Cancer Biol Med 2026. doi: 10.20892/j.issn.2095-3941.2025.0762

REVIEW

Microbiota-host interaction in colorectal cancer: emerging computational technology, multi-omics integration, and mechanisms

Yinghong Lu, Jun Yu
Institute of Digestive Disease and Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China

ABSTRACT Colorectal cancer (CRC) remains a major global health burden with the gut microbiome emerging as a critical contributor to tumor initiation and progression. Advances in high-throughput sequencing have deepened our understanding of host-microbe interactions across genomic, transcriptomic, epigenomic, and metabolomic levels. This review synthesizes current knowledge on how microbial communities shape colorectal carcinogenesis, including induction of genomic instability, remodeling of host transcriptional and epigenetic landscapes, and reprogramming of metabolic pathways within the tumor microenvironment. Integrative multi-omics strategies and advanced computational tools are powerful means for dissecting these complex biological systems. However, analytical challenges, such as data compositionality, sparsity, and high dimensionality, still hinder meaningful interpretation. Emerging technologies, like long-read sequencing and bacterial single-cell spatial transcriptomics, are enhancing the resolution and accuracy of microbiota profiling. Finally, the convergence of advanced experimental models, artificial intelligence-driven computational integration, and precision microbiome medicine are highlighted as key avenues for translating microbiome insights into preventive, diagnostic, and therapeutic innovations in CRC.

KEYWORDS Colorectal cancer; microbiota; high-throughput sequencing; multi-omics integration; bioinformatics

Introduction

Colorectal cancer (CRC) ranks third as the most commonly diagnosed cancer and second as the leading cause of cancer-related deaths worldwide[1]. The development of CRC involves a complex interplay between genetic, environmental, and lifestyle factors with chronic inflammation serving as a key contributor. Within this context, the gut microbiota, which constitutes a vast community of bacteria, fungi, archaea, and viruses, has emerged as a dynamic component of the tumor microenvironment (TME) and actively influences cancer initiation and progression[2,3]. This complex ecosystem has a crucial role in maintaining human health through nutrient metabolism, immune modulation, and epithelial barrier integrity[4]. However, perturbations of this ecosystem, referred to as dysbiosis, can convert commensal populations into drivers of tumorigenesis. Compelling evidence from fecal microbiota transplantation (FMT) studies demonstrates that germ-free mice receiving microbiota from CRC patients exhibit increased cell proliferation, a higher number of polyps, greater dysplasia, and elevated inflammatory markers compared to mice receiving microbiota from healthy individuals[5]. Large-scale multi-kingdom metagenomic analyses further identified specific microbiome signatures associated with the adenoma-carcinoma sequence, underscoring the critical role of microbiota in CRC evolution[6-8].

The mechanisms by which the microbiota influences CRC initiation and progression are multifaceted and primarily mediated through direct microbial contact with host cells and the production of bioactive metabolites[9]. For example, some pathogens, including *Fusobacterium nucleatum*, can directly bind to host epithelial E-cadherin *via* surface virulence factors [e.g., fusobacterial adhesin A (FadA)], activating Wnt/β-catenin signaling to promote cellular proliferation[10,11]. Concurrently, microbial components [e.g., lipopolysaccharide (LPS)], engage host pattern-recognition receptors, such as Toll-like receptor 4 (TLR4), triggering downstream NF-κB-mediated inflammation and fostering a tumor-promoting microenvironment[12]. In addition to direct

Correspondence to: Jun Yu
E-mail: junyu@cuhk.edu.hk
ORCID ID: https://orcid.org/0000-0001-5008-2153
Received November 28, 2025; accepted January 12, 2026;
published online February 23, 2026.
Available at www.cancerbiomed.org
©2026 The Authors. Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

contact, microbiota-derived metabolites exert profound and context-dependent influences. Secondary bile acids, such as deoxycholic acid (DCA), which are generated *via* microbial biotransformation of primary bile acids, have been shown to impair cytotoxic $CD8^+$ T cell functionality and thereby promote immune evasion in CRC[13]. Likewise, trimethylamine N-oxide (TMAO), a microbial product of dietary choline metabolism, is associated with increased CRC risk through pro-inflammatory and -angiogenic effects[14]. Collectively, these findings indicate that the gut microbiome functions not as a passive passenger within the TME but as an active trigger and sustained modulator of CRC pathogenesis.

In the past two decades the application of culture-independent, high-throughput genetic sequencing technologies has driven expansion of the human microbiome database and exponential growth in '-omics' data[15]. These advances have shifted the field beyond simple compositional profiling toward integrative analyses that interrogate microbial function, ecology, and molecular interactions with the host. Therefore, dysbiosis in CRC encompasses not only alterations in microbial abundance but also profound functional remodeling, including disruptions in short-chain fatty acid (SCFA) biosynthesis, bile acid metabolism, and the production of genotoxic or pro-inflammatory metabolites[16]. Statistical, computational, and mathematical frameworks now enable researchers to identify core microbial pathways that may disrupt the delicate homeostatic balance with the host, providing mechanistic insight into how microbiota compromise epithelial barrier integrity, modulate antitumor immunity, and contribute to oncogenic signaling[17,18]. Understanding these complex and multilayered interactions is essential for developing targeted therapeutic strategies in the era of microbiome-aware precision oncology.

In this review contemporary advances in microbiota-host interactions in CRC are synthesized across multiple omics layers, highlighting mechanistic insights into how microbes and microbe products modulate tumorigenesis. The prevalent computational frameworks and persistent methodologic challenges that shape current microbiome research are discussed. Finally, innovative technologies and future directions that may ultimately enable the clinical translation of microbiome-informed diagnostics, prognostics, and therapeutics in CRC are explored.

Microbiome and host interplay at multi-omics layers

The intricate interplay between the microbiome and host during CRC development can be systematically characterized across multiple molecular layers, including the genome, transcriptome, epigenome, and metabolome (**Figure 1**). To this end, researchers use a diverse suite of profiling techniques to generate multidimensional datasets and jointly analyze these data with microbiome features to uncover key insights into host-microbiota interactions in CRC (**Table 1**).

Microbiome and host genetic mutation

The gut microbiota contributes to CRC initiation and clonal evolution by directly inducing genomic instability and modulating host oncogenic pathways. Whole-genome sequencing (WGS) and whole-exome sequencing (WES) of CRC tumors have identified distinct mutational signatures that are mechanistically attributed to bacterial genotoxins[36]. The pks^+ island from *Escherichia coli* encodes a series of enzymes responsible for synthesizing colibactin, a potent genotoxin capable of inducing DNA damage in epithelial cells[37]. This damage manifests as interstrand crosslinks and double-strand breaks, ultimately leading to a characteristic mutational signature (i.e., the "colibactin damage motif") that is characterized by preferentially damaging sequences (AAWWTT, where W represents A or T) that lead to genomic instability, which is a hallmark of cancer progression[19]. Colibactin-associated mutations have been identified in > 12% of CRC cases[38], underscoring the substantial contribution to disease burden. Similar genotoxic effects have been observed with other bacterial toxins, including *Bacteroides fragilis* toxin (BFT), which is produced by enterotoxigenic *Bacteroides fragilis* (ETBF), and promotes DNA damage through spermine oxidase-dependent generation of reactive oxygen species (ROS), further establishing direct molecular links between specific gut bacteria and somatic mutations in CRC[39,40].

In addition to inducing *de novo* mutations, the microbiome can reshape the function of key tumor suppressor genes. *TP53*, the most frequently mutated gene in human cancers, is altered in approximately 43% of CRC cases[41]. Mutant p53 typically promotes tumor development, enhances cancer cell survival, and is associated with treatment resistance and poor prognosis. However, when mutant p53 is exposed to gallic acid, a polyphenol metabolite produced by gut commensal microbes, mutant p53 loses the tumor-suppressive property and switches to perform oncogenic functions by over-activating the Wnt signaling pathway, significantly enhancing tumor cell proliferation and invasion[20]. In addition, some microbes can promote p53 ubiquitination and degradation, further attenuating the tumor-suppressive function[42].

Host genetics reciprocally influence microbial colonization, establishing a bidirectional feedback loop that reinforces tumor-promoting interactions. For example, KRAS mutations promote intratumoral colonization of ETBF in CRC by regulating the miRNA3655/SURF6/IRF7/IFNβ axis[21], whereas *F. nucleatum* is preferentially enriched in KRAS p.G12D mutant CRC tumor tissues and contributes to tumorigenesis *via* adhesion to the RNA helicase, DHX15[22]. Similarly, BRAF V600E mutations render intestinal stem cells susceptible

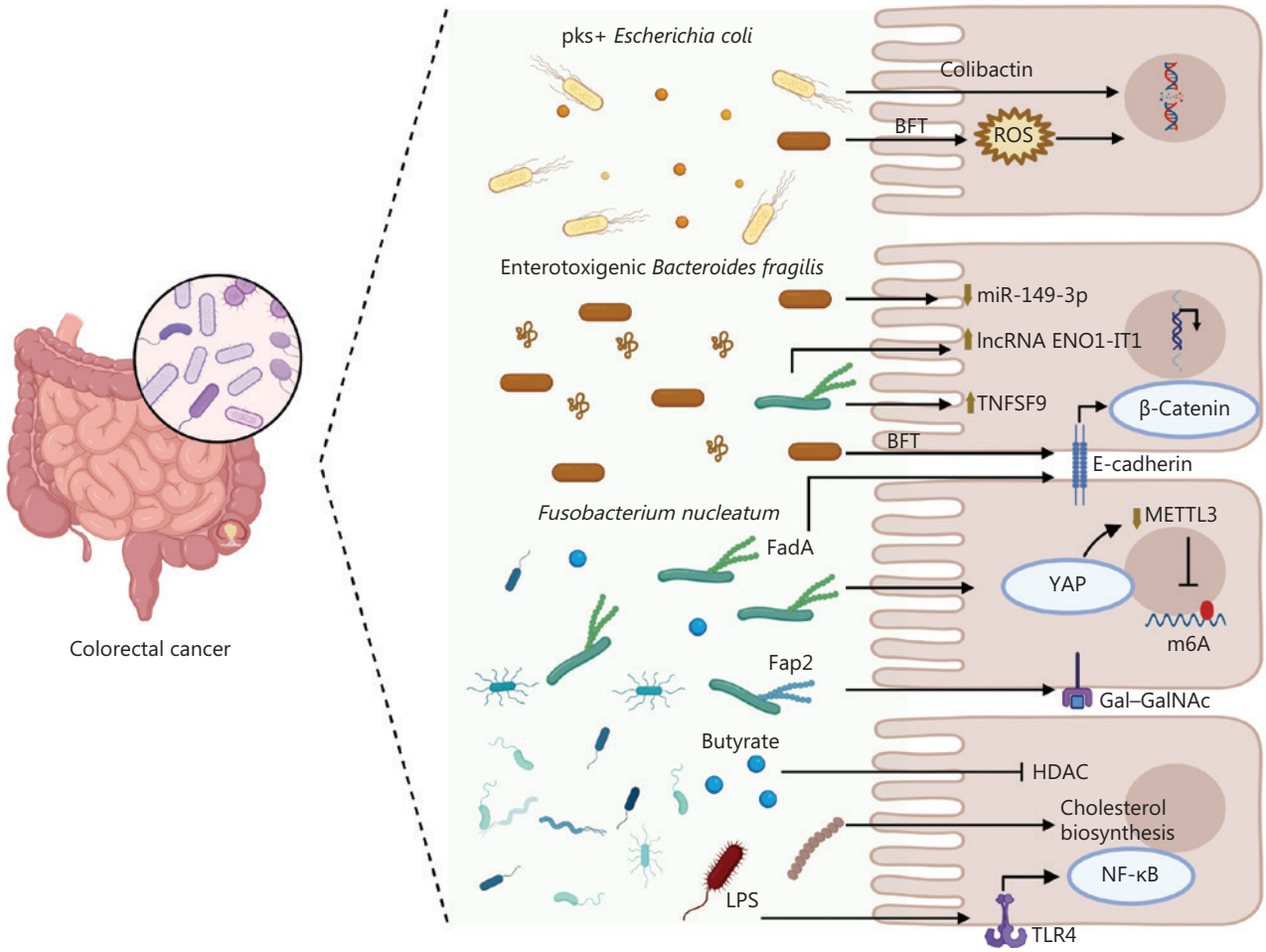


Figure 1 Microbiome and host interaction in CRC. **pks+ *Escherichia coli***: Produces the genotoxin colibactin, which induces DNA double-strand breaks and genomic instability. **Enterotoxigenic *Bacteroides fragilis* (ETBF)**: Secretes BFT, which contributes to ROS generation and induces DNA damage; BFT also disrupts E-cadherin-mediated cell adhesion, activating β-catenin signaling. ETBF downregulates the tumor-suppressive microRNA, miR-149-3p. ***Fusobacterium nucleatum***: Utilizes virulence factors (FadA and Fap2) to bind host E-cadherin and Gal-GalNAc, respectively. F. nucleatum promotes TNFSF9 gene expression, upregulates the lncRNA ENO1-IT1, and suppresses the m6A "writer" enzyme (METTL3) *via* the YAP signaling pathway. Other bacteria promote cholesterol biosynthesis. Commensal-derived metabolites, such as butyrate, function as HDAC inhibitors, while microbial LPS activates TLR4 and NF-κB signaling, collectively fostering an inflammatory and pro-tumorigenic microenvironment. BTF, *Bacteroides fragilis* toxin; CRC, colorectal cancer; Gal-GalNAc, galactose-N-acetyl-d-galactosamine; HDAC, histone deacetylase; LPS, lipopolysaccharide; ROS, reactive oxygen species; TLR4, Toll-like receptor 4.

to colonization with toxigenic bacteria[43], highlighting how common oncogenic drivers can create a niche for specific tumor-promoting microbes.

The search for microbiome-genome interactions is increasingly powered by microbiome-wide association studies (mGWAS), which correlate host genetic variation with microbial abundance[44]. The single nucleotide polymorphism (SNP), rs2355016, which is located in the intron of ATP-sensitive inward rectifier potassium channel 11 (KCNJ11), was shown to be significantly associated with the abundance of *F. nucleatum* in a cohort of 748 Chinese CRC patients. Decreased KCNJ11 expression enhances bacterial adhesion through galactose-N-acetyl-d-galactosamine (Gal-GalNAc) binding and promotes CRC growth[45]. A two-sample Mendelian randomization (MR) study, which tests for potential causal relationships between gut microbiota and health outcomes by comparing summary statistics from a separate microbiome and outcome genome-wide association study (GWAS), has also identified a causal relationship between six bacterial taxa and CRC at a locus-wide significance level[46]. Additional studies indicated that early-life microbial exposures interact with host polymorphisms (e.g., DUOX2 variants) that collectively modulate CRC susceptibility through microbiota-mediated pathways[23].

Table 1 Overview of the application of multi-omics technologies and the main findings in the CRC-related microbiome research

Project	Sample type	Sample size	Sequencing	Integration strategy	Main findings
Dziubańska-Kusibab et al.[19]	Human CRC tissue	*n* = 1372	Whole exome seq, whole genome seq	Correlation, Mann–Whitney *U*-tests	Mutation rates in AAWWTT motifs were enriched compared to all other WWWWWW motifs in CRCs with pks$^+$ *E. coli* infection.
Kadosh et al.[20]	Mouse CRC tissue; stool	*n* = 41	RNA-seq; ChIP-seq; 16S rRNA seq	One-sided student's *t*-test	The gut microbiome switches mutant p53 from tumor-suppressive-to-oncogenic.
Chen et al.[21]	Human CRC tissue; stool	*n* = 40	Metagenomic seq; RNA-seq	PCoA, Pearson correlation, network	KRAS mutations affect the intratumoral colonization of ETBF in CRC through the miR3655/SURF6/IRF7/IFNβ axis.
Zhu et al.[22]	Human and mouse CRC tissue	*n* = 278	16S rRNA seq; targeted gene seq	Correlation, clustering	*F. nucleatum* promotes tumor progression in KRAS p.G12D-mutant CRC by binding to DHX15.
Li et al.[23]	Mouse CRC tissue; stool	*n* = 426	16s rRNA seq; RNA-seq	Mediation analysis; consensus clustering, network	Mediation analysis revealed gut microbiome as mediators partially exerting the effect of SNP UNC3869242 within Duox2 on colorectal tumor susceptibility.
Zou et al.[24]	Human CRC tissue; stool	*n* = 41	Metagenomic DNA seq; exome DNA seq; RNA seq	Spearman correlation, clustering	*F. nucleatum* modified the tumor immune environment by *TNFSF9* gene expression.
Galeano Niño et al.[25]	Human CRC tissue	*n* = 11	16S rRNA seq; 10x Visium spatial single-cell RNA seq	PCoA, clustering, Spearman correlation	Bacteria localize within specific intratumoral microniches characterized by the upregulation of immunosuppressive pathways, and a specific microorganism, including *Fusobacterium* and *Treponema*, are predominantly associated with epithelial and macrophage cell types, driving transcriptional changes linked to metastasis and inflammation.
Hong et al.[26]	Human CRC tissue	*n* = 289	RNA-seq; ChIP-seq	Spearman correlation	*F. nucleatum* activated lncRNA ENO1-IT1 transcription *via* upregulating the binding efficiency of transcription factor, SP1, to the promoter region of lncRNA ENO1-IT1.
Yu et al.[27]	Human CRC tissue	*n* = 296	RNA-seq	Cox regression, clustering	*F. nucleatum* targeted TLR4 and MYD88 innate immune signaling and specific microRNAs to activate the autophagy pathway and alter colorectal cancer chemotherapeutic response.
Ansari et al.[28]	Mouse CRC tissue	*n* = 15	Whole-genome bisulfite seq; ATAC-seq; RNA-seq; ChIP-seq	Hidden Markov modelling	Exposure to commensal microbiota induced localized DNA methylation changes at regulatory elements, which are TET2/3-dependent.
Xia et al.[29]	Human CRC tissue	*n* = 33	Methylated-DNA capture seq	Zero-inflated negative binomial regression, Spearman correlation	*F. nucleatum* and *H. hathewayi* upregulated DNA methyltransferase. *H. hathewayi* inoculation also promoted colonic epithelial cell proliferation in germ-free and conventional mice.

Table 1 Continued

Project	Sample type	Sample size	Sequencing	Integration strategy	Main findings
Liu et al.[30]	Human CRC tissue; stool	*n* = 24	16s rRNA seq; shotgun metagenomic seq; whole-genome bisulfite seq	Lasso penalized regression, network	Gut microbiota and pathogenic bacteria in dynamically shaping DNA methylation patterns, impacting physiologic homeostasis, and contributing to CRC tumorigenesis.
Chen et al.[31]	CRC cell line	*n* = 2	m6A seq, RNA-seq	Clustering	*F. nucleatum* induces a dramatic decline of m6A modifications in CRC cells and PDX tissues by downregulation of an m6A methyltransferase METTL3, contributing to induction of CRC aggressiveness.
Gao et al.[32]	Human CRC stool	*n* = 225	Metagenomic seq; mass spectrometry	PCoA, clustering, random forest, Spearman correlation, network	CRC-associated metabolites were linked to cross-cohort gut microbiome signatures of the disease.
Xu et al.[33]	Human CRC tissue	*n* = 30	16S rRNA; RNA-seq	Spearman correlation, clustering	Intestinal microbiota can affect CRC progression through arginine catabolism.
Chen et al.[34]	Human CRC tissue	*n* = 905	Single cell RNA seq; RNA seq; 16S rRNA seq; metagenomic seq	PCoA, clustering, random forest, Spearman correlation	Host urea cycle metabolism is significantly activated during colorectal tumorigenesis, accompanied by the absence of beneficial bacteria with ureolytic capacity, such as *Bifidobacterium*, and the overabundance of pathogenic bacteria lacking ureolytic function.
Liu et al.[35]	CCSCs	*n* = 3	RNA-seq	Clustering	*F. nucleatum* directly manipulates colorectal cancer cell fate and reveals the mechanism of lipid droplet-mediated Numb degradation for activating Notch signaling.

CCSCs, colorectal cancer stem-like cells; CRC, colorectal cancer; ETBF, enterotoxigenic Bacteroides fragilis; m6A, mRNA N^6-methyladenosine; PCoA, principal coordinates analysis; PDX, patient-derived xenograft.

In summary, the microbiota contributes to CRC-associated genetic instability through both direct genotoxic activity and functional interference with canonical tumor suppressors, while host genetic alterations also shape microbial colonization patterns. This bidirectional interplay solidifies the role of the microbiome as an active contributor to the mutational and ecological landscape of CRC.

Microbiome and host transcriptomic alteration

The gut microbiota exerts significant influence on host gene expression programs during CRC development, reshaping signaling pathways, immune responses, and cellular phenotypes[47]. Researchers have used a wide range of techniques, including bulk RNA sequencing (RNA-seq), single-cell RNA sequencing (scRNA-seq), and spatial transcriptomics, to profile host gene expression, and jointly analyze these data with microbiome features to uncover key insights into host-microbiota interactions. Early bulk RNA-seq of CRC tumor tissues has been instrumental in revealing broad associations between specific microbial presence or dysbiosis and distinct host gene expression programs, identifying links between pathogenic mucosal bacteria and the expression of host genes involved in inflammation, proliferation, and epithelial barrier dysfunction[48]. For example, colonization with *F. nucleatum* activates TNFSF9, a co-stimulatory molecule that modulates tumor-associated immune responses[24].

More recently, advances in single-cell technologies have significantly transformed our understanding of host-microbe interactions by enabling cell type-specific resolution of these complex relationships. Spatial host-microbiome sequencing, a novel technique that combines spatial transcriptomics and 16S rRNA gene amplicon sequencing, allows simultaneous profiling of host gene expression and microbial composition in tissues with spatial resolution. Distinct spatial niches were identified using this approach in the mouse gut. Bacterial genera, such as *Pseudobutyrivibrio* and *Oscillibacter*, were shown to influence the expression of host genes (e.g., *Muc2* and *Ceacam20*), which are involved in mechanisms critical for maintaining gut barrier integrity and immune signaling[49]. The development of invasion-adhesion-directed expression sequencing (INVADEseq) represents a major technological leap, enabling precise identification of bacterial invasion events at the single-cell resolution[25]. *F. nucleatum*-infected epithelial cells exhibit robust induction of inflammatory genes (e.g., *CXCL1*) and matrix-remodeling factors (e.g., MMP9) in CRC tissues, implicating the role of bacteria in driving metastatic potential. Concurrently, distinct immune cell populations, such as tumor-associated macrophages and T cells, exhibit profoundly altered expression profiles in pathways related to DNA repair and cellular dormancy upon bacterial exposure. These findings suggest sophisticated, cell-type-specific mechanisms of bacterial modulation that are entirely obscured in bulk tissue analyses.

In addition to protein-coding mRNAs, the microbiota also exerts influence through the regulation of non-coding (nc)RNAs. For example, *F. nucleatum* upregulates the long non-coding (lnc)RNA, ENO1-IT1, which in turn promotes chemoresistance by activating autophagic pathways that allow cancer cells to survive therapeutic stress[26,27]. In contrast, ETBF downregulates the tumor-suppressive micro (mi)RNA, miR-149-3p, enhancing Th17 cell differentiation and creating a pro-tumorigenic environment. Moreover, microbial suppression of the lncRNA, Snhg9, disrupts the role in stabilizing p53 *via* the SIRT1–CCAR2 complex, illustrating how microbes exploit ncRNA circuits to reinforce oncogenic signaling[50].

In summary, these findings revealed that the gut microbiota shapes CRC progression through cell-type-specific transcriptional reprogramming, direct modulation of cancer-relevant pathways, and sophisticated regulation of ncRNA networks. This comprehensive rewiring of the host transcriptome represents a fundamental mechanism through which microbial communities influence tumor behavior and therapeutic responses.

Microbiome and host epigenomic modulation

Epigenetics refers to the study of heritable changes in gene expression that occur without alterations to the underlying DNA sequence. The primary epigenetic mechanisms include DNA methylation, various histone modifications (e.g., acetylation and methylation), and regulatory ncRNAs, which collectively mediate the interplay between genetic predisposition and environmental factors. The gut microbiota, as a dynamic and abundant environmental factor resident within the host, is a potent modulator of this epigenetic landscape[51,52]. Advances in technologies, such as chromatin immunoprecipitation sequencing (ChIP-seq) for histone modifications, whole-genome bisulfite sequencing for DNA methylation, and assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) for chromatin accessibility, have begun to illuminate how microbial signals directly reshape the host epigenome in CRC.

A growing body of evidence delineates a direct causal link between CRC-associated dysbiosis and aberrant epigenetic patterning. Commensal bacteria have been shown to induce TET2/3-dependent DNA demethylation in intestinal epithelial cells, thereby reprogramming transcriptional responses associated with colitis and CRC[28]. FMT from CRC patients-to-mice induces hypermethylation of the promoter regions of key tumor suppressor genes, such as *SFRP1*, *PENK*, and *WIF1*[53]. Several pathogens, such as *F. nucleatum*, *Hungatella hathewayi*, and *Streptococcus* species, which are consistently enriched in CRC tumors, have been reported to promote hypermethylation of tumor suppressor gene promoters by upregulating the

expression of DNA methyltransferases (DNMTs)[29]. Moreover, a population-based study revealed that CRC tissues infected with a high abundance of *F. nucleatum* are significantly associated with the CpG island methylator phenotype (CIMP)-high subtype and correlate with a poorer patient prognosis[54]. Supporting these findings, integrated multi-omics analyses combining 16S/metagenomic sequencing with whole-genome bisulfite sequencing have consistently identified significant methylome differences between CRC tumors and adjacent normal tissues, pinpointing specific microbial-associated CpG methylation patterns in gene promoter regions[30].

The microbial influence extends beyond DNA methylation to the rapidly growing field of RNA epigenetics, in which bacteria and bacteria-derived metabolites can modulate mRNA N^6-methyladenosine (m6A) modification. For example, *F. nucleatum* has been shown to suppress the m6A "writer" enzyme, METTL3, *via* the YAP signaling pathway, leading to reduced m6A modification and subsequent stabilization of pro-metastatic genes, like *Kif26b*[31]. In contrast, butyrate, a metabolite derived from commensal bacteria, such as *Clostridium*, functions not only as a potent endogenous histone deacetylase inhibitor (HDACi) but also as a key regulator of epi-transcriptomic modifications. Butyrate inhibits CRC development by downregulating METTL3 expression and downstream targets, such as cyclin E1, thereby attenuating processes, like epithelial-mesenchymal transition (EMT)[55,56].

Therefore, the gut microbiota could serve as a master regulator of the host epi-transcriptome, influencing CRC progression through both classical epigenetic mechanisms (DNA methylation) and emerging epi-transcriptomic pathways (m6A modification). While these insights reveal promising therapeutic targets for epigenetic therapy, this field remains nascent and warrants further investigation to fully exploit the potential of targeting the microbiota-epigenome axis for CRC intervention.

Microbiome and host metabolomic crosstalk

Metabolomic crosstalk represents one of the most direct and dynamic interfaces of the microbiota-host interaction in CRC. The gut microbiota functions as a vast, diverse, and highly adaptable bioreactor, transforming dietary and host-derived compounds into a wide array of bioactive metabolites that directly influence host cellular processes. Mass spectrometry-based metabolomics has enabled systematic profiling of these interactions, enabling comprehensive profiling of thousands of small molecules in biological samples, such as serum and stool. Integrated multi-omics approaches have linked specific metabolite profiles to CRC-associated microbiome signatures, identifying both enriched metabolites, such as leucylalanine, serotonin, and imidazole propionate, and depleted species, including perfluorooctane sulfonate and sphingadienine, in CRC patients[32].

Maladaptation in host-microbiota metabolic crosstalk has a critical role in colorectal carcinogenesis. An arginine succinyltransferase (AST)-deficient strain of *E. coli*, Nissle 1917 (ΔacEcN), inhibits intestinal arginine catabolism, leading to arginine accumulation that promotes M2 macrophage polarization and activates Wnt/β-catenin signaling, ultimately accelerating tumor growth[33]. Similarly, multi-omics analyses involving single-cell transcriptomics, microbiome profiling, metabolomics, and clinical data have revealed significant activation of the host urea cycle as a hallmark of colorectal tumorigenesis. This metabolic shift is accompanied by a loss of ureolytic beneficial bacteria, such as *Bifidobacterium*, and an expansion of non-ureolytic pathobionts, collectively disrupting intestinal nitrogen homeostasis and local immune responses[34].

The gut microbiome also profoundly modulates lipid metabolism in CRC. For example, *F. nucleatum* infection reduces lipid accumulation in CRC stem-like cells (CCSCs) by enhancing fatty acid oxidation, thus promoting CCSC self-renewal. *F. nucleatum* infection induces fatty acid formation and promotes lipid accumulation in non-CCSCs, which in turn drives cancer stemness *via* activation of the Notch/Numb signaling pathway[35]. Similarly, *Peptostreptococcus anaerobius* interacts with TLR2 and TLR4 on colon epithelial cells to increase intracellular levels of reactive oxidative species, which promote cholesterol synthesis and cell proliferation *via* SREBP2 and PI3K-Akt-mTOR signaling[57]. Bile acid metabolism constitutes another major node of microbiota-driven metabolic reprogramming. Genes involved in secondary bile acid synthesis are strongly associated with CRC progression[58]. A landmark shotgun metagenomics-metabolomics study demonstrated that the levels of the carcinogenic secondary bile acid, DCA, are markedly elevated in patients with advanced adenomas and *Bilophila wadsworthia* is the only species consistently correlated with DCA abundance[59].

In summary, the metabolome provides a functional readout of the microbiome biochemical activity within the TME. Microbial metabolites act as signaling molecules, immune modulators, and metabolic substrates that collectively shape tumorigenesis. Deciphering this complex crosstalk is essential for developing metabolism-focused interventions and biomarker strategies in CRC.

Computational approaches for microbiome multi-omics integration

The complexity of microbiota-host interactions necessitates computational methods that can integrate disparate omics datasets to infer meaningful biological relationships. These approaches move beyond analyzing each data type in isolation, aiming to construct a holistic model of the interplay between microbial communities and host biology in cancer (**Figure 2**).

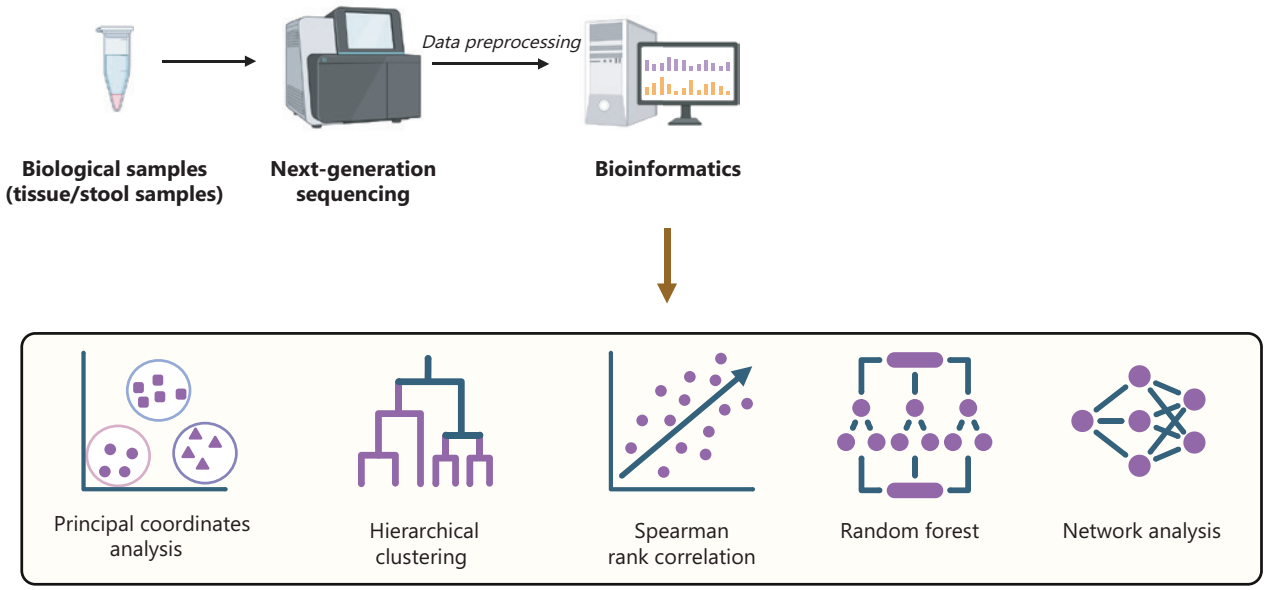


Figure 2 Overview of the multi-omics integration method to reveal host-microbiota interaction. **Principal coordinates analysis** is a fundamental linear method that transforms the original high-dimensional data into a set of orthogonal axes capturing major sources of variation. **Hierarchical clustering** organizes microbiota samples into dendrograms through pairwise distance metrics, such as Bray-Curtis or Jensen-Shannon divergence. The **Spearman rank correlation** is widely used due to robustness of the non-normal distributions typical of microbial data. **Random forest** models can handle high-dimensional data and provide measures of feature importance, ranking microbes by the predictive power. **Network analysis** provides a systems-level framework for integrating multi-omics data by representing complex microbiota-host interactions as unified graphs.

Dimensionality reduction methods

Microbiome and transcriptome datasets are typically high-dimensional, characterized by numerous features [e.g., operational taxonomic units (OTUs) and genes] relative to sample size. Dimensionality reduction techniques provide an essential first step for exploratory data analysis, providing a high-level overview of sample similarities and underlying patterns[60]. Principal component analysis (PCA) and an extension for distance matrices, principal coordinates analysis (PCoA), are fundamental linear methods that transform the original high-dimensional data into a set of orthogonal axes capturing major sources of variation[61]. These techniques are widely used to visualize the overall structure of microbiome data (e.g., based on beta-diversity metrics, like Bray-Curtis or UniFrac distances) and host omics data (e.g., gene expression profiles from RNA-seq), enabling researchers to assess sample relationships, detect potential batch effects, and identify the main sources of biological variation. For more complex, non-linear structures that linear methods fail to capture, non-linear dimensionality reduction techniques, such as t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and projection (UMAP), have gained prominence[62,63]. These methods are particularly powerful for revealing subtle sample clusters that may correspond to biologically or clinically distinct patient subtypes, including those defined by specific microbiome-host interaction states.

Clustering methods

Clustering methods are essential for deciphering complex microbiome data structures by grouping samples or microbial features based on similarity measures[64]. Hierarchical clustering, one of the most established approaches, organizes microbiota samples into dendrograms through pairwise distance metrics, such as Bray-Curtis or Jensen-Shannon divergence. This method enables multi-resolution visualization of sample relationships and is particularly valuable for identifying broad ecological patterns in microbial communities[65]. K-means clustering offers an alternative partitioning approach that groups samples into a predefined number of clusters by minimizing within-cluster variance, providing efficient processing of large datasets despite requiring prior specification of cluster numbers[66]. Partitioning around medoids (PAM) clustering with Jensen-Shannon distance is widely adopted for enterotype analysis due to robustness to outliers, effectively identifying stable states of the gut ecosystem across populations[67]. Recent advances have incorporated machine learning to enhance clustering performance.

The integration of microbial data with host transcriptomic classifications has yielded significant insights into CRC heterogeneity. A landmark study by Guinney et al. established the consensus molecular subtypes (CMS) classification, which categorizes CRC tumors into four distinct subgroups based on gene expression patterns: CMS1 (immune invasive, 14%); CMS2 (canonical, 37%); CMS3 (metabolic dysregulation, 13%); and CMS4 (mesenchymal, 23%)[68]. Subsequent research has revealed critical interactions between these molecular subtypes and specific microbial signatures. For example, *F. nucleatum* enrichment has been identified as a prognostic factor for patients within the CMS4 subtype with studies showing that mesenchymal tumors (CMS4) exhibiting high levels of *Fusobacteriales* are associated with an approximately two-fold higher risk of poor clinical outcomes[69]. Building on this finding, more recent frameworks have integrated microbial community data directly into stratification models, leading to the proposal of onco-microbial community subtypes (OCSs). This classification system stratifies CRCs into three distinct subgroups based on the microbial composition, each demonstrating unique clinic-molecular features and patient outcomes, thereby highlighting the power of integrated omics approaches for refining cancer taxonomy[70]. Furthermore, multi-omics profiling of fecal samples has identified five enterotypes associated with immunotherapy responses in cancer patients, including patients with CRC[71]. These approaches not only reveal microbial community structures but also enable the development of novel classification systems that combine microbial and molecular features, offering new perspectives for personalized cancer diagnostics and therapeutics.

Correlation-based methods

Correlation-based analyses represent a fundamental approach for identifying specific pairwise associations between microbial features (e.g., species abundance) and host molecular features (e.g., gene expression and metabolite concentration). Standard non-parametric tests, like Spearman's rank correlation, are widely used due to the robustness of non-normal distributions, which are typical of microbial data. However, a critical limitation arises from the compositional nature of microbiome data, in which the relative abundance of each taxon is intrinsically dependent on other taxa, often leading to spurious correlations that reflect data structure rather than true biological relationships. To address these limitations, more sophisticated multivariate approaches have been developed. Multivariate statistical methods, including partial least squares regression (PLS), orthogonal partial least squares, and non-metric multidimensional scaling (NMDS), enable the identification of key features contributing to associations across multiple omics datasets. Canonical correlation analysis (CCA) and the sparse variant (sCCA) enable the identification of complex associations between groups of microbial and host features. Priya et al. demonstrated the power of this approach by applying sCCA to colonic mucosal samples from CRC patients, revealing coordinated microbial-host gene networks with both shared and disease-specific interaction patterns[72]. Other advanced methodologies, including weighted correlation network analysis (WGCNA), facilitate the construction of co-expression networks that capture complex interaction patterns, while Procrustes analysis coupled with Mantel testing provides a framework for assessing overall concordance between microbial and host data matrices. This method rotates, scales, and translates one dataset (e.g., microbiome PCoA coordinates) to maximize the similarity with another dataset (e.g., transcriptome PCoA coordinates) and the statistical significance of the association is typically assessed using a Mantel test[73]. Tools, such as MaAsLin2, are specifically designed to identify multivariate associations, while controlling for confounding factors, like age, BMI, and technical batch effects, providing robust statistical frameworks for high-dimensional data[74]. Despite the challenges posed by high-dimensional multi-omics data, these correlation-based methods remain essential tools for generating initial hypotheses in microbiota-host interaction studies. The correlation-based methods provide a crucial foundation for subsequent experimental validation of identified relationships, bridging the gap between observational associations and mechanistic understanding.

Regression and classification methods

Regression and classification methods represent a critical advancement beyond correlation-based analyses, moving from identifying associations-to-building microbiome-based predictive models in which microbial features forecast host-related outcomes[75]. These techniques often involve feature selection, probabilistic inference, model optimization, and performance assessment. Regularized regression techniques, particularly least absolute shrinkage and selection operator (LASSO) and Elastic Net, are well-suited to high-dimensional microbiome data. Regularized regression techniques perform variable selection by penalizing regression coefficients, effectively identifying a small subset of microbial taxa that are most predictive of a continuous host outcome (regression), such as immune cell infiltration score, or a categorical outcome (classification), such as response to immunotherapy. For more complex, non-linear relationships, machine learning algorithms, like random forests, are widely used. These models can handle high-dimensional data and provide measures of feature importance, ranking microbes by predictive power. CRC research has demonstrated particularly successful applications of these methods. For example, a random forest-based multiclass model combining gut bacterial abundances with metabolite markers significantly improved diagnostic performance in distinguishing CRC from other gastrointestinal diseases, showcasing the clinical potential of integrated

microbiome-host feature analysis[76]. The MetaNN framework, a neural network-based deep learning approach, demonstrates superior classification accuracy for host phenotype prediction by leveraging both synthetic and real metagenomic data[77]. These frameworks illustrate the growing potential of predictive modeling to translate microbiome-host interactions into actionable biomarkers.

Network methods

Network analysis provides a systems-level framework for integrating multi-omics data by representing complex microbiota-host interactions as unified graphs[78]. Nodes correspond to entities across different biological layers in these networks (e.g., bacterial species, host genes, and metabolites), while edges represent statistically significant associations between the nodes, including correlations, partial correlations, or more advanced dependency measures. This approach moves beyond pairwise analyses to reveal larger, interconnected functional modules that may drive biological processes. Methods, like graphical LASSO, enable the inference of conditional dependency networks that estimate direct interactions between nodes after accounting for all other variables in the system. Such estimates provide a more accurate representation of potential direct biological relationships than simple correlation analyses. Once constructed, these networks can be analyzed to identify highly interconnected modules, which are clusters of nodes that may represent functional units, such as a group of co-occurring bacteria interacting with host genes involved in specific signaling pathways (e.g., JAK-STAT or Wnt/β-catenin signaling)[79]. Network topology analysis further revealed biologically significant patterns through centrality measures. Nodes with high-degree centrality (many connections) or high-betweenness centrality (positioned on many shortest paths) may represent keystone species or critical host factors central to network stability and function. For example, microbial networks have been shown to vary spatially within tumors and across disease states with specific modules enriched in cancer samples compared to healthy tissue[80]. Functional enrichment analysis [e.g., Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG)] of host genes within these modules helps elucidate the mechanistic pathways through which microbial communities may influence tumor biology. Overall, network methods offer a powerful integrative approach for dissecting multi-omics interactions and identifying system-level drivers of cancer progression.

Challenges in microbiome multi-omics integration

Current approaches to microbiome profiling, primarily through 16S rRNA amplicon sequencing or shotgun metagenomics, have revolutionized our ability to characterize microbial communities. While 16S sequencing provides a cost-effective method for taxonomic classification, particularly valuable in low-biomass environments, shotgun metagenomics provides deeper functional and strain-level insights. However, integrating these microbial data with host multi-omics layers (genome, transcriptome, epigenome, and metabolome) present substantial computational and biological challenges that complicate the extraction of meaningful biological inferences.

Compositionality

A fundamental analytical challenge in microbiome studies arises from the compositional nature of sequencing data. Samples often vary in sequencing depths (library sizes) in high-throughput sequencing results, which yields relative rather than absolute measures of microbial abundance[81]. As a result, an apparent increase in the relative abundance of one microbe necessarily corresponds to a proportional decrease across all other microbes, even if the actual counts remain unchanged. This phenomenon, known as compositionality, can distort common analytical tasks, such as diversity estimation, correlation analysis, and differential abundance testing[82]. Standard statistical techniques assuming absolute measurements may therefore yield misleading results. For example, a significant negative correlation between two bacteria may reflect this mathematical artifact rather than a true biological interaction. To address this issue, data normalization methods, such as rarefaction, scaling, and additive or centered log ratio transform, have been developed to transform compositional data into a form that can be readily analyzed with non-compositional analysis techniques, such as linear models[83].

Sparsity and zero-inflation

Microbiome datasets typically exhibit high sparsity, meaning that many microbial taxa are absent from most samples, leading to an abundance of zero values in the data matrix. These zeros can arise from the following two distinct sources: true biological absence, in which a microbe is genuinely not present; or technical limitations, in which a microbe is present but not detected due to insufficient sequencing depth. The prevalence of zeros violates the assumptions of many conventional statistical models, biases the estimate of diversity and correlation, and complicates the application of machine learning algorithms. Distinguishing between structural and sampling zeros is essential for accurate biological interpretation. Statistical approaches, such as zero-inflated negative binomial and hurdle models, are often used to account for this characteristic, although improved study design, increased sequencing depth, and technical replicates remain the most effective strategies to reduce technical zeros and improve inference.

High dimensionality and biological variability

Microbiome-host studies often face the "small n, large p" problem, in which the number of features (e.g., thousands of microbial taxa, genes, or metabolites) vastly exceeds the number of samples (e.g., dozens or hundreds of patients). In this high-dimensional space, models are highly prone to overfitting, in which the model memorizes noise and patterns specific to the training dataset but fails to generalize to new, independent data. This situation can lead to the identification of numerous false-positive associations. Utilizing regularization techniques (e.g., LASSO, ridge regression), performing rigorous cross-validation, and seeking validation in independent cohorts are necessary steps to mitigate this risk. In addition, the gut microbiota exhibits substantial interindividual variation influenced by diet, genetics, age, medication, and other factors, which can confound true associations. Including these variables as covariates in models and expanding cohort sizes are necessary to distinguish genuine microbiome-host interactions from confounding effects.

Emerging technologies for studying host-microbiota interplay

Advances in sequencing and imaging technologies have expanded our ability to characterize host-microbiota interplay at unprecedented resolution. Long-read sequencing platforms, including PacBio single-molecule real-time (SMRT) sequencing and Oxford Nanopore Technologies (ONT), generate long reads spanning thousands-to-millions of base pairs, overcoming key limitations of short-read sequencing. While short-read approaches remain valuable for affordability, the limited read length often compromises assembly accuracy and taxonomic resolution[84]. In contrast, long-read technologies enable more contiguous genome assemblies, improved detection of structural variants, and direct sequencing of epigenetic modifications, including bacterial DNA methylation patterns. These advances significantly enhance metagenome-assembled genomes (MAGs) in microbiome research by resolving challenging genomic regions, such as repetitive elements, mobile genetic elements, and loci with extreme GC content[85,86]. Recent applications have successfully reconstructed complete bacterial genomes and plasmids from complex microbial communities, providing unprecedented insights into strain-level variation, horizontal gene transfer dynamics, and functional adaptation within host-associated microbiota[87,88].

Beyond bulk taxonomic characterization, emerging single-cell technologies now enable investigation of microbial community function at the resolution of individual cells. The field has witnessed transformative advances with the adaptation of single-cell gene expression profiling techniques to microbial systems, revealing how cellular heterogeneity underpins population-level bacterial dynamics[89,90]. A prime example is bacterial-multiplexed error-robust fluorescence *in situ* hybridization (MERFISH), which integrates expansion microscopy with high-throughput RNA imaging to achieve spatial transcriptomic profiling of individual bacterial cells within complex communities. This methodology overcomes the fundamental challenges of bacterial cell size and high RNA density, enabling researchers to map gene expression dynamics with subcellular resolution under various environmental conditions[91]. Such technologies provide unprecedented opportunities to investigate microbial subpopulations, host-microbe spatial interactions, and transcriptional responses at the single-cell level within native biological contexts, opening new frontiers in understanding the functional basis of host-microbiota interactions in health and disease.

Despite the transformative potential, these emerging technologies present practical limitations that must be considered. Long-read sequencing remains relatively costly and computationally intensive, requiring specialized bioinformatics pipelines for data processing and analysis. Similarly, spatial host-microbiome methods, such as bacterial-MERFISH, involve complex experimental workflows and currently lack standardized analytical frameworks, which can hinder reproducibility and cross-study comparisons. Addressing these challenges will be essential to fully realize the potential of high-resolution technologies in advancing microbiome research and clinical translation.

Conclusions and future directions

In conclusion, this systematic review has delineated the complex and multifaceted interplay between the microbiota and the host in CRC across genomic, transcriptomic, epigenomic, and metabolomic layers. It has been established using integrative multi-omics approaches that the microbiome functions not as a passive bystander but as an active contributor to carcinogenesis, influencing genomic instability, reshaping transcriptional and epigenetic landscapes, and creating a metabolite-rich TME that promotes cancer progression. The application of advanced computational models to multi-omics data has been instrumental in identifying key microbiota-host interactions that may be mechanistically linked to CRC pathophysiology. Despite significant progress, this field continues to face analytical challenges related to data dimensionality, compositionality, and sparsity. However, emerging technologies in long-read sequencing, single-cell microbial omics, spatial transcriptomics, and high-resolution microbial imaging are poised to greatly refine our understanding of microbiota-driven processes in CRC.

Future research should prioritize bridging the gap between correlation and causation through establishing sophisticated experimental models that are capable of recapitulating

human pathophysiology while permitting precise manipulation of microbial and host variables. Germ-free and gnotobiotic mouse systems colonized with defined microbial consortia provide a robust platform for causal inference, while human-derived organoids and organ-on-chip systems co-cultured with specific bacterial strains offer unprecedented opportunities to dissect host-microbe interactions within a patient-relevant context[92]. Specifically, microfluidic-based gut-on-chip systems provide fine control over chemical gradients, shear stress, oxygen tension, and microbial spatial organization, enabling detailed interrogation of microbial behaviors, biofilm formation, and host responses under near-physiologic conditions[93]. Integrating these experimental systems with longitudinal and spatially resolved multi-omics profiling will be essential to elucidate the temporal dynamics of microbiota influence on tumorigenesis and therapy response.

Concurrently, rapid advances in computational methodologies, especially artificial intelligence (AI), machine learning (ML), and deep learning (DL), are reshaping microbiome research[94]. Traditional association-based pipelines are increasingly limited by the complexity, scale, and multimodality of microbiome datasets. In contrast, AI-powered frameworks offer transformative capabilities in modeling non-linear relationships, integrating cross-kingdom omics data, predicting host-microbial interactions, and uncovering latent biological patterns from high-dimensional datasets[95]. For example, protein large language models, such as ProteoGPT, enable high-throughput screening of antimicrobial peptides with high predicted efficacy and minimal toxicity[96], while explainable AI approaches, such as Shapley Additive Explanations, have revealed reproducible microbial signatures of CRC, including *Fusobacterium*, *Peptostreptococcus*, and *Parvimonas*, and improved subtype stratification[97,98]. These computational advances are not only enhancing mechanistic insight but also paving the way for clinically oriented prediction frameworks capable of stratifying patients, forecasting treatment outcomes, and identifying microbial biomarkers with greater accuracy and robustness[99].

Efforts are also underway to translate microbiome insights into preventive, diagnostic, and therapeutic innovations in CRC, including probiotic administration and elimination of deleterious microorganisms, such as *F. nucleatum*, *E. coli*, or *B. fragilis*, from the TME[100]. Advances in precision microbiome medicine requires the development of comprehensive computational frameworks that integrate multi-omics data into predictive models of individual host-microbiome interactomes[101]. Such "digital twin" approaches could revolutionize personalized cancer care by forecasting responses to dietary interventions, prebiotics, and next-generation live biotherapeutic products[102]. Engineering bacteria capable of modulating specific cancer-relevant pathways, such as immune activation, barrier function restoration, or metabolic pathway manipulation, offers a promising avenue for microbiome-based therapeutics[103]. However, by realizing these potential demands large-scale, longitudinal cohort studies with standardized methodologies for sample processing, sequencing, and computational analysis can be conducted to ensure robust and reproducible biomarker discovery[104]. Beyond these technical challenges, clinical translation faces additional challenges, including substantial inter-individual variability in microbiome composition and function, the need for standardized protocols for microbiome modulation therapies, and the development of clear regulatory pathways for live biotherapeutic products. Addressing these barriers will be essential for translating microbiome research into clinically actionable interventions for CRC prevention and treatment.

In summary, the future of microbiome research in CRC lies at the convergence of advanced experimental models, AI-driven computational integration, and large-scale clinical data integration. By addressing current challenges in mechanistic validation, methodologic standardization, and clinical translation, the potential of microbiome research can be harnessed to develop novel strategies for CRC prevention, diagnosis, and treatment, ultimately improving patient outcomes and reshaping our understanding of cancer biology.

Grant support

This work was supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (Grant No. 2023ZD0500200), the Strategic Seed Funding Collaboration Research Scheme CUHK (Grant No. 3133344), the Strategic Impact Enhancement Fund CUHK (Grant No. 3135509), and the Impact Case for RAE CUHK (Grant No. 3134277).

Conflict of interest statement

No potential conflicts of interest are disclosed.

Author contributions

Conceived and designed the analysis: Jun Yu.
Wrote the paper: Yinghong Lu.

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024; 74: 12-49.
2. Janney A, Powrie F, Mann EH. Host-microbiota maladaptation in colorectal cancer. Nature. 2020; 585: 509-17.
3. Wong CC, Yu J. Gut microbiota in colorectal cancer development and therapy. Nat Rev Clin Oncol. 2023; 20: 429-52.

4. White MT, Sears CL. The microbial landscape of colorectal cancer. Nat Rev Microbiol. 2024; 22: 240-54.
5. Wong SH, Zhao L, Zhang X, Nakatsu G, Han J, Xu W, et al. Gavage of fecal samples from patients with colorectal cancer promotes intestinal carcinogenesis in germ-free and conventional mice. Gastroenterology. 2017; 153: 1621-33.e6.
6. Nakatsu G, Zhou H, Wu WKK, Wong SH, Coker OO, Dai Z, et al. Alterations in enteric virome are associated with colorectal cancer and survival outcomes. Gastroenterology. 2018; 155: 529-41.e5.
7. Li T, Coker OO, Sun Y, Li S, Liu C, Lin Y, et al. Multi-cohort analysis reveals altered archaea in colorectal cancer fecal samples across populations. Gastroenterology. 2025; 168: 525-38.e2.
8. Lin Y, Lau HC, Liu Y, Kang X, Wang Y, Ting NL, et al. Altered mycobiota signatures and enriched pathogenic aspergillus rambellii are associated with colorectal cancer based on multicohort fecal metagenomic analyses. Gastroenterology. 2022; 163: 908-21.
9. Long X, Wong CC, Tong L, Chu ESH, Ho Szeto C, Go MYY, et al. Peptostreptococcus anaerobius promotes colorectal carcinogenesis and modulates tumour immunity. Nat Microbiol. 2019; 4: 2319-30.
10. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. Cell Host Microbe. 2013; 14: 195-206.
11. Rubinstein MR, Baik JE, Lagana SM, Han RP, Raab WJ, Sahoo D, et al. *Fusobacterium nucleatum* promotes colorectal cancer by inducing Wnt/β-catenin modulator Annexin A1. EMBO Rep. 2019; 20: e47638.
12. Zhang M, Liu K. Mechanisms of lipopolysaccharide-driven progression and metastasis of colorectal cancer. Int J Biol Macromol. 2025; 329: 147856.
13. Cong J, Liu P, Han Z, Ying W, Li C, Yang Y, et al. Bile acids modified by the intestinal microbiota promote colorectal cancer growth by suppressing $CD8^+$ T cell effector functions. Immunity. 2024; 57: 876-89.e11.
14. Jalandra R, Dalal N, Yadav AK, Verma D, Sharma M, Singh R, et al. Emerging role of trimethylamine-N-oxide (TMAO) in colorectal cancer. Appl Microbiol Biotechnol. 2021; 105: 7651-60.
15. Integrative HMP (iHMP) Research Network Consortium. The Integrative Human Microbiome Project. Nature. 2019; 569: 641-48.
16. Sun Y, Zhang X, Hang D, Lau HC, Du J, Liu C, et al. Integrative plasma and fecal metabolomics identify functional metabolites in adenoma-colorectal cancer progression and as early diagnostic biomarkers. Cancer Cell. 2024; 42: 1386-400.e8.
17. Bertocchi A, Carloni S, Ravenda PS, Bertalot G, Spadoni I, Lo Cascio A, et al. Gut vascular barrier impairment leads to intestinal bacteria dissemination and colorectal cancer metastasis to liver. Cancer Cell. 2021; 39: 708-24.e11.
18. Liu W, Zhang X, Xu H, Li S, Lau HC, Chen Q, et al. Microbial community heterogeneity within colorectal neoplasia and its correlation with colorectal carcinogenesis. Gastroenterology. 2021; 160: 2395-408.
19. Dziubańska-Kusibab PJ, Berger H, Battistini F, Bouwman BAM, Iftekhar A, Katainen R, et al. Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. Nat Med. 2020; 26: 1063-69.
20. Kadosh E, Snir-Alkalay I, Venkatachalam A, May S, Lasry A, Elyada E, et al. The gut microbiome switches mutant p53 from tumour-suppressive to oncogenic. Nature. 2020; 586: 133-38.
21. Chen Y, Liu S, Tan S, Zheng Y, Chen Y, Yang C, et al. KRAS mutations promote the intratumoral colonization of enterotoxigenic bacteroides fragilis in colorectal cancer through the regulation of the miRNA3655/SURF6/IRF7/IFNβ axis. Gut Microbes. 2024; 16: 2423043.
22. Zhu H, Li M, Bi D, Yang H, Gao Y, Song F, et al. *Fusobacterium nucleatum* promotes tumor progression in *KRAS* p.G12D-mutant colorectal cancer by binding to DHX15. Nat Commun. 2024; 15: 1688.
23. Li D, Zhong C, Yang M, He L, Chang H, Zhu N, et al. Genetic and microbial determinants of azoxymethane-induced colorectal tumor susceptibility in Collaborative Cross mice and their implication in human cancer. Gut Microbes. 2024; 16: 2341647.
24. Zou S, Yang C, Zhang J, Zhong D, Meng M, Zhang L, et al. Multi-omic profiling reveals associations between the gut microbiome, host genome and transcriptome in patients with colorectal cancer. J Transl Med. 2024; 22: 175.
25. Galeano Niño JL, Wu H, LaCourse KD, Kempchinsky AG, Baryiames A, Barber B, et al. Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. Nature. 2022; 611: 810-17.
26. Hong J, Guo F, Lu SY, Shen C, Ma D, Zhang X, et al. *F. nucleatum* targets lncRNA ENO1-IT1 to promote glycolysis and oncogenesis in colorectal cancer. Gut. 2021; 70: 2123-37.
27. Yu T, Guo F, Yu Y, Sun T, Ma D, Han J, et al. Fusobacterium nucleatum promotes chemoresistance to colorectal cancer by modulating autophagy. Cell. 2017; 170: 548-63.e16.
28. Ansari I, Raddatz G, Gutekunst J, Ridnik M, Cohen D, Abu-Remaileh M, et al. The microbiota programs DNA methylation to control intestinal homeostasis and inflammation. Nat Microbiol. 2020; 5: 610-19.
29. Xia X, Wu WKK, Wong SH, Liu D, Kwong TNY, Nakatsu G, et al. Bacteria pathogens drive host colonic epithelial cell promoter hypermethylation of tumor suppressor genes in colorectal cancer. Microbiome. 2020; 8: 108.
30. Liu Z, Zhang Q, Zhang H, Yi Z, Ma H, Wang X, et al. Colorectal cancer microbiome programs DNA methylation of host cells by affecting methyl donor metabolism. Genome Med. 2024; 16: 77.
31. Chen S, Zhang L, Li M, Zhang Y, Sun M, Wang L, et al. Fusobacterium nucleatum reduces METTL3-mediated m^6A modification and contributes to colorectal cancer metastasis. Nat Commun. 2022; 13: 1248.
32. Gao R, Wu C, Zhu Y, Kong C, Zhu Y, Gao Y, et al. Integrated analysis of colorectal cancer reveals cross-cohort gut microbial signatures and associated serum metabolites. Gastroenterology. 2022; 163: 1024-37.e9.
33. Xu S, Zhang Y, Ding X, Yang Y, Gao J, Zou N, et al. Intestinal microbiota affects the progression of colorectal cancer by

participating in the host intestinal arginine catabolism. Cell Rep. 2025; 44: 115370.
34. Chen H, Tong T, Lu SY, Ji L, Xuan B, Zhao G, et al. Urea cycle activation triggered by host-microbiota maladaptation driving colorectal tumorigenesis. Cell Metab. 2023; 35: 651-66.e7.
35. Liu H, Du J, Chao S, Li S, Cai H, Zhang H, et al. Fusobacterium nucleatum promotes colorectal cancer cell to acquire stem cell-like features by manipulating lipid droplet-mediated numb degradation. Adv Sci (Weinh). 2022; 9: e2105222.
36. Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, van Hoeck A, Wood HM, Nomburg J, et al. Mutational signature in colorectal cancer caused by genotoxic pks$^+$ E. coli. Nature. 2020; 580: 269-73.
37. Wilson MR, Jiang Y, Villalta PW, Stornetta A, Boudreau PD, Carrá A, et al. The human gut bacterial genotoxin colibactin alkylates DNA. Science. 2019; 363: eaar7785.
38. Cornish AJ, Gruber AJ, Kinnersley B, Chubb D, Frangou A, Caravagna G, et al. The genomic landscape of 2,023 colorectal cancers. Nature. 2024; 633: 127-36.
39. Boleij A, Hechenbleikner EM, Goodwin AC, Badani R, Stein EM, Lazarev MG, et al. The bacteroides fragilis toxin gene is prevalent in the colon mucosa of colorectal cancer patients. Clin Infect Dis. 2015; 60: 208-15.
40. Goodwin AC, Destefano Shields CE, Wu S, Huso DL, Wu X, Murray-Stewart TR, et al. Polyamine catabolism contributes to enterotoxigenic bacteroides fragilis-induced colon tumorigenesis. Proc Natl Acad Sci U S A. 2011; 108: 15354-9.
41. Liebl MC, Hofmann TG. The role of p53 signaling in colorectal cancer. Cancers (Basel). 2021; 13: 2125.
42. Wei J, Nagy TA, Vilgelm A, Zaika E, Ogden SR, Romero-Gallo J, et al. Regulation of p53 tumor suppressor by *Helicobacter pylori* in gastric epithelial cells. Gastroenterology. 2010; 139: 1333-43.
43. DeStefano Shields CE, White JR, Chung L, Wenzel A, Hicks JL, Tam AJ, et al. Bacterial-driven inflammation and mutant *BRAF* expression combine to promote murine colon tumorigenesis that is sensitive to immune checkpoint therapy. Cancer Discov. 2021; 11: 1792-807.
44. Hughes DA, Bacigalupe R, Wang J, Rühlemann MC, Tito RY, Falony G, et al. Genome-wide associations of human gut microbiome variation and implications for causal inference analyses. Nat Microbiol. 2020; 5: 1079-87.
45. Yu J, Liang Y, Zhang Q, Ding H, Xie M, Zhang J, et al. An interplay between human genetics and intratumoral microbiota in the progression of colorectal cancer. Cell Host Microbe. 2025; 33: 657-70.e6.
46. Xiang Y, Zhang C, Wang J, Cheng Y, Wang L, Tong Y, et al. Identification of host gene-microbiome associations in colorectal cancer patients using mendelian randomization. J Transl Med. 2023; 21: 535.
47. Richards AL, Muehlbauer AL, Alazizi A, Burns MB, Findley A, Messina F, et al. Gut microbiota has a widespread and modifiable effect on host gene regulation. mSystems. 2019; 4: e00323-18.
48. Flemer B, Lynch DB, Brown JM, Jeffery IB, Ryan FJ, Claesson MJ, et al. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. Gut. 2017; 66: 633-43.
49. Lötstedt B, Stražar M, Xavier R, Regev A, Vickovic S. Spatial host-microbiome sequencing reveals niches in the mouse gut. Nat Biotechnol. 2024; 42: 1394-403.
50. Wang M, Liu K, Bao W, Hang B, Chen X, Zhu X, et al. Gut microbiota protect against colorectal tumorigenesis through lncRNA Snhg9. Dev Cell. 2025; 60: 1008-17.e7.
51. Tian S, Chen M. Global research progress of gut microbiota and epigenetics: bibliometrics and visualized analysis. Front Immunol. 2024; 15: 1412640.
52. Pepke ML, Hansen SB, Limborg MT. Unraveling host regulation of gut microbiota through the epigenome-microbiome axis. Trends Microbiol. 2024; 32: 1229-40.
53. Sobhani I, Bergsten E, Couffin S, Amiot A, Nebbad B, Barau C, et al. Colorectal cancer-associated microbiota contributes to oncogenic epigenetic signatures. Proc Natl Acad Sci U S A. 2019; 116: 24285-95.
54. Koi M, Okita Y, Carethers JM. *Fusobacterium nucleatum* infection in colorectal cancer: Linking inflammation, DNA mismatch repair and genetic and epigenetic alterations. J Anus Rectum Colon. 2018; 2: 37-46.
55. Zhang K, Dong Y, Li M, Zhang W, Ding Y, Wang X, et al. Clostridium butyricum inhibits epithelial-mesenchymal transition of intestinal carcinogenesis through downregulating METTL3. Cancer Sci. 2023; 114: 3114-27.
56. Zhu W, Si Y, Xu J, Lin Y, Wang JZ, Cao M, et al. Methyltransferase like 3 promotes colorectal cancer proliferation by stabilizing CCNE1 mRNA in an m6A-dependent manner. J Cell Mol Med. 2020; 24: 3521-33.
57. Tsoi H, Chu ESH, Zhang X, Sheng J, Nakatsu G, Ng SC, et al. Peptostreptococcus anaerobius induces intracellular cholesterol biosynthesis in colon cells to induce proliferation and causes dysplasia in mice. Gastroenterology. 2017; 152: 1419-33.e5.
58. Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nat Med. 2019; 25: 679-89.
59. Devkota S, Wang Y, Musch MW, Leone V, Fehlner-Peach H, Nadimpalli A, et al. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in *IL10*$^{-/-}$ mice. Nature. 2012; 487: 104-8.
60. Armstrong G, Rahman G, Martino C, McDonald D, Gonzalez A, Mishne G, et al. Applications and comparison of dimensionality reduction methods for microbiome data. Front Bioinform. 2022; 2: 821861.
61. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci. 2016; 374: 20150202.
62. Yang Y, Sun H, Zhang Y, Zhang T, Gong J, Wei Y, et al. Dimensionality reduction by UMAP reinforces sample heterogeneity analysis in bulk transcriptomic data. Cell Rep. 2021; 36: 109442.
63. Xu X, Xie Z, Yang Z, Li D, Xu X. A t-SNE based classification approach to compositional microbiome data. Front Genet. 2020; 11: 620143.

64. Shi Y, Zhang L, Peterson CB, Do KA, Jenq RR. Performance determinants of unsupervised clustering methods for microbiome data. Microbiome. 2022; 10: 25.
65. Berrazeg M, Drissi M, Medjahed L, Rolain JM. Hierarchical clustering as a rapid tool for surveillance of emerging antibiotic-resistance phenotypes in *Klebsiella pneumoniae* strains. J Med Microbiol. 2013; 62: 864-74.
66. Koo HK, Min J, Kim HW, Ko Y, Oh JY, Jeong YJ, et al. Cluster analysis categorizes five phenotypes of pulmonary tuberculosis. Sci Rep. 2022; 12: 10084.
67. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. Nature. 2011; 473: 174-80.
68. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015; 21: 1350-6.
69. Salvucci M, Crawford N, Stott K, Bullman S, Longley DB, Prehn JHM. Patients with mesenchymal tumours and high *Fusobacteriales* prevalence have worse prognosis in colorectal cancer (CRC). Gut. 2022; 71: 1600-12.
70. Mouradov D, Greenfield P, Li S, In EJ, Storey C, Sakthianandeswaren A, et al. Oncomicrobial community profiling identifies clinicomolecular and prognostic subtypes of colorectal cancer. Gastroenterology. 2023; 165: 104-20.
71. Zhu X, Hu M, Huang X, Li L, Lin X, Shao X, et al. Interplay between gut microbial communities and metabolites modulates pan-cancer immunotherapy responses. Cell Metab. 2025; 37: 806-23.e6.
72. Priya S, Burns MB, Ward T, Mars RAT, Adamowicz B, Lock EF, et al. Identification of shared and disease-specific host gene-microbiome associations across human diseases using multi-omic integration. Nat Microbiol. 2022; 7: 780-95.
73. Quilodrán CS, Currat M, Montoya-Burgos JI. Benchmarking the Mantel test and derived methods for testing association between distance matrices. Mol Ecol Resour. 2025; 25: e13898.
74. Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, et al. Multivariable association discovery in population-scale meta-omics studies. PLoS Comput Biol. 2021; 17: e1009442.
75. Marcos-Zambrano LJ, Karaduzovic-Hadziabdic K, Loncar Turukalo T, Przymus P, Trajkovik V, Aasmets O, et al. Applications of machine learning in human microbiome studies: a review on feature selection, biomarker identification, disease prediction and treatment. Front Microbiol. 2021; 12: 634511.
76. Coker OO, Liu C, Wu WKK, Wong SH, Jia W, Sung JJY, et al. Altered gut metabolites and microbiota interactions are implicated in colorectal carcinogenesis and can be non-invasive diagnostic biomarkers. Microbiome. 2022; 10: 35.
77. Lo C, Marculescu R. MetaNN: Accurate classification of host phenotypes from metagenomic data using neural networks. BMC Bioinformatics. 2019; 20: 314.
78. Bodein A, Scott-Boyer MP, Perin O, Lê Cao KA, Droit A. Interpretation of network-based integration from multi-omics longitudinal data. Nucleic Acids Res. 2022; 50: e27.
79. Silva-García O, Valdez-Alarcón JJ, Baizabal-Aguirre VM. Wnt/β-catenin signaling as a molecular target by pathogenic bacteria. Front Immunol. 2019; 10: 2135.
80. Dai Z, Coker OO, Nakatsu G, Wu WKK, Zhao L, Chen Z, et al. Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. Microbiome. 2018; 6: 70.
81. Quinn TP, Erb I, Richardson MF, Crowley TM. Understanding sequencing data as compositions: an outlook and review. Bioinformatics. 2018; 34: 2870-78.
82. Lin H, Peddada SD. Analysis of microbial compositions: a review of normalization and differential abundance analysis. NPJ Biofilms Microbiomes. 2020; 6: 60.
83. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome. 2017; 5: 27.
84. Gounot JS, Chia M, Bertrand D, Saw WY, Ravikrishnan A, Low A, et al. Genome-centric analysis of short and long read metagenomes reveals uncharacterized microbiome diversity in Southeast Asians. Nat Commun. 2022; 13: 6044.
85. Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. Nat Biotechnol. 2020; 38: 701-7.
86. Tamburini FB, Maghini D, Oduaran OH, Brewster R, Hulley MR, Sahibdeen V, et al. Short- and long-read metagenomics of urban and rural South African gut microbiomes reveal a transitional composition and undescribed taxa. Nat Commun. 2022; 13: 926.
87. Lai S, Wang H, Bork P, Chen WH, Zhao XM. Long-read sequencing reveals extensive gut phageome structural variations driven by genetic exchange with bacterial hosts. Sci Adv. 2024; 10: eadn3316.
88. Slizovskiy IB, Oliva M, Settle JK, Zyskina LV, Prosperi M, Boucher C, et al. Target-enriched long-read sequencing (TELSeq) contextualizes antimicrobial resistance genes in metagenomes. Microbiome. 2022; 10: 185.
89. Zheng W, Zhao S, Yin Y, Zhang H, Needham DM, Evans ED, et al. High-throughput, single-microbe genomics with strain resolution, applied to a human gut microbiome. Science. 2022; 376: eabm1483.
90. Wu Y, Zhuang J, Song Y, Gao X, Chu J, Han S. Advances in single-cell sequencing technology in microbiome research. Genes Dis. 2024; 11: 101129.
91. Sarfatis A, Wang Y, Twumasi-Ankrah N, Moffitt JR. Highly multiplexed spatial transcriptomics in bacteria. Science. 2025; 387: eadr0932.
92. Zhang Y, Song W, Zhou L, Wei S, Xu L, Li X, et al. Bacteria-metabolite butyrate boosts anti-PD1 efficacy in colorectal cancer patient-derived organoids through activating autologous tumour-infiltrating *GNLY*$^{+}$CD8^{+} T cells. Gut. 2025; 74: 1755-58.
93. Ugolini GS, Wang M, Secchi E, Pioli R, Ackermann M, Stocker R. Microfluidic approaches in microbial ecology. Lab Chip. 2024; 24: 1394-418.

94. Asnicar F, Thomas AM, Passerini A, Waldron L, Segata N. Machine learning for microbiologists. Nat Rev Microbiol. 2024; 22: 191-205.
95. Zhou T, Zhao F. AI-empowered human microbiome research. Gut. 2025. DOI: 10.1136/gutjnl-2025-335946
96. Wang Y, Zhao L, Li Z, Xi Y, Pan Y, Zhao G, et al. A generative artificial intelligence approach for the discovery of antimicrobial peptides against multidrug-resistant bacteria. Nat Microbiol. 2025; 10: 2997-3012.
97. Novielli P, Romano D, Magarelli M, Bitonto PD, Diacono D, Chiatante A, et al. Explainable artificial intelligence for microbiome data analysis in colorectal cancer biomarker identification. Front Microbiol. 2024; 15: 1348974.
98. Rynazal R, Fujisawa K, Shiroma H, Salim F, Mizutani S, Shiba S, et al. Leveraging explainable AI for gut microbiome-based colorectal cancer classification. Genome Biol. 2023; 24: 21.
99. Caminero A, Tropini C, Valles-Colomer M, Shung DL, Gibbons SM, Surette MG, et al. Credible inferences in microbiome research: ensuring rigour, reproducibility and relevance in the era of AI. Nat Rev Gastroenterol Hepatol. 2025; 22: 788-803.
100. Ding X, Ting NL, Wong CC, Huang P, Jiang L, Liu C, et al. Bacteroides fragilis promotes chemoresistance in colorectal cancer, and its elimination by phage VA7 restores chemosensitivity. Cell Host Microbe. 2025; 33: 941-56.e10.
101. Kuntz TM, Gilbert JA. Introducing the microbiome into precision medicine. Trends Pharmacol Sci. 2017; 38: 81-91.
102. Lam KN, Alexander M, Turnbaugh PJ. Precision medicine goes microscopic: engineering the microbiome to improve drug outcomes. Cell Host Microbe. 2019; 26: 22-34.
103. Gurbatri CR, Arpaia N, Danino T. Engineering bacteria as interactive cancer therapies. Science. 2022; 378: 858-64.
104. Liu S, Rodriguez JS, Munteanu V, Ronkowski C, Sharma NK, Alser M, et al. Analysis of metagenomic data. Nat Rev Methods Primers. 2025; 5: 5.

Cite this article as: Lu Y, Yu J. Microbiota-host interaction in colorectal cancer: emerging computational technology, multi-omics integration, and mechanisms. Cancer Biol Med. 2026; 23: 580-595. doi: 10.20892/j.issn.2095-3941.2025.0762