Colorectal Cancer Biomarkers Discovery Approach: A Proteomic and Genomic Perspective

Maryam Hassan Al-Zahrani*

Received: 09 June 2023 / Received in revised form: 30 August 2023, Accepted: 04 September 2023, Published online: 26 September 2023
© Biochemical Technology Society 2014-2023
© Sevas Educational Society 2008

Abstract

Colorectal cancer is a major global health concern, necessitating the identification of biomarkers for early detection, prognosis, and therapeutic targeting. In recent years, proteomic and genomic approaches have revolutionized cancer research, providing valuable insights into the molecular mechanisms underlying CRC. Epidemiological and clinical studies point to a connection between inflammation and development of cancer. Incidences of (CRC) have increased globally during the past decade. In this review, we will discuss the biomarker discovery approach for CRC, which involves a combination of proteomic and genomic perspectives. There are numerous promising biomarkers that must improve life expectancy or quality of life to be evaluated for use in clinical practice. We will cover the important steps such as sample collection, protein extraction, separation, mass spectrometry (MS) analysis, genomic profiling, data analysis, validation, and functional characterization. By utilizing both proteomics and genomics, researchers can identify potential biomarkers for CRC, leading to better diagnostics and personalized treatment options.

Keywords: Colorectal cancer, Biomarkers, Proteomic, Genomic, Gene therapy

Introduction

Cancer is a catastrophic global public health issue, regardless of a country's degree of development (Gari et al., 2021). One million people have diagnosed annually with (CRC), which accounts for 30% of all malignancies.

Incidence of (CRC) have increased globally during the past decade. The third most common adult cancer is CRC, which is also the third leading cause of cancer-related mortality in the United States.

Additionally, the prevalence of CRC is rising in younger people, and by 2030, more people aged 20 to 49 are anticipated to have the disease. CRC prevalence is on the rise in the Middle East, particularly among young people (Coppola et al., 2021). In the Arab world, some variations in disease prevalence and epidemiology have also been found. The Arab population has been influenced by Western lifestyles, which have increased the prevalence of CRC and affected younger generations (Guraya, 2018; Makhlof et al., 2021).

Most CRCs are sporadic and are defined by a sequenced carcinogenesis process involving the progressive accumulation of mutations over 10–15 years on average. This long evolution interval permits the successful use of screening, early cancer identification, and treatment of premalignant lesions, decreasing incidence and death (Aghagozladeh & Radpour, 2016).

Detecting CRC early is the most effective method for reducing cancer mortality. The fecal occult blood test (FOBT), colonoscopy, sigmoidoscopy, and immunological FOBT can detect CRC early enough for effective disease management (Quintero & Salido, 2009; Ettarh, 2012).

Even if there is a potential for an early diagnosis, 20–25% of CRC cases are found at stage IV, when patients have already displayed distant metastasis and the 5-year survival probability is less than 10%. For people with early restricted conditions for whom surgical resection is possible, the 5-year survival rate may exceed 90%. Genetic pathways have a major impact on CRC. More than 25% of patients diagnosed with the condition have a family history of it (Corbo et al., 2012; Zygulska & Pierzchalski, 2022).

The large intestine's epithelial cells are the starting point of the prolonged process that leads to CRC development. In reality, these cells lose their normal biological behavior and develop the characteristics of cancer cells as a result of the accumulation of mutations and the subsequent modification in gene function. Depending on the disease's stage, there are three strategies to halt the progression of CRC. The first pertains to cancer or adenoma at an early stage (Gassler et al., 2010).

The last five years have witnessed a significant increase in research aimed at finding biomarkers that can enhance the current diagnostic and prognostic case for CRC screening and therapy (Alkhayyat et al., 2021).

Types of Biomarkers

Maryam Hassan Al-Zahrani
Department of Biochemistry, Faculty of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia.

*E-mail: mhsalzahrani@kau.edu.sa

© 2023 Journal of Biochemical Technology. Open Access - This article is under the CC BY license https://creativecommons.org/licenses/by/4.0/.
A biomarker is an objectively measurable biological molecule found in bodily fluids or tissues that may be used to identify a pathological state or to indicate whether a biological process is normal or abnormal. Biomarkers can be used to diagnose disease, predict prognosis, and predict pharmacologic responses to therapeutic interventions. Typically, a biomarker must improve life expectancy or quality of life to be evaluated for use in clinical practice (Goossens et al., 2015).

There are three primary categories of biomarkers, depending on the function they serve. Diagnostic biomarkers are likely the most significant biomarkers and are valuable for detecting recurrent disorders. The purpose of prognostic biomarkers is to anticipate the likely course of a disease; they may indicate the aggressiveness and the chance for metastasis. These biomarkers can be used to assess the prognosis of the disease and influence therapy and care decisions. Predictive biomarkers can aid in identifying subpopulations of patients who may benefit from a certain treatment. A predictive biomarker can predict the potential treatment outcomes and can also be employed as a therapy target. It can also signify a "predisposition," or an elevated risk of developing a particular disease. A potential cancer biomarker is any detectable molecular change at the DNA, RNA, protein, or metabolite level in a cancer cell (Atkin, 2003; Diakos et al., 2015).

Yamamoto et al. performed liquid chromatography / (MS) on formalin-fixed and paraffin-embedded (FFPE) CRC tissue using a global proteome method, demonstrating greater expression levels of cyclophilin A, annexin A2, and aldolase A in cancer compared to non-cancerous regions (Yamamoto et al., 2016).

Blood-based biomarkers are perhaps the ideal matrix for early diagnosis and surveillance of CRC due to the ease with which non-invasive, low-cost specimens may be collected. Using targeted liquid chromatography-tandem MS (Clarke et al., 2012; Loktionov, 2020).

These results demonstrate the limitations of existing diagnostic screening and the difficulties of generating surrogate markers for early disease identification. Current non-invasive stool screening methods are not sensitive enough to detect precancerous lesions and may miss early-stage CRC. Therefore, a low threshold must be maintained for more intrusive colonoscopies in these patients, and other technologies are necessary to promote early CRC detection.

It is possible to use prognostic biomarkers to predict disease progression, including early recurrence and mortality. KRAS is a member of the RAS proto-oncogene GTPase family, which inhibits cell growth. Mutations in KRAS are associated with a greater likelihood of metastatic CRC recurrence after curative resection, as well as a lower overall survival following hepatic metastasectomy in metastatic CRC (Bonnot & Passot, 2019).

In clinical practice, the primary predictive biomarker is a carcinoembryonic antigen (CEA), a glycoprotein with a high-molecular-weight produced in embryonic tissue and CRCs. This antigen was identified in 1965, but it continues to be the most extensively utilized blood-based biomarker for CRC (Amilca-Seba et al., 2021; Chen & Ke 2021). Predictive biomarkers are used to personalize therapies based on molecular subtypes. The increasing rise of adjuvant and neoadjuvant therapeutic techniques necessitates the immediate development of predictive biomarkers to guide treatment decisions. An illustration of the significance of predictive biomarkers is the ability of medications to inhibit the epidermal growth factor receptor in patients with KRAS-wild malignancies. The development of this targeting therapy made determining the KRAS status of patients with advanced CRC a prerequisite for determining the efficacy of chemotherapy (Amilca-Seba et al., 2021; Sarkar, 2023).

**Genomics and Proteomics**

Proteomics encompasses a vast array of techniques used for the large-scale identification, measurement, characterization, and analysis of proteins. The bulk of biomarker discovery research uses quantitative MS-based approaches to identify and validate dysregulated proteins as disease biomarker candidates (Anderson & Anderson, 1998). A genomic biomarker is a detectable DNA or RNA characteristic that serves as an indicator of normal biological activities, pathogenic processes, and/or responsiveness to therapeutic or other interventions (Kim & Hahn, 2007; Bodaghi et al., 2023). A genomic biomarker could, for example, be a measurement of gene expression, function, or regulation (Eltayeb et al., 2022).

**Biomarkers Based on Epigenetic Changes for CRC**

Epigenetic modifications cause heritable changes in cellular phenotypes and DNA-coded information. These modifications are independent of DNA sequence and susceptible to chromatin-modifying enzymes. Four DNA modifications and sixteen histone modifications have been identified, with cytosine methylations being the most widely described. Complex diseases like cancer, autoimmune disorders, and mental disorders are linked to altered methylation patterns (Schweiger et al., 2013; Zygulska & Pierzchalski, 2022).

In conjunction with posttranscriptional changes of histones, cytosine methylations are arranged in extensive epigenetic silencing areas (LRES). Genes inside these regions are transcriptionally repressed; for instance, a 4-Mb region on chromosome 3p22 containing the MLH1 gene causes MSI-H CRC (Yamashita et al., 2003).

**Gene Therapy in CRC**

New cancer treatments and tools to analyze genes are leading to a need for dependable biomarkers. Most cancer drugs fail clinical trials, which are expensive and lengthy. To address this, the FDA is prioritizing the use of biomarkers to identify which subtypes of cancer respond best to which treatments. This review covers current trends and challenges in developing effective cancer biomarkers for clinical use (Fearon & Vogelstein, 1990; Jung et al., 2007).
The treatment of CRC is dependent on the TNM staging of cancer, patient health, and curative versus palliative purposes. This includes surgical intervention, chemotherapy, and immunotherapy. The necessity and kind of adjuvant therapy are determined by stage, circumferential resection margin, lymphovascular invasion, perineural invasion, and genotyping (Guetz et al., 2007). 5-FU, commonly used in colon cancer treatment, may harm those with MSI or DPYD. It can improve disease-free survival by 2-4% in stage II CRC, but up to 25% still experience relapse. KRAS wildtype is now being used for better response rates to cetuximab and bevacizumab, while anti-PD-1 drugs treat metastatic CRC. Nivolumab and ipilimumab have demonstrated efficacy in MSI and mismatch repair defective genotypes, resulting in their approval for patients whose cancer progresses after first-line treatment (Lenz et al., 2022).

Genomics and Biomarker Discovery: Strategies and Their Limitations

Adding anti-EGFR biological medicines to chemotherapy treatment for CRC patients with KRAS mutations can improve survival rates and reduce cancer progression. However, it's unclear if KRAS/BRAF wild-type tumors and KRAS wild-type/BRAF mutant cancers respond differently to anti-EGFR therapy. Studies also conflict with the effectiveness of EGFR-targeted therapies for BRAF-mutant CRC (Garcia-Carbonero et al., 2020).

The P53 gene is crucial for preventing uncontrolled cell growth in cancer. CRC is affected by abnormalities in the TP53 pathway, which can impact treatment effectiveness. More research is needed to determine TP53's potential as a biomarker for CRC (McHugh et al., 2009). MSI status is another indicator of high clinical value. Microsatellites are small DNA sequences that repeat throughout the genome. MSI status is often induced by the inactivation of the four MMR genes (Suzuki et al., 2002). Dysbiosis in the intestinal microbiota may contribute to colorectal carcinogenesis by affecting inflammation, DNA damage, and metabolites involved in tumor progression. This can result from impaired intestinal epithelial barrier function, pro-inflammatory responses, genotoxic biosynthesis, and toxic metabolites produced by pathogens (Tanaka et al., 2010; Goossens et al., 2015).

Genomics Discovery Techniques

Next-generation sequencing (NGS) techniques have changed both genomic and transcriptome analyses. NGS platforms provide deep sequencing, which can detect extremely rare genetic variants, and massively parallel sequencing, which can fast and exhaustively cover the human genome. Different NGS approaches are used for whole-genome sequencing, whole-exome sequencing, targeted sequencing, and RNA-seq. These tools are then used to detect changes in both coding and non-coding genomic regions as well as aberrant dynamics in the transcriptome. High sensitivity and massively parallel sequencing enable quick detection of somatic and germline mutations, propelling NGS to the forefront of cancer biomarker research (McDermott et al., 2013; Marks et al., 2018) MS is the primary enabling tool for proteome discovery. Whatever the case, ionization technique, or performance characteristics, all mass spectrometers create mass spectra, which plot the mass-to-charge ratio of the observed ions (x-axis) against the measured ion abundance (y-axis) (Fan et al., 2012).

Technology platforms use pattern-based and identity-based techniques to discover proteomic biomarkers. Pattern-based methods generate protein patterns using techniques like SELDI, MALDI, or electrospray. Identity-based methods use LC-MS/MS analysis to identify peptide sequences from differential protein displays like 2D-PAGE. LC-MS/MS-based techniques have proven to be more sensitive, repeatable, and efficient than 2D-PAGE (Mischak et al., 2009). Using methods such as loss of heterozygosity screening and comparative genomic hybridization, the evaluation of the human genome has become extremely efficient (Fanelli et al., 2020). The expression of genes and expressed sequence tags (ESTs) in laboratory and clinical tumor tissues were compared and are now routinely performed using microarray technology (Roos & Byron, 2019).

By amplifying RNA with fluorescent labels and transferring the tagged transcripts on array slides with a large number of oligonucleotides or cDNAs, six thousand genes can be assessed simultaneously. Expression of the fluorescent label indicates the presence and quantity of a certain cDNA transcript in the test population. By combining microarrays with comparative analysis, patterns of gene expression can be detected by logging differences in gene expression (Subramanian et al., 2005).

Proteomics Discovery Techniques

Proteome Marc Wilkins developed the term "proteome" in 1994 by combining the words "protein" and "genome." The proteome encompasses all of the proteins expressed in an organism, tissue, cell, or biological system. Proteomics is the extensive study of all proteins, with an emphasis on their structures and activities (Anderson & Anderson, 1998).

Proteomics is a breakthrough in protein chemistry, focusing on studying the entire proteome as a single analyte for cellular molecular pathways. This approach allows for an accurate representation of the proteome in a given cell state. However, proteome analysis faces challenges such as protein concentrations, detection of post-translational changes (PTMs), and sample complexity (Jungblut et al., 1999). Even though the separation processes vary, the ultimate phase of each strategy is (MS) analysis, which assigns a name to each protein. Several of the most prevalent technologies utilized in CRC research are listed below (Fallas et al., 2006). This area of proteomics employs the methods outlined below and includes (Engwegen et al., 2006):
1. 1-dimensional electrophoresis
2. 2-dimensional electrophoresis
3. In-gel differential electrophoresis
4. Electrophoretic microarrays of proteins
5. Mass Spectrometry

Utilizing Mass Spectrometry
MS has enabled the development of proteomics despite the challenge of detecting low-abundance proteins. Techniques like SILAC, TMT, and iTRAQ have improved sensitivity and allowed for simultaneous analysis and peptide quantification of multiple specimens. TMT LC-MS/MS has high throughput capability and is vital for lab standardization. It has become a prominent technique in identifying cancer biomarkers, resulting in the subclassification of ovarian, breast, and CRCs. CPTAC has successfully used MS to identify cancer-specific proteins and unique protein patterns (Perry et al., 2008).

Antibody-based techniques for targeted proteomics are not the only ones. MS approaches such as Selected Reaction Monitoring (SRM) and Multiple Reaction Monitoring (MRM) are developing as dependable, high-throughput cancer biomarker tests. Specific peptides coming from the protein of interest are identified using a triple quadrupole mass spectrometer and broken down into smaller components, which are then measured to assess protein abundance.

**Advances in MS**

Through MS-based discovery studies, numerous potential biomarkers for specific diseases have been identified using various technologies. At present, the focus is on creating MS-based MRM scanning techniques to accurately measure the absolute quantity of established proteins in intricate clinical samples. To discover a practical biomarker for therapeutic purposes, customized quantitative proteome profiling methods are necessary, and recent advancements make this increasingly achievable. Nonetheless, the cost of MS instruments and the lack of highly specific antibodies for a significant number of proteins in MS-based biomarker validation methods must be further addressed (Han et al., 2001; Melle et al., 2005; Liu et al., 2006).

**Two-Dimensional Gel Electrophoresis**

Proteomic biomarker discovery typically uses 2-DE, which separates proteins by charge and size on a gel. However, comparing different gels can be difficult due to slight variations. DIGE minimizes variability, but transferring data between labs is problematic (Ünlü et al., 1997). Secretagogin may indicate cancer biological variations due to sample treatment, staining procedures, or image acquisition, and (b) biological variations due to the sample’s production, processing, and preservation environment. Working with several biological and analytical replicates helps reduce these differences, but this increases the complexity of the investigation (Friedman et al., 2004).

**Matrix-Assisted Laser Desorption/Ionization – Time of Flight**

New methods are being used to identify biomarkers, including the Matrix-Assisted Laser Desorption/Ionization – Time of Flight (MALDI-TOF) technology. This involves blending the sample with a matrix molecule that absorbs light at a specific wavelength, then using a laser to transform the sample into gas and eject peptide ions from the surface (de Noo et al., 2006). These ions are separated in a vacuum chamber based on their flight time, and a three-dimensional algorithm is constructed to identify protein clusters. MALDI-TOF has been used to distinguish (CRC) patients from healthy controls, but further validation is needed due to the small sample size and age differences between groups. The technique has also been used to predict metastases in CRC patients, identifying Hsp 27 overexpression as a potential marker for predicting metastatic behavior. These findings are a promising starting point for larger investigations (Liao et al., 2010; Balluff et al., 2011; Kirana et al., 2019).

**Surface-Enhanced Laser Desorption Ionization/Time of Flight**

SELDI-TOF technology has identified 14 protein peaks that could potentially distinguish RCT responders from nonresponders in rectal cancer patients. These peaks were observed 24-48 hours after the initiation of RCT and remained unchanged at baseline (Seibert et al., 2005; Gemoll et al., 2010).

Sadly, it would be unduly optimistic to believe that targeting these separate proteins will boost patient sensitivity in non-responders, given that a particular tumor's resistance to chemotherapeutic drugs probably involves numerous pathways of resistance. Combination therapy targeting various proteins to sensitize the drug-resistant patient is an aspirational goal for the future of cancer treatment, but the technology is not yet advanced enough to make this a reality (Seibert et al., 2005).

**Limitations & Challenges**

Gene-expression profiles have limitations as direct biomarkers, as driver genes may not be differentially expressed at the mRNA level, potentially affecting cancer progression. Furthermore, the differentially expressed genes in a signature may not resolve to one or two distinct gene ontological processes, or the pathways to which they map are unclear, limiting their value as mechanistic study guides. In addition, the expression of mRNA is not always proportional to the expression level of the protein, which is the immediate determinant of cellular phenotype. In these instances, gene transcription level may not necessarily have a significant effect on disease. These restrictions should not be interpreted to suggest that genome-wide assessments of protein-coding mRNAs are no longer useful as indicators of dysregulation, which may play a role in illness. Rather, it is important to emphasize that these data are most likely to be useful when combined with all of the essential information we know about the cell.
Conclusion

Gene and protein changes are only part of the complex cellular changes behind colorectal cancer. Proposed gene expression patterns have limited use in predicting the disease. However, systems biology-based methods show promise in identifying markers for various human disorders, despite the challenges of the 'omics revolution. The combination of high-dimensional results from genomes and proteomics, along with legacy data supporting interatomic databases, has the potential to pave the way for more accurate disease classifiers.

Advances in genetics have led to molecular marker assays for colon cancer screening, but current methods fall short of the ideal. FIT and colonoscopy remain the preferred technique. Blood-based screening with the septin9 biomarker has been approved, but its use for precancerous lesions is under review.

Iscoveries in genetics and cancer development have changed how we treat colorectal cancer. Testing for KRAS, BRAF, and MSI status is important for planning therapy. Immunotherapy and liquid biopsies offer new treatment options. However, there are no biomarkers available yet for early diagnosis, treatment, prognosis, or monitoring. Challenges include small study sizes and difficulties in data analysis and interpretation, and the need for confirmation in larger populations.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References


