important genes behind collated variance.

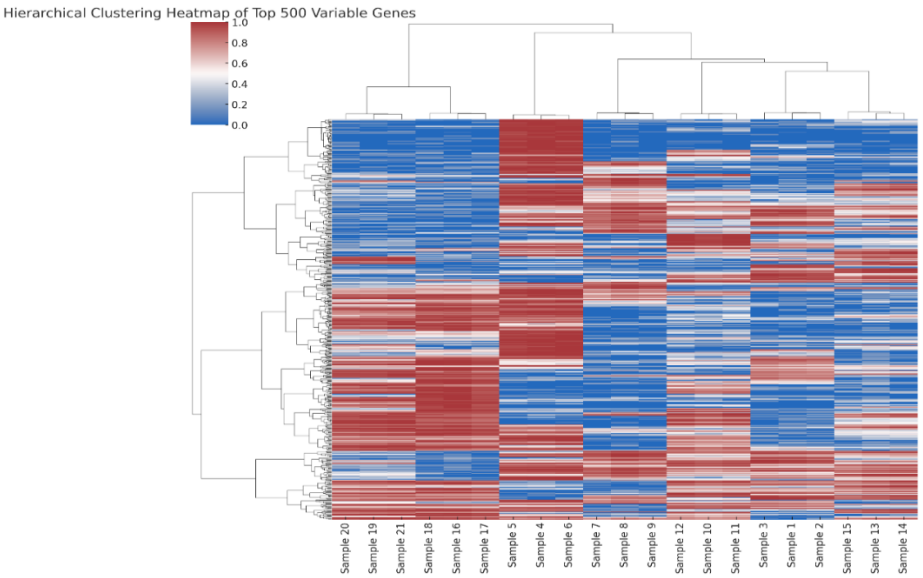

Figure 3: Hierarchical Clustering Heatmap of Top 500 Variable Genes

The expression profiles of the 21 colorectal cancer samples were analyzed using the top 500 most variable genes, in terms of a hierarchical clustering analysis. In this type of analysis, expression patterns of samples and genes are determined and grouped according to their similarity. Heat map columns indicate individual sample, groups of which are stratified by the similarity of their gene's expression. Rows correspond to the various genes that were taken as the ones with the greatest variance across the samples; the clusters are determined by the expression profiles. The color tool of the heat map shows the levels of expression with red as high level and blue as low level. Sample clusters indicate the existence of similar molecules profiles, meaning that they might be related to different biological conditions or clinical diagnoses. On the same note, networks of genes can be biologically similar in functions or pathways. This discussion will provide a glimpse on the nature of the underlying structure of the information and enhance subsequent research on important genes, and sample groups linked to colorectal cancer.

Differential expression analysis was also done to select genes that show a great extent of differences in expression levels of at least two groups of the colorectal cancer samples which were grouped using the hierarchical clustering method. Each gene gave a t-test of expression contrasting Cluster 1 (Samples 16-21) against Cluster 2 (Samples 1-15). The multiplicity of hypothesis testing was corrected by using the False Discovery Rate (FDR) procedure. The results also demonstrated that a number of genes exhibited statistically significant difference in expression (adjusted p-value < 0.05). The table below is a summary of the top 10 most significant genes based on the adjusted p-values:

Table 1: Top 10 Differentially Expressed Genes Between Colorectal Cancer Sample Clusters with Statistical Significance

Gene ID	t-statistic	Raw p-value	Adjusted p-value	Significant
58608	38.14	6.73×10^{-19}	1.75×10^{-14}	Yes
31596	-27.57	8.68×10^{-17}	1.13×10^{-12}	Yes
20058	26.17	2.76×10^{-16}	2.40×10^{-12}	Yes
14910	36.53	4.15×10^{-16}	2.70×10^{-12}	Yes
23446	-22.66	3.32×10^{-15}	1.73×10^{-11}	Yes
20297	23.56	4.26×10^{-15}	1.85×10^{-11}	Yes
25918	22.22	6.92×10^{-15}	2.57×10^{-11}	Yes
52595	-20.22	3.21×10^{-14}	1.04×10^{-10}	Yes
46902	19.59	4.78×10^{-14}	1.38×10^{-10}	Yes
51060	-20.18	5.70×10^{-14}	1.49×10^{-10}	Yes

CONCLUSION

This analysis using differential gene expression was able to identify a sub-population of genes that exhibit considerable differences between the two sample groups based on the colorectal cancer data. These genes can most likely be of critical importance in the underlying molecular determinants of tumor heterogeneity and should be considered possible prognostic biomarkers or

therapeutic targets. To further ascertain their biological importance and usefulness, they have to be functional validated and have been integrated with corresponding clinical data. Further research using bigger groups of patients with an array of omics strategies will add deeper insight and refine diagnosis and treatment of colorectal cancer.

FUTURE DIRECTIONS

In continuation of the present work, pending studies ought to be directed towards ascertaining biological activities of the revealed differentially expressed genes in experimental assays. It will be possible to increase the sample size and include various patient groups to increase the generalizability of the results. Furthermore, the inclusion of multi-omics data like proteomics, epigenomics, and metabolomics has the potential to give a more detailed picture of molecular processes behind colorectal cancer development and response to therapy. Lastly, clinical trials testing of targeted therapy based on these gene signatures could make a substantial difference in personalized approaches to treatment and clinical outcome.

Reference

1. Yang Z, Cespedes-Acuña CL, Yang Z, Abu Bakar MZ, Chan KW, Deng X. Plant foods and their bioactives as dietary enhancers for colon cancer treatment with 5-fluorouracil. Food Rev Int. 2025:1–50. http://dx.doi.org/10.1080/87559129.2024.2446366
2. Coelho D, Estêvão D, Oliveira MJ, Sarmento B. Radioresistance in rectal cancer: can nanoparticles turn the tide? Mol Cancer. 2025;24(1):35. https://doi.org/10.1186/s12943-025-02232-x
3. Omran MM, Fouda MS, Mekkawy SA, Tabll AA, Abdelaziz AG, Omran AM, Emran TM. Molecular biomarkers and signaling pathways of cancer stem cells in colorectal cancer. Technol Cancer Res Treat. 2024;23:15330338241254061. https://doi.org/10.1177/15330338241254061
4. Zhang X, Yang L, Liu S, Cao LL, Wang N, Li HC, Ji JF. Interpretation on the report of global cancer statistics 2022. Zhonghua Zhong Liu Za Zhi. 2024;46(7):710–721. https://doi.org/10.3760/cma.j.cn112152-20240416-00152
5. Cañellas-Socias A, Sancho E, Batlle E. Mechanisms of metastatic colorectal cancer. Nat Rev Gastroenterol Hepatol. 2024;21(9):609–625. https://doi.org/10.1038/s41575-024-00934-z
6. Razumovskaya A, Silkina M, Poloznikov A, Kulagin T, Raigorodskaya M, Gorban N, et al. Predicting patient outcomes with gene-expression biomarkers from colorectal cancer organoids and cell lines. Front Mol Biosci. 2025;12:1531175. https://doi.org/10.3389/fmolb.2025.1531175
7. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular biology of the cell. 6th ed. New York: Garland Science; 2015.
8. World Health Organization (WHO). Cancer [Internet]. Geneva: World Health Organization; [cited 2021 Jul 14]. Available from: https://www.who.int/news-room/fact-sheets/detail/cancer
9. Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F. Global Cancer Observatory: Cancer Today (Version 1.1) [Internet]. Lyon, France: International Agency for Research on Cancer; 2024 [cited 2025 Jul 26]. https://gco.iarc.who.int/today
10. Hossain MS, Karuniawati H, Jairoun AA, Urbi Z, Ooi J, John A, Lim YC, Kibria KMK, Mohiuddin AKM, Ming LC, Goh KW, Hadi MA. Colorectal cancer: A review of carcinogenesis, global epidemiology, current challenges, risk factors, preventive and treatment strategies. Cancers. 2022;14(7):1732. https://doi.org/10.3390/cancers14071732

11. Matsuda T, Fujimoto A, Igarashi Y. Colorectal cancer: Epidemiology, risk factors, and public health strategies. Digestion. 2025;106(2):91–99. https://doi.org/10.1159/000543921
12. Pizzamiglio M, Soulabaille A, Lahlou W, Pilla L, Zaanan A, Taieb J. Advances and challenges in targeted therapies for HER2-amplified colorectal cancer. Eur J Cancer. 2025;222:115471. https://doi.org/10.1016/j.ejca.2025.115471
13. Delle Cave D. Advances in molecular mechanisms and therapeutic strategies in colorectal cancer: A new era of precision medicine. Int J Mol Sci. 2025;26(1):346. https://doi.org/10.3390/ijms26010346
14. Xie YH, Chen YX, Fang JY. Comprehensive review of targeted therapy for colorectal cancer. Signal Transduct Target Ther. 2020;5(1):22. https://doi.org/10.1038/s41392-020-0116-z
15. Alam A, Farooq U, Singh R, Dubey VP, Kumar S, Kumari R, et al. Chemotherapy treatment and strategy schemes: A review. Open Access J Toxicol. 2018;2(5):555600. http://dx.doi.org/10.19080/OAJT.2018.02.555600
16. Paly VF, Dasari A, Hubbard J, Bekaii-Saab T, Padukkavidana T, Hernandez L. Adverse event costs of systemic therapies for metastatic colorectal cancer previously treated with fluoropyrimidine-, oxaliplatin- and irinotecan-based chemotherapy and biologics in the US. J Comp Eff Res. 2024;13(8):e240084. https://doi.org/10.57264/cer-2024-0084
17. Grothey A, Marshall JL, Bekaii-Saab T. Sequencing beyond the second-line setting in metastatic colorectal cancer. Clin Adv Hematol Oncol. 2019;7(3):1–19. https://pubmed.ncbi.nlm.nih.gov/31730588/
18. FRUZAQLA (fruquintinib) [Package insert]. [Takeda Pharmaceuticals America, Inc, Lexington, MA (2023). https://www.accessdata.fda.gov/drugsatfda_docs/label/2023/217564s000lbl.pdf
19. Del Val C, Díaz de la Guardia-Bolívar E, Zwir I, Mishra PP, Mesa A, Salas R, et al. Gene expression networks regulated by human personality. Mol Psychiatry. 2024;29(7):2241–2260. https://doi.org/10.1038/s41380-024-02484-x
20. Rashid OF, Mohammed MT, Tuama SA. RNA encoding and Raita algorithm for misuse intrusion detection method using KDD-Cup99 dataset. In: Proceedings of the 2024 International Jordanian Cybersecurity Conference (IJCC 2024). 2024. pp. 1–6. http://dx.doi.org/10.1109/IJCC64742.2024.10847287
21. Zwir I, Del-Val C, Hintsanen M, Cloninger KM, Romero-Zaliz R, Mesa A, et al. Evolution of genetic networks for human creativity. Mol Psychiatry. 2022;27(1):354–376. https://doi.org/10.1038/s41380-021-01097-y
22. Rashid OF, Subhi MA, Hussein MK, Mahdi MN. Enhancing internet data security: A fusion of cryptography, steganography, and machine learning techniques. Baghdad Sci J. 2025;22(3):1005–15. https://doi.org/10.21123/bsj
23. National Center for Biotechnology Information. GSE251958: RNA-seq of colorectal cancer samples. Gene Expression Omnibus; 2024. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE251958