







## Computers in Biology and Medicine

Available online 11 February 2023, 106639

In Press, Journal Pre-proof [? What's this?](#)

---

# Integrated analysis of multi-omics data for the discovery of biomarkers and therapeutic targets for colorectal cancer

[Nima Zafari](#)<sup>a</sup>, [Parsa Bathaei](#)<sup>a</sup>, [Mahla Velayati](#)<sup>a</sup>, [Fatemeh Khojasteh-Leylakoochi](#)<sup>a b c</sup>,  
[Majid Khazaei](#)<sup>a b</sup>, [Hamid Fiuji](#)<sup>a</sup>, [Mohammadreza Nassiri](#)<sup>d</sup>, [Seyed Mahdi Hassanian](#)<sup>a b</sup>,  
[Gordon A. Ferns](#)<sup>e</sup>, [Elham Nazari](#)<sup>a b</sup>  , [Amir Avan](#)<sup>a b c</sup>  

[Show more](#) 

 Share  Cite

---

<https://doi.org/10.1016/j.compbimed.2023.106639>

[Get rights and content](#)

---

### Abstract

The considerable burden of colorectal cancer and the rising trend in young adults emphasize the necessity of understanding its underlying mechanisms, providing new diagnostic and prognostic markers, and improving therapeutic approaches. Precision medicine is a new trend all over the world and identification of novel biomarkers and therapeutic targets is a step forward towards this trend. In this context, multi-omics data and integrated analysis are being investigated to develop personalized medicine in the management of colorectal cancer. Given the large amount of data from multi-omics approach, data integration and analysis is a great challenge. In this Review, we summarize how statistical and machine learning techniques are applied to analyze multi-omics data and how it contributes to the discovery of useful diagnostic and prognostic biomarkers and therapeutic targets. Moreover, we discuss the importance of these biomarkers and therapeutic targets in the clinical management of colorectal cancer in the future. Taken together, integrated analysis of multi-omics data has great potential for finding novel diagnostic and prognostic biomarkers and therapeutic targets, however, there are still challenges to overcome in future studies.

## Introduction

Colorectal cancer (CRC) caused about 0.94 million deaths in 2020 in the world and ranked as the second most common cause of mortality due to cancer. CRC also caused approximately 1.93 million incidence cases in 2020 worldwide [1]. Additionally, its incidence is increasing and currently includes 11% of total cancer cases. Increasing trend in CRC incidence and mortality has occurred predominantly in developed countries [2]. Nowadays, colonoscopy is known as the gold standard screening method for CRC, however, it is an expensive and semi-invasive method. Therefore, poor patient adherence to colonoscopy remains a major challenge in this field. On the other hand, fecal blood tests with a suboptimal accuracy, has led to the late diagnosis of CRC [3]. Tumor Node Metastasis (TNM) staging is now the foundation of prognostication in CRC. Although biological, genetic, and other molecular and clinical information affect the prognosis of patients, a more powerful tool is needed for personalized prognostication of CRC [4].

A better understanding of the molecular pathophysiology of CRC in recent years has contributed to the identification of novel biomarkers for screening, diagnosis, and prognosis of CRC. Unfortunately, these biomarkers are often disappointing in outcomes [3]. However, advances in high-throughput techniques have provided new opportunities to study CRC at different molecular levels and to improve our knowledge about CRC that finally resulted in generation of significant amounts of data. As we know, CRC formation is very complex and thousands of molecules alter during this process. Among these molecules, some might be valuable markers of cancer diagnosis and prognosis. These dysregulated molecules could serve as potential targets that help scientists develop novel targeted drugs for the treatment of CRC. The availability of multi-omics data consisting of proteins (proteomics), genes (genomics), RNAs (transcriptomics), and metabolites (metabolomics) necessitate the development of tools to facilitate the analysis and interpretation of these generated data. Independent analysis of omics-data from each molecular level can be performed using many statistical methods. However, such individual analysis omits the relationship between different molecular levels, which makes it necessary to better identify the pathological processes behind CRC formation in a biologically meaningful context based on integrated analysis of data from various omics studies. Accordingly, integrated analysis of multi-omics data which is called integromics can provide useful insight into CRC pathogenesis and help researchers to discover novel therapeutic targets, and diagnostic and prognostic biomarkers. Integromics reveals interactions between different molecular levels, thus biomarker candidates which have been found based on integromics would be more biologically relevant. In this review, we describe how integrated analysis of multi-omics data led to the discovery of promising CRC biomarkers.

---

## Section snippets

### Online resources of omics data and tools for CRC research

Online resources of omics data are important parts of cancer research and help scientists to identify cellular processes involved in cancer pathogenesis. Additionally, these online resources facilitate the systematic integration of different datasets. Online resources and tools which has been mentioned in our paper are provided in Table 1. These online tools and databases are aimed to improve the data analysis methods and allow researchers to obtain more accurate and more valuable results from...

### Statistical analysis, machine learning models, and validation methods

Statistical analyses encompass the organization and description of data, the discovery of the correlations and interactions between factors, and the interpretation of acquired results [7]. Traditional statistical data analysis includes converting data into analytical variables and selecting an appropriate statistical test such as t-tests, analysis of variance (ANOVA), and Pearson and Spearman correlations. t-tests are used for comparing differences between two groups while the ANOVA test is...

### Biomarker discovery based on integrated analysis and statistical methods

Omics data can be integrated leading to the identification of new biomarkers in different levels of genomics, epigenomics, transcriptomics, and proteomics. This integration can be performed by numerous statistical methods. The focus of this section is to review the statistical-based integrated analysis of multi-omics data for the discovery of CRC biomarkers (Fig. 1)....

### Machine learning analysis: a step forward in analyzing multi-omics data

In recent years, the advancement of analysis technologies led to an increasing amount of data obtained by omics analysis. In addition, developing modern high-throughput omic platforms contributed to the necessity of undertaking an integrative approach for analyzing these data [61,62]. On the other hand, given that precision medicine is a new trend all over the world, oncology is actively adopting this approach. In this context, the field of oncology is focusing on molecular information to...

## Conclusions

Recent advances in technology have contributed to the development of omics data which enables us to gather data from different molecular levels and integrate them. Combining multiple experimental approaches and integrating the results make it possible to generalize colorectal cancer's complexity from experimental models. The information obtained from these integrated analyses helps us better understand CRC pathophysiology, provides opportunities for developing diagnostic and prognostic...

### Declarations of competing interest

Nima Zafari: None Declared.

Parsa Bathaei: None Declared.

Mahla Velayati: None Declared.

Fatemeh khojasteh-Leylakoochi: None Declared.

Majid Khazaei: None Declared,

Hamid Fiuji: None Declared,

Mohammadreza Nassiri: None Declared.

Seyed Mahdi Hassanian: None Declared.

Gordon A Ferns: None Declared,

Elham Nazari: None Declared,

Amir Avan: None Declared....

---

### References (190)

Z. Ye *et al.*

[Integrated bioinformatics identifies the dysregulation induced by aberrant gene methylation in colorectal carcinoma](#)

Genes & Diseases (2021)

P. Gao *et al.*

[Integrated analysis of gene expression signatures associated with colon cancer from three datasets](#)

Gene (2018)

Y.-R. Wang *et al.*

## Insights regarding novel biomarkers and the pathogenesis of primary colorectal carcinoma based on bioinformatic analysis

Comput. Biol. Chem. (2020)

E. Eskandari *et al.*

## An integrated study on TFs and miRNAs in colorectal cancer metastasis and evaluation of three co-regulated candidate genes as prognostic markers

Gene (2018)

M.J. Hossain *et al.*

## Machine learning and network-based models to identify genetic risk factors to the progression and survival of colorectal cancer

Comput. Biol. Med. (2021)

P.S. Reel *et al.*

## Using machine learning approaches for multi-omics data analysis: a review

Biotechnol. Adv. (2021)

N. Zafari *et al.*

## Role of gut bacterial and non-bacterial microbiota in alcohol-associated liver disease: molecular mechanisms, biomarkers, and therapeutic prospective

Life Sci. (2022)

G. Zhu *et al.*

## Proteomics of post-translational modifications in colorectal cancer: discovery of new biomarkers

Biochim. Biophys. Acta Rev. Canc (2022)

G.M. Chalbatani *et al.*

## Integrative and in-vitro analysis reveal hsa\_circ\_001787 can act as a diagnostic biomarker for colorectal cancer

Saudi J. Biol. Sci. (2021)

L. Chen *et al.*

## Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis

Gene (2019)



View more references

## Cited by (0)

---

## Recommended articles (0)

---

[View full text](#)

© 2023 Published by Elsevier Ltd.



Copyright © 2023 Elsevier B.V. or its licensors or contributors.  
ScienceDirect® is a registered trademark of Elsevier B.V.

