

# 114538\_Auto\_Edited.docx

---

WORD COUNT

5136

TIME SUBMITTED

19-DEC-2025 10:29AM

PAPER ID

119556291

**Name of Journal:** *World Journal of Gastroenterology*

**Manuscript NO:** 114538

**Manuscript Type:** MINIREVIEWS

**Noninvasive diagnosis of colorectal adenoma: The emerging potential of blood-based biomarkers**

Qi CY *et al.* blood biomarkers for CRA

Cen-Yi Qi, Ran Wang, Jia-Wei Wang, Guo-Liang Ye, Ping Yang, Yu-Ping Zhou

## **Abstract**

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and the second leading cause of cancer-related death globally. Most CRCs arise from colorectal adenomas (CRAs), particularly advanced adenomas (AAs), which are recognized as critical precancerous lesions. Early detection and intervention at the adenoma stage are essential for alleviating the global disease burden of CRC. However, conventional screening methods such as colonoscopy are invasive and have poor compliance, underscoring the urgent need for efficient, noninvasive diagnostic alternatives. Blood-based biomarkers have gained substantial attention because of their accessibility, reproducibility, and potential for early detection. Advances in multiomics technologies—including proteomics, metabolomics, transcriptomics, and epigenomics—have led to the identification of numerous plasma- and serum-derived biomarkers. These include noncoding RNAs (*e.g.*, miRNAs, circRNAs, piRNAs), DNA methylation signatures, disease-specific proteins, and metabolic profiles. Moreover, emerging platforms such as liquid biopsy, extracellular vesicle profiling, and machine learning further expand the landscape of early CRA detection. The integration of multiomics data holds promise for substantially increasing the sensitivity and specificity of early adenoma detection, offering a transformative framework for precise CRC screening and risk stratification.

**Key Words:** Advanced adenoma; Colorectal adenoma; Circulating biomarkers; Noninvasive screening; Early detection; Colorectal cancer

**Core Tip:** Most colorectal cancers (CRC) arise from adenomas, particularly advanced adenomas (AAs). Blood-based biomarkers identified through multiomics approaches, including proteomics, metabolomics, transcriptomics, and epigenomics, offer a promising noninvasive strategy to enhance the sensitivity and specificity of adenoma detection and colorectal cancer risk assessment.

## INTRODUCTION

Colorectal cancer (CRC) is a major global public health challenge, ranking third in incidence and second in mortality among all malignancies worldwide[1]. In the United States, over 150,000 new CRC cases were projected for 2024, accounting for approximately 7%–8% of all newly diagnosed cancers[2]. Extensive evidence suggests that 70%–90% of CRCs arise through the adenoma–carcinoma sequence, which typically begins with the formation of adenomatous polyps and progresses to invasive carcinoma driven by a series of genetic and epigenetic alterations[3]. Colorectal adenomas (CRAs), particularly those with high-grade dysplasia or specific histological features, are thus recognized as the most clinically relevant precursors of CRC[4]. Early detection and removal of adenomas at the precancerous stage not only interrupt the neoplastic process but also represent the most direct and effective strategy for primary prevention, significantly reducing both CRC incidence and mortality.

Colonoscopy and fecal immunochemical testing (FIT) are the two most widely utilized CRC screening methods[5, 6]. Although colonoscopy offers excellent sensitivity and specificity, its invasiveness, burdensome bowel preparation, and risk of complications substantially limit adherence in average-risk populations[7]. In contrast, FIT noninvasively detects human hemoglobin in the stool to estimate CRC risk. However, its limited sensitivity for the detection of CRAs—particularly advanced adenomas (AAs)—constrains its utility for early-stage detection[8, 9]. Compared with FIT, multitarget stool DNA testing, which integrates DNA-based molecular markers with hemoglobin measurement, has demonstrated superior sensitivity for identifying both CRC and AAs, but this comes at the expense of lower specificity[10]. Collectively, these limitations highlight a critical diagnostic gap, underscoring the need for a modality that combines high sensitivity for precancerous lesions with broad patient acceptance.

In this context, blood-based biomarker testing has emerged as a compelling solution, offering not only greater acceptability but also unique clinical utility in scenarios where conventional methods fall short. In daily clinical practice, physicians frequently

encounter patients who present with vague or nonspecific gastrointestinal symptoms—such as abdominal discomfort or altered bowel habits—that are difficult to interpret definitively. In such instances, FIT may yield false-negative results due to the intermittent or absent bleeding of many precancerous adenomas, leading to a diagnostic dilemma. Unlike stool-based assays that rely on intermittent hemoglobin shedding, blood biomarkers reflect dynamic molecular signals associated with tumor cell turnover and tumor–microenvironment interactions[11]. This capability renders blood testing a valuable tool for objective risk stratification, enabling the identification of high-risk individuals who might otherwise be overlooked due to the absence of overt signs or bleeding. From a logistical perspective, venipuncture is highly convenient and time-efficient[12], requiring neither the bowel preparation demanded by colonoscopy nor the sample collection burdens associated with stool-based testing, making it a more feasible and scalable option for widespread clinical implementation. In recent years, the rapid evolution of omics technologies has fueled increasing interest in the potential of blood-derived biomarkers for the early detection of CRC and the clinical diagnosis of colorectal adenomas.

This review provides a comprehensive synthesis of recent advances in blood-based biomarker discovery for CRAs, focusing on developments across multiple omics disciplines including transcriptomics, epigenetics, proteomics, and metabolomics (Figure 1). By integrating insights from these emerging fields, we seek to facilitate the development of more effective, noninvasive, and scalable diagnostic strategies and to offer a scientific foundation for future translational research.

### NONCODING RNAS

Noncoding RNAs (ncRNAs), including <sup>1</sup>microRNAs (miRNAs), circular RNAs (circRNAs), and PIWI-interacting RNAs (piRNAs), are involved in distinct biogenetic pathways and have broad functional versatility in cancer regulation[13,14]. These transcripts act as critical orchestrators of tumorigenesis and metastasis. Notably, the remarkable stability of circulating ncRNAs in blood renders them promising

noninvasive biomarkers for early CRC screening. Here, we systematically assess their diagnostic potential and prospects for clinical translation.

### *MicroRNAs*

MicroRNAs are endogenous transcripts of approximately 19–25 nucleotides that regulate gene expression posttranscriptionally by binding to the 3' untranslated regions (3'UTRs) of target mRNAs, thereby inducing mRNA degradation or translational repression[15]. Circulating miRNAs exhibit exceptional stability owing to protective mechanisms such as encapsulation within extracellular vesicles (EVs) or complexing with Argonaute 2 (AGO2), which shield them from RNase-mediated degradation[16].

Multiple analytical platforms, including quantitative reverse transcription PCR (qRT-PCR), droplet digital PCR (ddPCR), and next-generation sequencing (NGS), have been applied to blood-based miRNA detection, and recent CRISPR/Cas13a-based systems have further enhanced sensitivity and specificity[17]. With advances in sequencing and computational analysis, an increasing number of clinically informative miRNAs has been identified, supporting their use as promising noninvasive biomarkers for early CRC screening and adenoma detection (Table 1).

Early studies demonstrated the diagnostic utility of plasma miRNAs in colorectal adenomas. Huang *et al.*[18] first reported elevated miR-29a and miR-92a levels in patients with CRA, with AUCs of 0.769 and 0.749, respectively, with a combined AUC of 0.773 when distinguishing AAs from healthy individuals. Subsequent work reinforced the role of miR-92a in adenoma–carcinoma progression[19]. Additional panels have shown discriminatory ability for neoplastic and nonneoplastic lesions, as exemplified by a four-miRNA signature with an AUC of 0.716 and strong validation in TCGA tissues[20]. More recently, a serum miR score comprising seven miRNAs effectively differentiated AAs from nonadvanced adenomas (NAAs) and correlated with progression risk (OR = 2.22; 95%CI: 1.06–4.64)[21]. These findings highlight the potential of serum miRNA profiling as a noninvasive strategy for identifying high-risk AAs and refining early cancer risk assessment.

EV-based liquid biopsy can be used for further advanced CRC screening. EVs are nanoscale lipid bilayer vesicles that carry miRNAs and other biomolecules and protect their RNA cargo from degradation, providing highly stable, analytically tractable targets. Shi *et al.*[22] reported that plasma EV-miR-185-5p levels progressively increased across AA and CRC, with AUCs of 0.737 for AA detection and 0.887 for early CRC detection. Integrative strategies combining WGCNA, RT-qPCR, and machine learning have recently produced optimized classifiers capable of detecting AAs and early CRC with high accuracy (sensitivity up to 99.0%; specificity 79.25%)[23].

Early detection of AAs is critical for reducing the burden of advanced CRC. Despite promising data, circulating miRNA assays remain limited by small cohorts, methodological heterogeneity and limited external validation. Large multicenter prospective studies are needed to standardize these protocols and evaluate their feasibility as noninvasive precolonoscopy screening tools.

### *Circular RNA*

CircRNAs, which are covalently closed loop structures and exhibit exceptional resistance to exonuclease degradation, have emerged as promising biomarkers for noninvasive CRC screening. Their back-splicing origin and absence of 5' caps and 3' poly(A) tails confer remarkable stability, and their broad distribution in blood, urine, and other fluids ensures excellent accessibility for detection. Vo JN and colleagues[24] further demonstrated their superior stability over linear RNAs and developed the MiOncoCirc database based on exon-capture RNA sequencing, providing a sensitive platform for translational circRNA research.

Given the limitations of single biomarkers, multiparametric circRNA models have shown improved diagnostic performance. A plasma CircPanel comprising hsa\_circ\_0001900, hsa\_circ\_0001178, and hsa\_circ\_0005927 effectively distinguished CRC from adenomas, with additional gains when combined with CEA[25].

Similarly, a three-circRNA signature (hsa\_circ\_001978, hsa\_circ\_105039, and hsa\_circ\_103627), identified by microarray and validated by qRT-PCR, achieved

excellent discriminatory power for CRC and CRA ( $AUC \geq 0.966$ )[26]. The integration of circ-CCDC66 with CEA and CA19-9 further enhanced the diagnostic accuracy, increasing the AUC to 0.991, with a sensitivity and specificity above 95%[27].

As a noninvasive modality, plasma circRNA testing can complement colonoscopy and is well suited for population-level early detection. However, clinical translation remains at an early stage. Key challenges include variability in circRNA expression across CRC populations, the absence of standardized diagnostic cut-offs, and the need for validation in large multicenter cohorts with longitudinal follow-up. Moreover, current enrichment methods such as RNase R treatment and Ribo-Zero depletion impose strict requirements for sample quality, involve labor-intensive workflows, and offer suboptimal cost-effectiveness. The development of high-throughput platforms with improved efficiency and sensitivity and broader compatibility is therefore essential for advancing circRNA-based diagnostics.

#### *PIWI-interacting RNA*

PiRNAs are single-stranded small ncRNAs of 24–31 nucleotides that were originally identified in germline cells and were later recognized to perform important regulatory functions in somatic tissues[28]. They are derived mainly from “piRNA clusters,” where long noncoding precursors transcribed by RNA polymerase II undergo sequential processing to generate mature piRNAs[29]. Through their roles in gene regulation, transposon silencing and epigenetic modulation, piRNAs have gained increasing attention as potential diagnostic and prognostic biomarkers across cancers[30], including early CRC and its precancerous stages.

Wang *et al.*[31] integrated small RNA sequencing with RT-qPCR to develop a diagnostic model comprising piR-020619 and piR-020450 for the identification of CRA, achieving an AUC of 0.779, with 72.5% sensitivity and 76.6% specificity. In parallel, serum piR-54265, quantified *via* RT-ddPCR, similarly demonstrated high accuracy for early-stage CRC diagnosis and showed additional potential for postoperative monitoring and risk prediction[32]. However, subsequent evidence demonstrated that

piR-54265 represents a 5' fragment of SNORD57 and that current primer designs may amplify the full-length transcript, raising concerns about detection specificity[33]. These observations highlight key challenges in piRNA research, including complex sequence architecture, strong tissue specificity, incomplete functional annotation, and inconsistent nomenclature and database standards. Technical limitations in sensitivity, specificity and reproducibility further complicate the discrimination of bona fide piRNAs from homologous small RNA species, hindering their clinical translation as reliable biomarkers.

#### *Other noncoding RNAs*

In addition to miRNAs, piRNAs, and circRNAs, other classes of ncRNAs are being increasingly implicated in the early detection of CRAs. Long noncoding RNAs (lncRNAs) influence colorectal tumorigenesis through transcriptional, epigenetic and signaling regulation, and several have shown value in distinguishing adenomas from healthy controls[34]. Recent studies have characterized several previously underexplored ncRNA species, including small nucleolar RNAs (snoRNAs)[35], Y RNAs[36], and enhancer RNAs (eRNAs)[37], many of which display tumor-associated, tissue-specific or fluid-specific expression patterns that underscore their potential as novel biomarkers. Moreover, structural RNAs such as tRNAs and rRNAs generate regulatory fragments –including tRNA-derived fragments (tRFs)– that participate in diverse pathological processes[38]. Although emerging evidence suggests the diagnostic importance of tRFs in CRC[39], their utility in adenomas remains largely unexamined. Advancing the characterization of these ncRNA classes and improving detection platforms will be essential for establishing more sensitive and specific noninvasive strategies for adenoma detection.

#### **DNA METHYLATION**

DNA methylation, a pivotal epigenetic mechanism regulating gene expression, plays a central role in the molecular pathogenesis of CRC. Among the numerous DNA

methylation biomarkers explored for noninvasive CRC detection, *SEPT9* methylation has emerged as one of the most clinically translatable candidates, attracting sustained interest in both research and diagnostic settings[40]. The *SEPT9* gene encodes septin-9, a cytoskeletal GTP-binding protein essential for a wide array of cellular processes, including mitosis, proliferation, apoptosis, vesicle trafficking, and cytoskeletal dynamics. As the first blood-based CRC screening assay approved by the U.S. FDA, the *SEPT9* methylation-based test—commercially available as Epi proColon—is recommended for average-risk individuals aged 50 years and above[41]. Compared with FIT, Epi proColon offers better patient acceptability and compliance[12,42]. Although the sensitivity of methylated *SEPT9* (mSEPT9) for CRC detection is slightly lower than that of FIT (77.0% vs. 88.0%), its specificity is greater (88.0% vs. 80.0%), and the AUC is comparable between the two modalities (0.82 vs. 0.83)[43]. Notably, for early-onset CRC (EOCRC), mSEPT9 alone yields a sensitivity of 90.8%, a specificity of 88.9%, and an AUC of 0.89[44]. However, its sensitivity for detecting AAs remains low at just 6.4%[45].

To improve adenoma detection, researchers have investigated combining mSEPT9 with multiple biomarkers. In a cross-sectional study involving 191 subjects, six additional markers—CEA, carbohydrate antigens CA19-9, CA125, CA72-4, fecal occult blood (FOB), and methylated *DC2*—were concurrently measured. This multimarker approach increased the sensitivity of AA detection to 64.5%[46]. While the inclusion of multiple biomarkers may improve sensitivity, it also raises the risk of false positives, necessitating careful clinical consideration.

Li *et al.*[47] developed a noninvasive screening model integrating three methylation biomarkers, including mSEPT9, coupled with machine learning, achieving a 55.0% detection rate for adenomas. However, the study defined "healthy" controls *via* multitarget stool DNA tests rather than histopathological gold standards, potentially overestimating the model's specificity. Despite existing limitations in sensitivity, specificity, and cost-effectiveness, the 2023 National Comprehensive Cancer Network (NCCN) guidelines classify mSEPT9 as a "nonstandard option". Nonetheless, as a

blood-based assay, mSEPT9 remains a valuable supplement to existing screening strategies, particularly for individuals prioritizing convenience[48].

Another widely studied noninvasive methylation biomarker is the gene *SFRP2*, which encodes a secreted protein that binds to Wnt ligands or Frizzled receptors, thereby inhibiting  $\beta$ -catenin-mediated oncogenic transcription, suppressing proliferation, and promoting apoptosis. Aberrant methylation of *SFRP2* is present in more than 60% of CRC patients[49]. When assessed in plasma, *SFRP2* methylation has a sensitivity of 50.0% for identifying AAs and 63.1% for CRC, with a specificity of 90.1%[50]. To this end, Zhao *et al.*[50] developed the SpecColon blood test by incorporating methylated *SDC2* (mSDC2) alongside *SFRP2*, thereby enhancing early-stage CRC detection performance, with the sensitivity for AAs increasing to 58.3%. These findings emphasize the central role of *SFRP2* methylation in CRC pathogenesis and support its utility as a biologically meaningful biomarker for early detection. Its presence in peripheral blood offers a promising avenue for blood-based screening strategies and warrants further clinical optimization and validation.

In recent years, there has been growing momentum in multigene methylation assays for CRC screening. Compared with single-locus approaches, these multiplexed assays offer superior sensitivity and specificity (Table 2). One notable example is the ColonAiQ test, which integrates six distinct methylation markers and has a sensitivity of 42% for AAs and 86% for CRC, while maintaining a high specificity of 92%[51]. Compared with FIT, ColonAiQ shows markedly improved performance in early-stage CRC detection (85.7% vs. 28.6%) and a higher overall positive predictive value across clinical stages (88.3% vs. 59.7%)[51].

A large-scale prospective study involving over 100,000 average-risk individuals further validated the utility of the ColonAiQ platform, confirming its robust detection capacity for both CRC and precancerous lesions and supporting its broader application in population-based screening programs[52]. Additionally, another study screened 193,000 CpG sites and identified 149 hypermethylated markers to construct a ColonSecure model, which effectively distinguished CRC patients from noncancer

controls in both the training and validation cohorts. At a fixed specificity of 90%, the model achieved sensitivities of 85.3% and 87.0%, respectively, demonstrating excellent classification performance[53]. However, the clinical implications of ColonSecure-positive results in colonoscopy-negative individuals remain unclear. Long-term follow-up data are currently lacking. Consequently, the risks of overdiagnosis, as well as the potential impact on downstream clinical decision-making, have yet to be fully elucidated.

DNA methylation is a promising biological biomarker for CRC and its precursors. However, the limited sensitivity of current methylation assays for detecting advanced adenomas remains a critical barrier. Future research should prioritize the identification of high-performance methylation loci, integrate multiomics data with artificial intelligence (AI)-powered modeling, conduct multicenter prospective validation, and explore the interplay between DNA methylation, the tumor microenvironment (TME), and immune responses. Such efforts are essential to advancing methylation-based assays from adjunctive roles toward routine implementation in CRA screening.

### **PROTEOMICS-DERIVED BIOMARKERS**

Proteins, as the core functional products of genes and key mediators of cellular regulation, are critically involved in tumor initiation and progression through aberrant expression and posttranslational modifications[54]. The activation of multiple signaling pathways in CRC—including the *Wnt/β-catenin*, *EGFR*, *PI3K/AKT*, and *TGF-β* pathways—is dependent on the dysregulated expression of specific proteins, with an imbalance in this protein network driving the malignant transformation of CRAs[55]. The continuous evolution of mass spectrometry (MS)-based proteomic technologies offers powerful tools for systematically identifying differentially expressed proteins characteristic of early disease stages (Table 3).

To optimize early diagnostic strategies for CRC, Zhang J *et al.*[56] developed a high-throughput and streamlined proteomics approach based on plasma-derived EVs, enabling in-depth profiling of protein expression changes throughout intestinal

tumorigenesis. Their study identified multiple circulating proteins with potential for early-stage screening of adenomas and early CRC. Using machine learning algorithms, the team subsequently selected key biomarkers and constructed a panel of 10 signature proteins, which formed the basis of a classification model that demonstrated excellent discriminative performance among healthy controls and patients with CRAs and CRC, achieving AUC values exceeding 0.94 [56]. Following a similar approach, another study integrated proteomics with machine learning (ML) to develop a neural network-driven diagnostic model that combined three EV-derived proteins (FIBG, PDGF- $\beta$ , and TGF- $\beta$ ) with CEA. <sup>4</sup> The model achieved AUCs of 0.822 and 0.767 in the training and validation cohorts, respectively, reflecting robust diagnostic capability[57]. The proteomic alterations observed in blood likely reflect the systemic host response to early neoplasia. For example, the identification of differential levels of inflammatory mediators (*e.g.*, TGF- $\beta$  pathway proteins) in EV cargo, together with the validation of elevated fibrinogen levels in both EVs and total plasma[57,58], suggests that adenomas, even at a precancerous stage, actively remodel the microenvironment and induce systemic immune-metabolic shifts. Understanding these pathway-level dysregulations provides not only diagnostic markers but also potential targets for chemoprevention. Another study expanded on targeted plasma protein detection by employing ML alongside multiple reaction monitoring (MRM) to construct a noninvasive diagnostic model consisting of seven plasma proteins capable of distinguishing patients with CRC from those with precancerous lesions and healthy individuals. In independent ( $n = 253$ ) and blinded ( $n = 84$ ) validation cohorts, the model demonstrated an overall sensitivity of 49% for advanced precancerous conditions (APCs), with a sensitivity approaching 60% for lesions exceeding 1.5 cm, underscoring its promise for early screening and clinical application[59].

At the tissue level, proteomics also elucidates adenoma progression features. Bech JM *et al.*[60] pioneered the use of proteomics to reveal marked molecular heterogeneity within high-grade adenomas and demonstrated that these distinct protein expression profiles predict a 5-10-year risk of synchronous advanced neoplasia. These findings

provide a novel molecular framework for understanding the adenoma-to-carcinoma sequence. However, this study utilized formalin-fixed, paraffin-embedded (FFPE) tissue samples rather than blood samples. Integrating key tissue-level discoveries with dynamic blood-based biomarkers could enable more comprehensive disease surveillance and facilitate earlier clinical intervention.

With advances in multiomics technologies, combined proteomic and metabolomic analyses have revealed multiple regulatory pathways and key molecular players involved in CRC. Vesicle-associated membrane protein-associated protein A (VAPA) was found to be significantly downregulated across CRC tissues and plasma and leukocyte samples, suggesting a strong association with CRC pathogenesis[61]. Further analyses revealed aberrant activation of VAPA-related cholesterol metabolism and the PI3K-Akt pathway at the precancerous (adenoma) stage, highlighting its potential as an early diagnostic marker[61]. Sensitive monitoring of VAPA levels and dynamics in blood could provide an efficient, noninvasive approach for early CRC screening and preventive intervention during the adenoma phase.

In recent years, proteomics-driven identification of differentially expressed biomarkers has shown encouraging results in distinguishing CRAs, including NAAs and AAs, with significant advances, particularly in tissue-based analyses. In contrast, plasma proteomics remains relatively underexplored, primarily owing to the complexity of plasma proteins, which span a dynamic concentration range of up to  $10^{10}$ . The low abundance of disease-relevant proteins, often masked by highly abundant background proteins, poses significant challenges to the sensitivity and specificity of detection. A major bottleneck in plasma proteomics has been this vast dynamic range of protein concentrations. However, recent shifts from traditional mass spectrometry to advanced data-independent acquisition (DIA) modes and proximity extension assays (PEA) have significantly increased the coverage of the low-abundance proteome[62,63]. These technologies allow for the quantification of subtle changes in tissue leakage proteins that were previously undetectable, thereby enhancing the sensitivity for small lesions such as adenomas.

## **METABOLOMICS-DERIVED BIOMARKERS**

Metabolic reprogramming is a hallmark of both cancer initiation and progression. Metabolomics, through high-throughput analytical technologies, enables systematic profiling of small-molecule metabolites in biofluids or tissues, thereby offering critical insights into disease-related metabolic dysregulation and underlying biological mechanisms[64]. Depending on the research objective, the field is broadly divided into targeted and untargeted metabolomics. Targeted metabolomics quantitatively analyzes predefined metabolites, whereas untargeted metabolomics enables comprehensive profiling of the metabolome and serves as a pivotal approach for novel biomarker discovery.

Several studies have utilized high-throughput metabolomics technologies to characterize metabolic differences between CRA and CRC. For example, <sup>1</sup>using ultra-performance liquid chromatography coupled with tandem mass spectrometry (UPLC-MS/MS), researchers have systematically analyzed serum metabolic profiles in CRA and CRC patients and revealed significantly higher serum levels of 3,4,5-trimethoxybenzoic acid in CRC patients[65]. This differential expression suggests its potential utility as a novel metabolic biomarker for distinguishing CRC from CRA in clinical settings. In another untargeted metabolomics study employing gas chromatography-mass spectrometry (GC-MS), analysis of 412 plasma samples revealed several significantly altered metabolites, including ribitol,  $\beta$ -hydroxybutyrate, and lactate. A diagnostic model constructed using these metabolites demonstrated promising discriminatory power for both CRC and CRA patients (AUC > 0.7)[66]. Together, these findings enhance our understanding of the metabolic reprogramming underlying the adenoma-carcinoma sequence and pave the way for the development of novel diagnostic strategies.

In recent years, increasing efforts have been made to integrate host metabolic profiles with the gut microbiota composition to gain deeper insights into the metabolic underpinnings of CRA and CRC while also exploring their potential in early,

noninvasive diagnosis. The integration of metabolomics with microbiome data is particularly relevant for CRA detection. Because the gut vascular barrier may be compromised during early tumorigenesis, microbial metabolites can be translocated into the circulation[67]. This 'gut leakage' phenomenon provides a mechanistic basis for using serum metabolomic profiles as a proxy for gut dysbiosis associated with adenomas, offering a unique advantage over tumor-derived DNA or protein markers, which may be scarce in early lesions. In 2022, Chen F *et al.*[68] conducted the first integrative analysis of serum metabolomics and fecal metagenomics data and identified 885 significantly altered metabolites in patients with CRA and CRC. These authors developed an eight-metabolite gut microbiome-associated serum metabolite panel (GMSM panel), which outperformed CEA testing in detecting both CRA (AUC = 0.84) and early-stage CRC (stage I/II, AUC = 0.93), indicating strong potential for noninvasive screening. This approach was subsequently validated in an independent cohort of 225 individuals from a Shanghai CRC study, where a combined diagnostic model achieved exceptional discriminatory power for both CRA (AUC = 0.912) and CRC (AUC = 0.994)[69]. Additionally, this study revealed elevated serum levels of N-methylproline and trigonelline, as well as significantly reduced levels of sphingosine and its reduced form, in patients with CRA, implicating these metabolites in adenoma-associated metabolic reprogramming. More recently, in 2024, Sun Y *et al.*[70] employed a combined plasma and fecal metabolomics strategy to elucidate the pivotal roles of lipid metabolism and bile acid metabolism in the adenoma-carcinoma transition. Based on these findings, they developed a high-performance plasma-based diagnostic model, providing novel insights into early CRC detection and its underlying metabolic mechanisms. Mechanistically, metabolic reprogramming in adenomas reflects the increased bioenergetic and biosynthetic demands of proliferating cells. The observed alterations in serum metabolites often map to dysregulated glycolysis, lipid metabolism, and the TCA cycle. For instance, the reduction in specific lipids and the accumulation of lactate precursors suggest a shift toward the 'Warburg effect' occurring as early as the adenoma stage[71,72]. In summary, metabolomics has emerged as a

pivotal tool in decoding the metabolic reprogramming underlying the transition from CRA to CRC. From serum and plasma profiling *via* UPLC-MS/MS and GC-MS platforms to multiomics integration with fecal metagenomics, these studies provide a solid foundation for understanding tumor metabolism and identifying clinically actionable biomarkers (Table 4). With continuous technological advancements and increasing cohort sizes, metabolomics-based multiomics strategies hold great promise for precise risk stratification, individualized intervention, and, ultimately, early detection and prevention of CRC.

### **LIMITATIONS OF CURRENT EVIDENCE AND THE PATH FORWARD**

Despite growing interest in blood-based biomarkers for colorectal adenoma, current evidence remains constrained by major methodological flaws and limited reproducibility. Most discovery-phase studies across genomic, proteomic and metabolomic platforms rely on small, retrospective, single-center cohorts, introducing spectrum bias and yielding inflated AUCs with poor external validity. Heterogeneous inclusion criteria, inconsistent lesion definitions and assay variability further hinder cross-study comparability, while asymptomatic screening populations remain markedly underrepresented. Moreover, performance estimates are frequently distorted by incomplete assessment of the full lesion spectrum; for example, a 2024 model integrating methylation, fragmentomic features and copy number variation (CNV) showed excellent accuracy for CRC but did not assess advanced adenomas, an omission that limits its relevance for true screening settings<sup>[73]</sup>. More broadly, even assays reporting strong overall diagnostic performance consistently demonstrate low sensitivity for advanced precursor lesions, underscoring that insufficient evaluation of key premalignant states remains the central barrier to clinical deployment. Together, these gaps in disease-spectrum evaluation, compounded by substantial methodological heterogeneity—from preanalytical handling and EV isolation to normalization strategies and machine-learning architectures—generate pervasive batch effects across omics platforms and severely restrict reproducibility and translation.

Across omics platforms, each modality interrogates a distinct layer of adenoma biology. DNA methylation assays primarily detect stable epigenetic alterations but detect only a minority of advanced adenomas (10–30%)[74]. Circulating ncRNAs capture dynamic regulatory states, proteomics quantifies pathway-level activity, and metabolomics characterizes metabolic and microbial phenotypes. However, the use of circulating ncRNAs remains limited by nonstandardized analytical methods and substantial preanalytical variability; moreover, proteomics is constrained by the extreme dynamic range of plasma proteins and the scarcity of screening-scale datasets, whereas metabolomics, which is informative for metabolic and microbial phenotypes, is highly susceptible to dietary, pharmacologic and circadian confounders. Therefore, these modalities offer complementary but individually incomplete biological insights, strengthening the rationale for integrated multiomics approaches.

Recognizing this need, Hui *et al.*[75] established a 15,000-participant multicenter prospective cohort to overcome retrospective biases and generate biomarker models with genuine population-level generalizability. In the future, substantive progress will require large prospective cohorts supported by standardized preanalytical procedures, transparent computational workflows and rigorous cross-laboratory quality control. Only with such a coordinated methodological discipline can blood-based biomarkers advance from exploratory signals to clinically reliable tools for colorectal adenoma screening.

## **CONCLUSION**

Colorectal adenomas, as established precursors to colorectal cancer, represent a critical target for early identification and risk stratification in efforts to reduce disease burden. A large-scale cohort study involving 296,170 participants revealed a consistent age-associated increase in adenoma incidence[76]. However, CRAs are often asymptomatic early on, making clinical detection difficult. While colonoscopy remains the gold standard for CRA and CRC detection, its invasiveness, low patient compliance, and substantial demands on health care resources limit its applicability for widespread

population screening. To address these challenges, the development of noninvasive, blood-based diagnostic models has emerged as a compelling complementary approach.

Blood testing offers intrinsic advantages such as being rapid, low-risk, and easily integrated into routine health check-ups, thereby increasing screening participation and adherence. Moreover, peripheral blood harbors a rich array of molecular biomarkers, providing multidimensional, noninvasive insights into early disease processes. Leveraging omics technologies, numerous studies have proposed CRA diagnostic models based on noncoding RNAs[77], proteins[78], and metabolites[79], some of which have demonstrated promising performance in small-scale validation cohorts, with favorable AUCs, sensitivity, and specificity.

However, defining the precise clinical utility of these biomarkers is paramount. Rather than serving as standalone replacements for endoscopy, blood-based assays should be positioned as efficient “triage” or risk stratification tools. In asymptomatic populations, a positive biomarker result provides the objective biological evidence necessary to justify invasive investigation. It is crucial to clarify that biomarker integration does not eliminate the need for colonoscopy; instead, it optimizes resource allocation. Noninvasive tests can be used to identify high-risk individuals, whereas subsequent mandatory colonoscopy provides visual confirmation and enables therapeutic polypectomy. This synergistic workflow ensures that invasive resources are targeted precisely toward patients most likely to benefit from intervention.

Despite this promising outlook, the field faces challenges. Most existing studies suffer from small sample sizes, retrospective designs, and single-center cohorts, resulting in limited reproducibility. Furthermore, heterogeneity in analytical platforms and algorithms restricts cross-comparability. Another major gap lies in the inadequate differentiation between AAs and NAAs. Most studies have focused on CRC or mixed polyp populations and have overlooked the specific clinical relevance of AA. However, prospective evidence from a cohort of more than 120,000 individuals has confirmed that AA significantly elevates CRC risk[80], underscoring the critical need for focused biomarker research in this high-risk subgroup.

To overcome these barriers, the next frontier of research lies in integrating multiomics data with artificial intelligence. Machine learning algorithms are essential for extracting meaningful patterns from high-dimensional molecular data. A prominent example of this potential is the DELFI approach<sup>6</sup>[81], which uses machine learning to analyze genome-wide cell-free DNA fragmentation profiles. By deciphering complex, nonintuitive biological features rather than relying solely on mutations, this method achieved a sensitivity of 73%, demonstrating the power of AI in early detection.

Nevertheless, the transition from discovery to clinical application reveals persistent challenges. The prospective PATHFINDER study[82], involving more than 6,000 participants, evaluated a novel blood-based assay that relies on machine learning algorithms to identify cancer-specific methylation patterns in cfDNA, confirming the feasibility of detection in asymptomatic individuals while also revealing critical limitations. Notably, the positive predictive value (PPV) was approximately 40%, implying a substantial burden of false positives that led to unnecessary invasive follow-ups and patient anxiety. Additionally, the “black-box” nature of deep learning models limits biological interpretability, posing a hurdle for clinician trust.

In conclusion, bridging the gap between discovery and clinical implementation requires a shift toward large-scale, prospective validation and standardized analytical frameworks. Concurrently, public health education is vital for fostering acceptance. Ultimately, the goal is to establish a robust two-step screening model that combines the accessibility of blood-based risk stratification with the therapeutic indispensability of colonoscopy. Achieving this paradigm will optimize health care resource allocation and significantly mitigate the global burden of colorectal cancer through earlier, more efficient detection.

2%

SIMILARITY INDEX

---

PRIMARY SOURCES

---

- 1 [www.mdpi.com](http://www.mdpi.com) 39 words — 1%  
Internet
- 2 Peizhu Wang, Zhaolei Zheng, Kui Wang, Jianhua Gu, Yanqing Li. "The burden of colorectal cancer attributable to high body mass index in China: findings from 1990 to 2021 and prediction to 2035", *International Journal of Colorectal Disease*, 2025 21 words — < 1%  
Crossref
- 3 [www.biorxiv.org](http://www.biorxiv.org) 19 words — < 1%  
Internet
- 4 Yang Xie, Min Wang, Haibin Xia, Huifang Sun, Yi Yuan, Jun Jia, Liangwen Chen. "Development and validation of a CECT-based radiomics model for predicting IL1B expression and prognosis of head and neck squamous cell carcinoma", *Frontiers in Oncology*, 2023 14 words — < 1%  
Crossref
- 5 [downloads.hindawi.com](http://downloads.hindawi.com) 13 words — < 1%  
Internet
- 6 [markets.businessinsider.com](http://markets.businessinsider.com) 13 words — < 1%  
Internet

---

EXCLUDE QUOTES ON

EXCLUDE SOURCES OFF

EXCLUDE BIBLIOGRAPHY ON

EXCLUDE MATCHES < 12 WORDS