Comprehensive analysis of transcriptome and microbiome in colorectal cancer with synchronous polyp patients

Yubin Wang[1†], Yongfeng Liu[2†], Xiaoqiang Liu[1], Pengwei Xu[2], Mingjie Luo[2], Anle Huang[3*] and Zhijun Su[4*]

[1]Department of Gastroenterology, Quanzhou First Hospital Affiliated to Fujian Medical University, Quanzhou, China, [2]Department of Scientific Research Cooperation GeneMind Biosciences Company Limited, Shenzhen, China, [3]Department of Gastrointestinal Oncology Surgery, The First Affiliated Hospital of Xiamen University, School of Medicine, Xiamen University, Xiamen, Fujian, China, [4]Department of Infectious disease, Quanzhou First Hospital Affiliated to Fujian Medical University, Quanzhou, China

Background: Colorectal cancer (CRC) is a prevalent and lethal malignancy, with the role of gut microbiota in its development still unclear. This study examines differences in gut microbiota between CRC patients and healthy controls and explores their association with host gene expression to identify potential diagnostic and therapeutic targets.

Methods: Fecal samples from 10 CRC patients and 13 healthy controls were subjected to 16S rRNA sequencing. Transcriptome sequencing of tumor tissues, normal mucosa, and colorectal polyps from same 10 CRC patients was performed to identify differentially expressed genes (DEGs). Pearson correlation analysis was employed to associate operational taxonomic units (OTUs) with host gene expression.

Results: β-diversity analysis showed significant differences in microbiota between CRC patients and controls (P < 0.01). LEfSe identified 38 distinct bacterial taxa, with genera such as *Bacteroides*, *Peptostreptococcus*, and *Parabacteroides* being enriched in CRC patients. Transcriptome analysis uncovered 1,026 DEGs. Notably, *TIMP1* and *BCAT1* were positively correlated (r > 0.76, P < 0.01) with pathogenic bacteria like *Fusobacterium nucleatum* and *Peptostreptococcus stomatis*. Tumor-related genes *TRPM4*, *MYBL2*, and *CDKN2A* were significantly upregulated and correlated with specific bacterial taxa.

Conclusion: This study underscores the significant alterations in gut microbiota associated with CRC and reveals novel correlations between specific microbes and host gene expression, offering potential diagnostic markers and therapeutic targets for CRC.

KEYWORDS

colorectal cancer, gut microbiota, 16S rRNA sequencing, OTU-gene correlation, diagnostic markers

Introduction

Colorectal cancer (CRC) ranks as the third most common cancer globally and is the second leading cause of cancer-related mortality. Data from the National Center for Health Statistics on Cancer for the year 2024 indicate that there are approximately 1.93 million new cases of CRC and 940,000 CRC-related deaths worldwide annually (Siegel et al., 2024). The incidence of synchronous polyps in CRC patients varies from 14% to 48% (Arenas et al., 1997), with a predilection for the right or proximal colon. These polyps often manifest as multiple adenomatous polyps, some of which may demonstrate high-grade dysplasia or possess malignant potential (Demetriades et al., 2004). The presence of synchronous polyps can complicate treatment, potentially requiring more extensive surgical interventions and signaling an increased risk for the development of subsequent metachronous lesions (Li et al., 2019). Therefore, preoperative and postoperative colonoscopic assessments are crucial for the detection and management of synchronous polyps, enabling optimized treatment strategies and improved patient outcomes (Piñol et al., 2004).

The gut microbiota, an essential component of the human gastrointestinal tract, plays a key role in maintenance of host immunity, metabolism, and barrier functions (Valdes et al., 2018; Tilg et al., 2020). Recent studies have demonstrated a strong correlation between gut microbiota dysbiosis and the development and progression of CRC (Wirbel et al., 2019; Wong and Yu, 2019; Yachida et al., 2019). Despite this, research exploring the interplay between the gut microbiota and host gene expression is relatively scarce (Gopalakrishnan et al., 2018). Prior investigations have predominantly compared microbiota and gene expression profiles between colorectal polyps and cancer by examining samples from distinct patient cohorts (Choi et al., 2023; Zou et al., 2024). However, there is a paucity of research that specifically addresses CRC patients with synchronous polyps, creating an obvious gap in our understanding of the pathological mechanisms specific to this subtype of CRC.

To bridge the identified research gap, we conducted a comprehensive analysis by integrating microbiome and transcriptome data. This approach allowed us to systematically evaluate the differences in gut microbiota composition between CRC patients presenting with synchronous polyps and a cohorts of healthy controls. Furthermore, we investigated the correlations between specific microbial taxa and differentially expressed genes (DEGs) to uncover the potential roles of the gut microbiota in CRC pathogenesis. Our study aims to offer fresh insights into the mechanisms specific to this distinct patient group and to pinpoint potential therapeutic targets and intervention strategies tailored to this population.

Methods

Sample collection and processing

In this study, tumor tissues (CC), adjacent normal mucosa (NM), and synchronous colorectal polyp tissues (PP) were collected from 10 patients diagnosed with both CRC and synchronous polyps for transcriptomic sequencing. All tissue samples were immediately flash-frozen in liquid nitrogen upon collection and stored at -80°C until RNA extraction. Additionally, fecal samples were collected from the same 10 CRC patients and 13 healthy controls for gut microbiome analysis. The inclusion criteria were: (1) age between 18-80 years; (2) CRC group: diagnosis of colorectal cancer confirmed by colonoscopy and pathology; control group: no abnormalities detected in colonoscopy; and (3) voluntary participation with signed informed consent. Exclusion criteria included: (1) familial colorectal cancer or familial polyposis; (2) history of diabetes; (3) use of antibiotics or probiotics in the past three months; (4) symptoms of infection within the last week; and (5) presence of other intestinal diseases. The study protocol was approved by the Ethics Committee of Quanzhou First Hospital (Ethics Approval Number: [2024] K189.

16S rRNA sequencing and analysis

Genomic DNA was meticulously extracted employing the CTAB/SDS protocol. The quantification and assessment of DNA purity were conducted using 1% agarose gel electrophoresis. Subsequently, the DNA was adjusted to a uniform concentration of 1 ng/µl using sterile water to standardize subsequent procedures. To construct sequencing libraries, the V3-V4 hypervariable regions of the 16S rRNA genes were selectively amplified. This was achieved using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, MA, USA), adhering closely to the manufacturer's guidelines. Unique index codes were incorporated to facilitate sample tracking. Post-quality control (QC), the libraries were subjected to high-throughput sequencing on an Illumina MiSeq platform, yielding 250 bp paired-end (PE) reads.

The sequences underwent stringent quality filtering using Trimmomatic V0.33 (Bolger et al., 2014), ensuring that only high-quality clean reads were retained for further analysis. Sequence assembly was performed using FLASH (v1.2.11), and the resulting contigs were clustered into operational taxonomic units (OTUs) at a 97% similarity threshold using the VSEARCH clustering algorithm (Rognes et al., 2016). For each OTU, a representative sequence was selected for taxonomic annotation. Taxonomic classification was conducted using QIIME2 (2019.4) with a confidence threshold of 0.7, leveraging the comprehensive Silva (Release 132) database for accurate species identification (Quast et al., 2013). Diversity indices, specifically the Shannon index, were calculated to assess alpha diversity for each sample. Beta diversity was analyzed using non-metric multidimensional scaling (NMDS) to visualize community structure differences between samples. To identify differentially abundant features, linear discriminant analysis effect size (LEfSe) was applied. The analysis was conducted using the Python LEfSe package, with a Wilcoxon p-value threshold set at 0.05 and a logarithmic LDA score cutoff at 2.0 to ensure statistical robustness. Finally, the results were visualized using bar plots and heatmaps generated with the R package v3.4.1, providing a clear and concise graphical representation of the data.

Transcriptome sequencing and analysis

Total RNA extraction from tumor tissue, normal mucosa, and colorectal polyps was conducted according to previously established protocols (Liu et al., 2021). Subsequently, RNA-sequencing (RNA-seq) libraries were meticulously prepared using the AHTS Universal V8 RNA-seq Library Prep Kit for Illumina (Vazyme, China), strictly adhering to the manufacturer's instructions. These libraries were then subjected to sequencing on the SURFSeq 5000 platform (GeneMind Biosciences LTD., China) employing a 150-cycle paired-end high-output sequencing protocol. Post-sequencing reads were meticulously filtered to remove adapters, poly-N sequences, and low-quality reads. The cleaned reads were subsequently aligned to the human reference genome (GRCh38/hg38) to ensure accurate mapping.

To explore the relationship between microbial community composition and host gene expression, the Pearson correlation coefficient was employed to assess the correlation between OTU abundance and differential gene expression levels for each OTU-gene pair across all 10 patients. To enhance computational efficiency and precision, OTUs present in fewer than 2 patients were excluded, resulting in a refined dataset of 298 OTUs that accounted for 98% of the initial abundance. DEG analysis was performed using DESeq2 to compare two groups per patient: normal mucosa versus colorectal tumor tissue (NM-VS-CC) and polyp tissue versus colorectal tumor tissue (PP-VS-CC) (Love et al., 2014). Genes were considered DEGs if the fold change ($|log2FC|$) between groups was >=1 and the differences were statistically significant (P-adjust <=0.05).

In correlating differential gene expression with OTU abundance, only the intersection genes of the two DEG sets were selected for analysis. This approach was taken to quantify the correlation between the host transcriptome and the microbiome. The Pearson's correlation was calculated using the corr.test function and the results were visualized using the Pheatmap package. The significance of each OTU-gene pair correlation was determined with a P-adjust threshold of <=0.05, ensuring that only robust and significant correlations were considered in the analysis.

Validation of key genes and gut microbiome

We validated the key genes and gut microbiome identified in our study using publicly available databases. Gene expression data were sourced from TCGA (The Cancer Genome Atlas https://www.cancer.gov/tcga) and GSE117606 from GEO (Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/geo) (https://doi.org/10.1093/nar/gks1193), while gut microbiome data were obtained from the GMrepo database (https://gmrepo.humangut.info/) (Dai et al., 2022).

Results

Patient cohort characteristics

A total of 10 patients diagnosed with both CRC and synchronous polyps were enrolled in this study, comprising 5 males and 5 females, with a mean age of 69.6 ± 9.3 years. Tumor samples were collected from the rectum, sigmoid colon, and ascending colon, and all patients were diagnosed with moderately differentiated adenocarcinoma. The cancer staging was distributed as follows: 1 patient in stage I, 5 in stage II, and 4 in stage IV. The healthy control group comprised 13 participants, including 10 males and 3 females, with a mean age of 47.6 ± 10.3 years (see Supplementary Table 1).

Differences in gut microbiota composition between CRC patients and healthy controls

In both CRC patients and healthy controls, four dominant bacterial phyla—Firmicutes, Bacteroidetes, Fusobacteria, and Proteobacteria—accounted for 94% to 98% of the OTUs on average (Figure 1A). Compared to the healthy controls, the relative abundance of Bacteroidetes and Fusobacteria was significantly increased in CRC patients, while the relative abundance of Firmicutes was significantly decreased (P < 0.05) (Figure 1B). Alpha diversity, evaluated using the Shannon diversity index, indicated a trend toward reduced gut microbial diversity in CRC patients, although this trend was not statistically significant (Figure 1C). Beta diversity analysis, conducted using NMDS, demonstrated a significant difference in microbial community composition between CRC patients and healthy controls (P < 0.01) (Figure 1D).

Identification of differentially abundant bacterial taxa

LEfSe analysis identified 38 significantly different taxa across five taxonomic levels (phylum, class, order, family, and genus) (Figure 2A). CRC patients showed significant enrichment of 30 taxa, including the genera *Bacteroides*, *Peptostreptococcus*, *Parabacteroides*, and the family Porphyromonadaceae. In contrast, healthy controls exhibited significant enrichment of 8 taxa, predominantly from the family Lachnospiraceae and genus *Blautia*. At the genus level, the Wilcoxon rank-sum test substantiated the differential abundance. The relative abundance of genera *Alistipes*, *Bacteroides*, and *Parabacteroides* was significantly higher in CRC patients (all P < 0.05), with *Bacteroides* and *Parabacteroides* showing highly significant
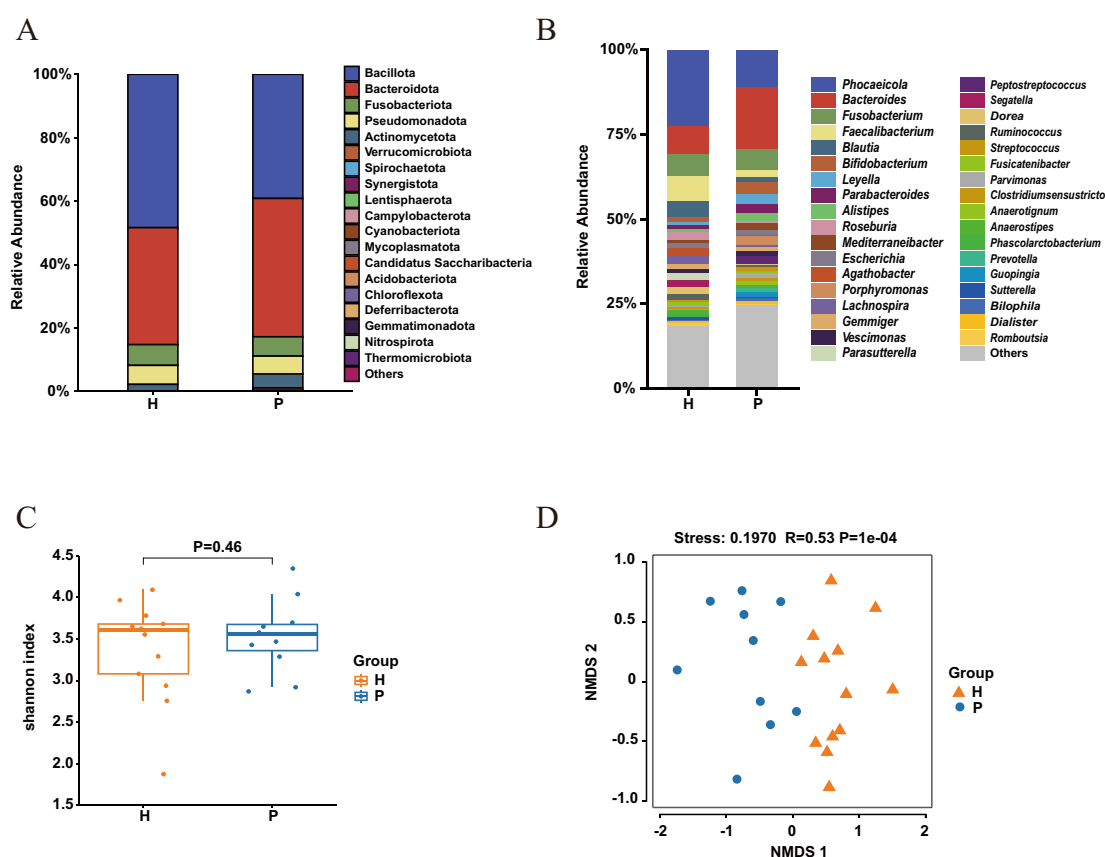
FIGURE 1

Gut bacterial composition and diversity in colorectal cancer (CRC) patients (P) and healthy controls (H). **(A)** Relative abundance at the phylum level and **(B)** genus level in CRC patients (blue) and healthy controls (yellow). **(C)** alpha-diversity assessed by the Shannon index for healthy controls (yellow) and CRC patients (blue), with p > 0.05 (T-test). **(D)** Non-metric multidimensional scaling (NMDS) of the 16S rRNA gene sequences from fecal samples in CRC patients (P) and healthy controls (H). The NMDS plot features a stress value of 0.1970 and a correlation coefficient (R) of 0.53, calculated using Bray-Curtis distances, with a highly significant P-value of 1e-04.

differences ($P < 0.01$) (Figure 2B). In the healthy controls, the relative abundance of genera *Blautia*, *Faecalibacterium*, *Phocaeicola*, and *Roseburia* was significantly elevated (all $P < 0.05$).

Differential gene expression across CRC tissues, normal mucosa, and polyp tissues

Transcriptome sequencing was performed on CC, NM, and PP samples from 10 CRC patients. The 2D PCA plot illustrates the gene expression disparities among CC, NM, and PP groups (Figure 3A). Notably, CC group exhibited a markedly distinct distribution in the principal component space when compared to both NM and PP groups, suggesting that the transcriptomic profiles of CC divergent from those of NM and PP. In the comparison between CC and NM, a total of 3,501 DEGs were identified, while 1,099 DEGs were identified between CC and PP (Figure 3B). After excluding genes unique to each comparison, 1,026 common DEGs were selected for further analysis.

GO analysis revealed a significant enrichment of genes related to bacterial molecular responses in CC tissue (Figures 3C, D). This finding suggests that the gut microbiota may play a crucial role in

cancer tissues, with distinct differences in bacterial-related responses between cancerous and non-cancerous tissues. These findings bolster our understanding of the potential role of the gut microbiome in CRC pathogenesis and provides new insights for exploring the interplay between the gut microbiota and cancer development.

Correlation analysis between gut microbiota and host gene expression

This study conducted Pearson correlation analysis between 1,026 DEGs and 298 OTUs, retaining data with adjusted P values (P-adjust) less than 0.05 for network visualization (Figure 4A). The network plot provided a clear view of the interactions between OTUs and genes, identifying 64 significant OTU-gene pairs, involving 29 mRNAs and 7 OTUs (Figure 4B). Within this set, 17 pairs had P-adjust <0.05, 17 pairs had P-adjust < 0.01, and 30 pairs had P-adjust < 0.001, with the majority of correlations being positive.

The study found that the genes tissue inhibitor of metalloproteinases 1 (*TIMP1*) and branched chain
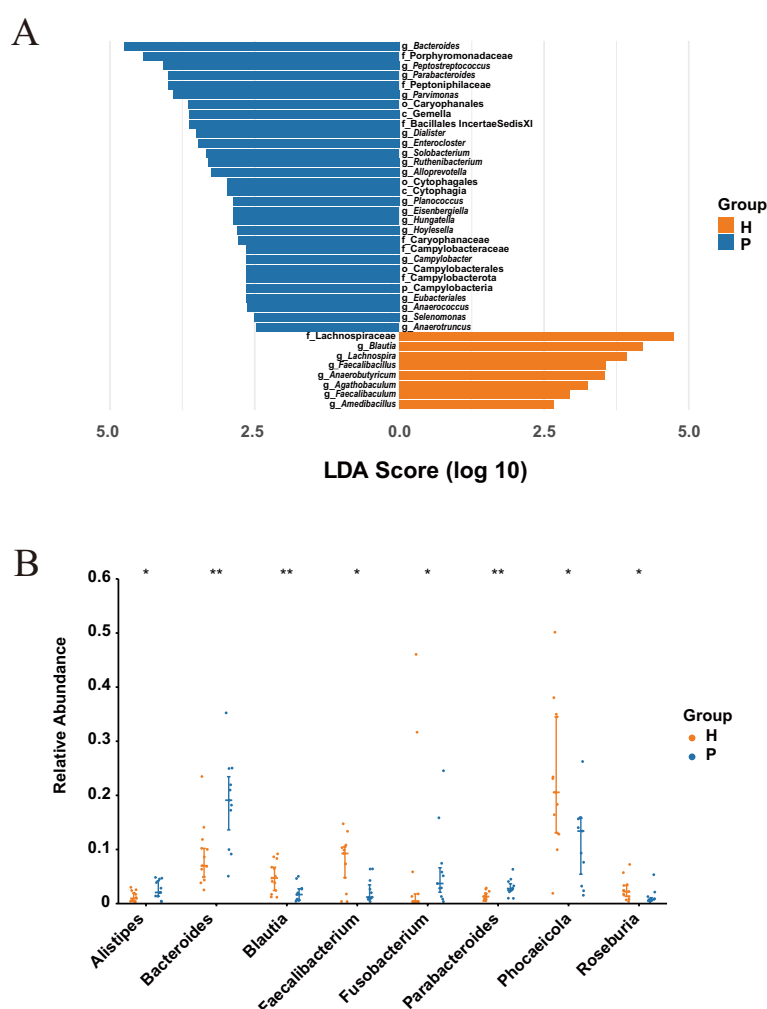
FIGURE 2
Differential abundance of microbiota between CRC patients (P) and healthy controls (H). **(A)** LEfSe analysis with an LDA score > 2 and P-value < 0.01 (Wilcoxon rank-sum test). **(B)** The top 8 differential genera in CRC patients (P) and healthy controls (H) with P-value < 0.05 (Wilcoxon rank-sum test). * indicates P-value < 0.05, ** indicates P-value < 0.01.

aminotransferase 1 (*BCAT1*) were significantly positively correlated (correlation coefficient r > 0.76, P < 0.01) with pathogenic bacteria such as *Fusobacterium nucleatum*, *Peptostreptococcus stomatis*, and *Parvimonas micra*. Additionally, the gene transient receptor potential cation channel subfamily M member 4 (*TRPM4*) showed significant correlations with multiple pathogenic bacteria. Other genes, including MYB proto-oncogene like 2 (*MYBL2*), cyclin dependent kinase inhibitor 2A (*CDKN2A*), and stathmin 3 (*STMN3*) were significantly upregulated in tumor tissues from CRC patients and were significantly associated with specific bacterial genera (Figure 4B).

Validation of key gene expression and gut microbiome

The results show that among the eight bacteria, *Bacteroides fragilis* exhibits a particularly strong association with colon

adenomas. Notably, the relative abundance of *Porphyromonas asaccharolytica* and *Porphyromonas somerae* is the highest among the analyzed species (Supplementary Figure S1).

Expression analysis of five key genes in CC, NM, and PP revealed that BCAT1, *MYBL2*, *CDKN2A*, and *TIMP1* were significantly upregulated in CC compared to NM and PP (all P < 0.05). Conversely, *TRPM4* was significantly downregulated in CC relative to NM and PP (P < 0.05). Analysis of TCGA database showed that *CDKN2A*, *MYBL2*, and *TIMP1* were significantly upregulated in cancer tissues, while *TRPM4* was significantly downregulated; *BCAT1* expression did not show a significant difference (Figure 5). GEO database analysis indicated that *TIMP1* and *BCAT1* were significantly upregulated in cancer tissues, whereas *TRPM4* was significantly downregulated, with no significant differences in *MYBL2* and *CDKN2A* expression. Additionally, TCGA data revealed that high expression of *CDKN2A* and *TIMP1* were significantly associated with poorer survival outcomes (P = 0.036 and P = 0.0045, respectively). In
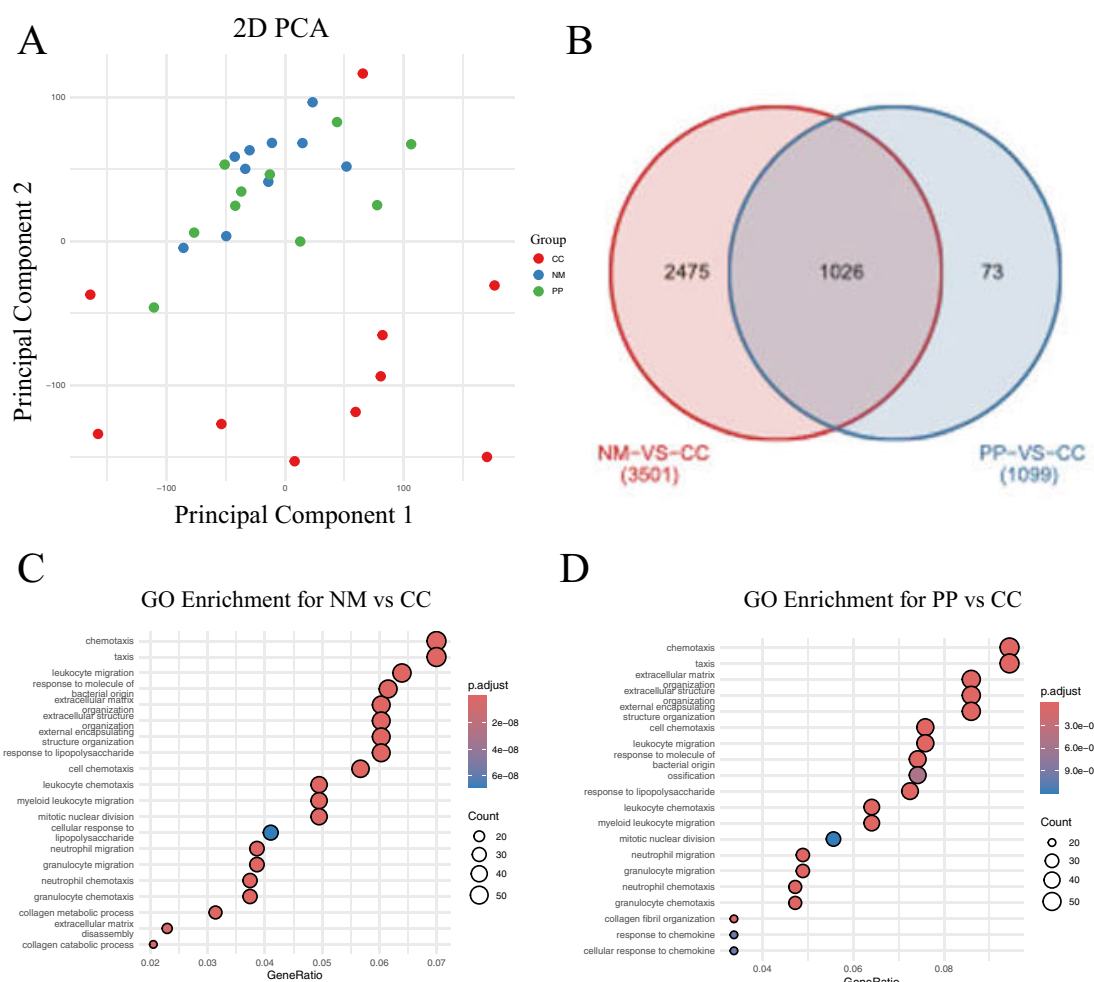
FIGURE 3

Gene expression analysis and GO enrichment in two comparisons. **(A)** The 2D plot shows the distribution of samples based on gene expression data. Points are colored to represent different groups: red for tumor, green for polyp, and blue for normal tissue. **(B)**. Venn diagram displaying the differentially expressed genes (DEGs) between the NM vs CC and PP vs CC comparisons. The overlapping region represents the common genes shared between both comparisons: tumor tissue (denoted as CC) versus normal mucosa (denoted as NM), and tumor tissue to colorectal polyps (denoted as PP). **(C)**. Top 20 GO enrichment for the NM vs CC comparison. **(D)**. Top 20 GO enrichment for the PP vs CC comparison. CC: tumor tissue, NM: normal mucosa, PP: colorectal polyps.

contrast, the expression of *TRPM4*, *BCAT1*, and *MYBL2* was not significantly correlated with survival (all P-values > 0.05).

Discussion

In this study, we conducted a comprehensive analysis of patients with CRC accompanied by synchronous polyps, revealing alterations in gut microbiota composition and host gene expression, and their interrelationships. To our knowledge, this is the first study to collect CC, NM, and PP from the same patient, thereby eliminating inter-individual variability and more precisely reflecting the biological changes during disease progression. This approach contrasts sharply with previous studies that treated CRC patients, polyp patients, and healthy controls as separate groups, providing fresh insights into the mechanisms underlying CRC initiation and progression.

Our findings demonstrated significant changes in the gut microbiota composition of CRC patients with synchronous polyps, characterized by an increase in pathogenic bacteria and a decrease in beneficial bacteria. This result is consistent with previous studies, further validating the crucial role of gut microbiota in the pathogenesis of CRC (Liu et al., 2020; Silva et al., 2021). For instance, we observed a significant increase in the abundance of phyla the Bacteroidetes and Fusobacteria, and a significant decrease in Firmicutes in these patients. This aligns with previous findings, which reported decreased gut microbial diversity and dysbiosis in CRC patients (Yeoh et al., 2020; Shariati et al., 2021; Cui et al., 2022). However, regarding α-diversity, our study showed a decreasing trend in gut microbial diversity in CRC patients with synchronous polyps, but it did not reach statistical significance. This finding contrasts with other studies that reported a significant reduction in α-diversity among CRC patients (Gataullin et al., 2021; Zhao et al., 2021), which reported a
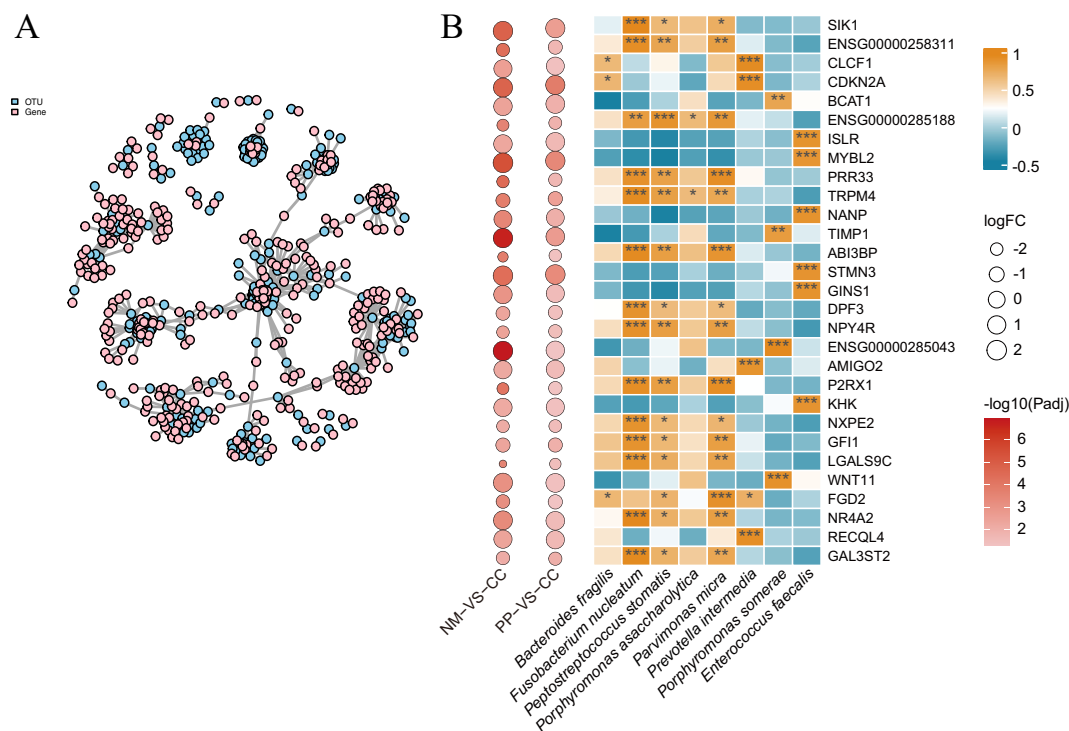
FIGURE 4
Correlation network and pearson correlation analysis. **(A)**. This network plot illustrates the correlations between OTUs and genes. The nodes in the plot represent OTUs and gene names, while the edges indicate the strength of the correlation, with edge thickness reflecting the magnitude of the correlation. OTU nodes are colored sky blue, and gene name nodes are colored pink. **(B)**. Pearson correlation coefficients of 100 OTU-gene pairs (P-adjust < 0.05). The intensity of the colors represents the degree of correlation; correlations were considered significant at P-adjust < 0.05. P-adjust < 0.001, P-adjust < 0.01, and P-adjust < 0.05 were labeled with "***", "**" and "*" , respectively. The left panels denote the fold changes and P-adjust of DEGs in two groups (NM vs CC and PP vs CC). CC: tumor tissue, NM: normal mucosa, PP: colorectal polyps.

significant reduction in α-diversity in CRC patients. The discrepancy may be attributed to our relatively small sample size, which could reduce the power of statistical tests.

Additionally, we confirmed the enrichment of *Fusobacterium nucleatum* in CRC patients with synchronous polyps, a bacterium associated with tumor invasiveness and poor prognosis, supporting its pivotal role in CRC progression (Xu et al., 2021a). *F. nucleatum* interacts with host E-cadherin through its adhesin FadA, activating the β-catenin signaling pathway and upregulating the oncogene *MYC* (Rubinstein et al., 2019). Additionally, it interacts with the TIGIT receptor on T cells via its surface protein Fap2, inhibiting the antitumor activities of natural killer cells and T cells, leading to immune evasion (Galaski et al., 2021). *F. nucleatum* can also induce the release of pro-inflammatory cytokines such as IL-6 and TNF-α, creating an inflammatory microenvironment that promotes tumor development (Bostanghadiri et al., 2023). Other pathogenic bacteria, such as *Parvimonas micra*, *Prevotella intermedia*, *Bacteroides fragilis*, and *Enterococcus faecalis*, are also believed to promote CRC through various mechanisms, consistent with previous studies (Lavoie et al., 2020; Kabwe et al., 2021; Osman et al., 2021; Lee et al., 2022). In contrast, beneficial bacteria enriched in the healthy control group, such as *Blautia*, *Faecalibacterium*, and *Roseburia*, typically exhibit anti-inflammatory effects and maintain intestinal barrier function (Mohebali et al., 2020; Xia et al., 2021).

Reducing these beneficial microbes may lead to an imbalance in the intestinal microenvironment, thereby promoting tumorigenesis.

Abnormal host gene expression was also observed. Transcriptome analysis revealed that genes such as *TIMP1*, *BCAT1*, and *MYBL2* were significantly upregulated in CRC tissues compared to normal mucosa and polyp tissues. Previous study has confirmed that overexpression of *BCAT1* in CRC is associated with tumor progression and poor prognosis (Symonds et al., 2022). Our study further emphasizes the critical role of this gene in CRC. *BCAT1*, as a key enzyme in branched-chain amino acid metabolism, promotes tumor cell proliferation (Mao et al., 2021). The methylation status of *BCAT1* has also been used as a biomarker for non-invasive CRC diagnosis, highlighting its potential in early tumor detection (Xu et al., 2021b). Conversely, we observed that *TRPM4* was significantly downregulated in cancer tissues compared to normal mucosa and polyp tissues. This finding is consistent with analyses from TCGA and GEO databases, suggesting that TRPM4 may act as a tumor suppressor in CRC. However, a previous study indicated that *TRPM4* is highly expressed in tumor buds of human colorectal tumors and is associated with proliferation, cell cycle regulation, and invasion of colorectal cancer cells (Kappel et al., 2019). This contradiction may stem from differences in sample types, detection methods, or patient populations, warranting further investigation in future studies.
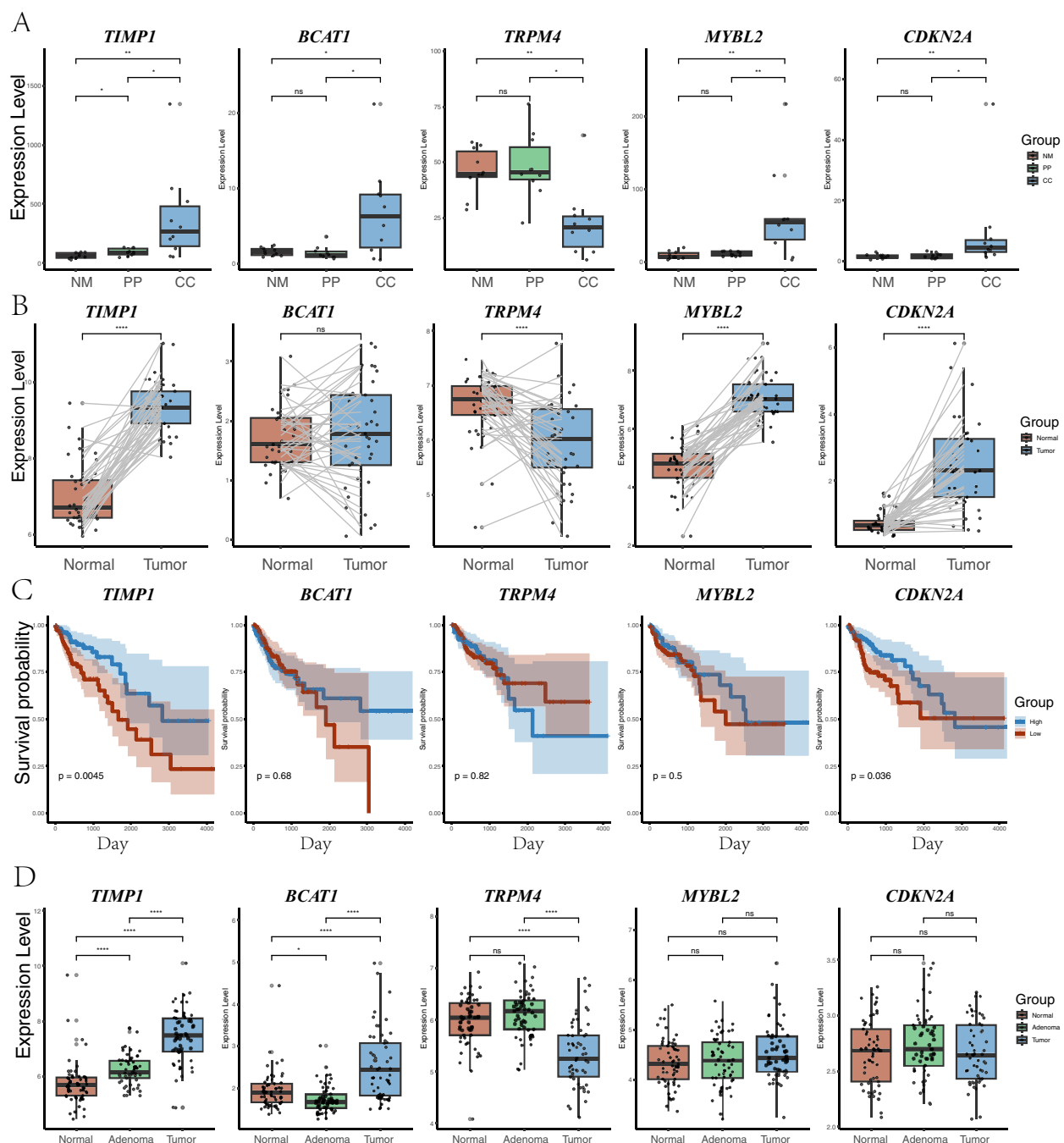
FIGURE 5
Gene expression analysis across different groups. **(A)** Expression levels of five genes (*TIMP1*, *BCAT1*, *TRPM4*, *MYBL2*, and *CDKN2A*) across three groups (CC: tumor tissue, NM: normal mucosa, PP: colorectal polyps). Statistical significance between the groups is indicated by P-value. P-value < 0.001, P-value < 0.001, P-value < 0.01, and P-value < 0.05 were labeled with "****", "***", "**" and "*" , respectively. **(B)** Distribution of the same five genes in TCGA paired samples between normal and tumor tissues, highlighting changes in gene expression between matched sample sets. **(C)** Survival analysis based on the expression levels of the five genes in TCGA samples. Kaplan-Meier plots show the relationship between high and low expression of each gene and overall survival probability. **(D)** Expression levels of the five genes (*TIMP1*, *BCAT1*, *TRPM4*, *MYBL2*, and *CDKN2A*) in the GSE117606 dataset, across normal, adenoma, and tumor groups.

Furthermore, we found a significant correlation between the expression of *TIMP1* and the abundance of *Porphyromonas somerae*. TIMP1, as an endogenous inhibitor of matrix metalloproteinases, participates in the degradation and remodeling of the extracellular matrix (Guccini et al., 2021). TIMP1 predicts colon cancer progression and metastasis through the FAK-PI3K/AKT and MAPK pathways (Song et al., 2016). Our results show that *TIMP1* expression in CRC tissues is higher than in polyp and normal mucosa groups, validated in TCGA and GEO databases, and that high *TIMP1* expression is significantly associated with poorer survival outcomes.

Functional enrichment analysis of high *TIMP1* expression suggests that *TIMP1* may influence tumor invasion and metastasis by regulating ECM remodeling, cell adhesion, cell migration, and related signaling pathways (such as ECM-receptor interaction and cytoskeleton regulation). *P. somerae*, a Gram-negative anaerobic bacterium belonging to the *Porphyromonas* genus, is commonly found in the oral microbiota and has been associated with various human diseases, particularly periodontitis and systemic inflammation (Lundmark et al., 2019). Study has found that in lean CRC, *P. somerae* is considered a characteristic microbe and is associated with functional metabolic pathways such as mucin O-glycan biosynthesis, glycosaminoglycan degradation, and butyrate metabolism (Zhu et al., 2024). Although direct studies linking *P. somerae* to CRC are relatively scarce, other species within the genus, such as *Porphyromonas gingivalis*, have been extensively studied and reported to play important roles in CRC pathogenesis (Wang et al., 2021; Kerdreux et al., 2023). *P. somerae* shares genomic similarities with *P. gingivalis*, including genes involved in energy metabolism and intracellular survival, suggesting that *P. somerae* may possess similar pathogenic mechanisms (Crooks et al., 2021). The gingipain virulence factors of *P. gingivalis* promote cell proliferation, inhibit apoptosis, and accelerate inflammatory responses by regulating the PI3K/AKT and MAPK signaling pathways, which are also involved in TIMP1-mediated colon cancer progression and metastasis (Liu et al., 2018). Additionally, gingipains can activate pro-inflammatory factors such as interleukins and tumor necrosis factors, further promoting tumor cell growth and invasion (Kerdreux et al., 2023). Regarding the impact of *P. gingivalis* on *TIMP1* expression, studies have shown that *P. gingivalis* infection can enhance the secretion of MMP-1 and TIMP1 in human periodontal ligament fibroblasts, leading to an imbalance in the MMP-1/TIMP1 ratio and further promoting tissue destruction (Herath et al., 2013). Based on the aforementioned studies on *P. gingivalis*, we speculate that *P. somerae* may influence CRC occurrence and progression through similar mechanisms. This provides a novel perspective for considering *P. somerae* or its related inflammatory pathways as potential therapeutic targets for CRC. However, further research is necessary to elucidate the specific molecular mechanisms by which *P. somerae* affects *TIMP1* expression and to verify its role in CRC pathogenesis.

Additionally, increasing evidence suggests that *P. somerae* can modulate other key genes via metabolic pathways. For instance, Crooks demonstrated in an endometrial cancer model that *P. somerae* can invade epithelial cells, evade immune recognition, and secrete metabolic intermediates such as succinate, which stabilize hypoxia-inducible factors (HIF) (Crooks et al., 2021). Through this process, *P. somerae* promotes carcinogenesis by inducing chronic inflammation and enhancing angiogenesis. With respect to *BCAT1*, branched-chain aminotransferases (*BCAT*) are pivotal in the catabolism of branched-chain amino acids (BCAAs), and *BCAT1* is the predominant isoform in human primary macrophages (Papathanassiu et al., 2017). Notably, *P. somerae* often exhibits esterase, esterase lipase, leucine arylamidase, and valine arylamidase activities (Summanen et al., 2005), the latter two being able to hydrolyze aromatic amine compounds containing leucine or valine, thereby supplying additional leucine or valine substrates to the host.

Under these conditions, the demand for *BCAT* by host cells likely increases, potentially resulting in *BCAT1* upregulation to facilitate BCAA transamination. Hence, we hypothesize that *P. somerae* indirectly induces *BCAT1* overexpression by altering the metabolic milieu or providing surplus substrates, in line with our observed positive correlation between *P. somerae* abundance and *BCAT1* expression. Further *in vitro* and *in vivo* investigations are warranted to validate this hypothesis and elucidate the precise molecular mechanisms involved.

Our study's findings have potential clinical significance. First, our innovative study design, comparing different pathological tissues within the same patient, avoids inter-individual differences and more accurately reflects the biological changes during disease progression. Second, detecting specific pathogenic bacteria such as *P. somerae* may serve as biomarkers for early diagnosis or risk assessment of CRC. Interventions targeting these pathogens, or their associated pathways could provide new strategies for CRC prevention and treatment. For example, probiotic therapies that modulate the gut microbiota or antimicrobial treatments targeting specific bacterial populations may influence host gene expression and slow tumor progression. Finally, expression levels of genes like *TIMP1* could serve as reference indicators for prognostic evaluation and personalized treatment.

Despite providing important insights, our study has several limitations. First, the sample size is relatively small, including only 10 CRC patients, which may limit the generalizability of the results. However, our findings are consistent with mainstream research, supporting the reliability of our conclusions. Second, the study primarily focuses on correlational analyses and lacks in-depth exploration of the causal relationships between microbiota and gene expression. Future studies should incorporate *in vivo* and *in vitro* experiments to investigate the specific mechanisms by which particular microbes affect host gene expression. Additionally, due to the absence of longitudinal data, we could not assess the dynamic changes of microbiota and gene expression during disease progression. Future longitudinal studies will help elucidate the roles of these changes in CRC development and progression.

Conclusion

Through a pioneering study design, our research conducted an in-depth analysis of the alteration in gut microbiota and host gene expression in CRC patients with synchronous polyps. We revealed, for the first time, a significant association between the abundance of *P. somerae* and the high expression of *TIMP1*. These findings shed new light on the mechanisms underlying the onset and progression of CRC and suggest potential avenues for early diagnosis, prevention, and the development of personalized treatment strategies.

Data availability statement

Publicly available datasets were utilized in this study as follows: Gene expression data were obtained from the TCGA (The Cancer

Genome Atlas, https://www.cancer.gov/tcga) and the GSE117606 dataset from GEO (Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/geo, https://doi.org/10.1093/nar/gks1193). Gut microbiome reference data were accessed from the GMrepo database (https://gmrepo.humangut.info/). The data generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Ethics statement

The studies involving humans were approved by Quanzhou First Hospital Ethics Committee. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

Funding

Conflict of interest

Authors YL, PX, ML was/were employed by GeneMind Biosciences Company Limited.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2025.1547057/full#supplementary-material

References

Arenas, R. B., Fichera, A., Mhoon, D., and Michelassi, F. (1997). Incidence and therapeutic implications of synchronous colonic pathology in colorectal adenocarcinoma. *Surgery* 122, 706–710. doi: 10.1016/S0039-6060(97)90077-5

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bostanghadiri, N., Razavi, S., Shariati, A., Talebi, M., Mirkalantari, S., Emami Razavi, A., et al. (2023). Exploring the interplay between Fusobacterium nucleatum with the expression of microRNA, and inflammatory mediators in colorectal cancer. *Front. Microbiol.* 14. doi: 10.3389/fmicb.2023.1302719

Choi, J. W., Lee, G.-Y., Kim, S., Ahn, K., Do, I.-G., Jung, K.-U., et al. (2023). RNA-seq-based molecular classification analyses in colorectal cancer and synchronous adenoma. *Cancers* 15, 4851. doi: 10.3390/cancers15194851

Crooks, T. A., Madison, J. D., Walsh, D. M., Herbert, W. G., Jeraldo, P. R., Chia, N., et al. (2021). Porphyromonas somerae invasion of endometrial cancer cells. *Front. Microbiol.* 12. doi: 10.3389/fmicb.2021.674835

Cui, Y., Zhang, L., Wang, X., Yi, Y., Shan, Y., Liu, B., et al. (2022). Roles of intestinal *Parabacteroides* in human health and diseases. *FEMS Microbiol. Lett.* 369, fnac072. doi: 10.1093/femsle/fnac072

Dai, D., Zhu, J., Sun, C., Li, M., Liu, J., Wu, S., et al. (2022). GMrepo v2: a curated human gut microbiome database with special focus on disease markers and cross-dataset comparison. *Nucleic Acids Res.* 50, D777–D784. doi: 10.1093/nar/gkab1019

Demetriades, H., Kanellos, I., Blouhos, K., Tsachalis, T., Vasiliadis, K., Pramateftakis, M. G., et al. (2004). Synchronous polyps in patients with colorectal cancer. *Tech Coloproctol* 8, s72–s75. doi: 10.1007/s10151-004-0117-x

Galaski, J., Shhadeh, A., Umaña, A., Yoo, C. C., Arpinati, L., Isaacson, B., et al. (2021). Fusobacterium nucleatum cbpF mediates inhibition of T cell function through CEACAM1 activation. *Front. Cell. Infect. Microbiol.* 11. doi: 10.3389/fcimb.2021.692544

Gataullin, B. I., Gataullin, I. G., Nga, N. T., Kolpakov, A. I., and Ilinskaya, O. N. (2021). Analysis of the intestinal microbiome in colorectal cancer. *Kazan Med. J.* 102, 185–191. doi: 10.17816/KMJ2021-185

Gopalakrishnan, V., Helmink, B. A., Spencer, C. N., Reuben, A., and Wargo, J. A. (2018). The influence of the gut microbiome on cancer, immunity, and cancer immunotherapy. *Cancer Cell* 33, 570–580. doi: 10.1016/j.ccell.2018.03.015

Guccini, I., Revandkar, A., D'Ambrosio, M., Colucci, M., Pasquini, E., Mosole, S., et al. (2021). Senescence reprogramming by TIMP1 deficiency promotes prostate cancer metastasis. *Cancer Cell* 39, 68–82.e9. doi: 10.1016/j.ccell.2020.10.012

Herath, T. D. K., Wang, Y., Seneviratne, C. J., Darveau, R. P., Wang, C.-Y., and Jin, L. (2013). The expression and regulation of matrix metalloproteinase-3 is critically modulated by Porphyromonas gingivalis lipopolysaccharide with heterogeneous lipid A structures in human gingival fibroblasts. *BMC Microbiol.* 13, 73. doi: 10.1186/1471-2180-13-73

Kabwe, M., Meehan-Andrews, T., Ku, H., Petrovski, S., Batinovic, S., Chan, H. T., et al. (2021). Lytic bacteriophage EFA1 modulates HCT116 colon cancer cell growth and upregulates ROS production in an enterococcus faecalis co-culture system. *Front. Microbiol.* 12. doi: 10.3389/fmicb.2021.650849

Kappel, S., Stokłosa, P., Hauert, B., Ross-Kaschitza, D., Borgström, A., Baur, R., et al. (2019). TRPM4 is highly expressed in human colorectal tumor buds and contributes to proliferation, cell cycle, and invasion of colorectal cancer cells. *Mol. Oncol.* 13, 2393–2405. doi: 10.1002/1878-0261.12566

Kerdreux, M., Edin, S., Löwenmark, T., Bronnec, V., Löfgren-Burström, A., Zingmark, C., et al. (2023). *Porphyromonas gingivalis* in Colorectal Cancer and its Association to Patient Prognosis. *J. Cancer* 14, 1479–1485. doi: 10.7150/jca.83395

Lavoie, S., Chun, E., Bae, S., Brennan, C. A., Gallini Comeau, C. A., Lang, J. K., et al. (2020). Expression of free fatty acid receptor 2 by dendritic cells prevents their expression of interleukin 27 and is required for maintenance of mucosal barrier and immune response against colorectal tumors in mice. *Gastroenterology* 158, 1359–1372.e9. doi: 10.1053/j.gastro.2019.12.027

Lee, C.-G., Hwang, S., Gwon, S.-Y., Park, C., Jo, M., Hong, J.-E., et al. (2022). Bacteroides fragilis toxin induces intestinal epithelial cell secretion of interleukin-8 by the E-cadherin/β-catenin/NF-κB dependent pathway. *Biomedicines* 10, 827. doi: 10.3390/biomedicines10040827

Li, S., Zhu, K., Yu, W., Wang, Y., Wang, T., Guo, S., et al. (2019). Synchronous neoplastic lesions in referred patients with colorectal cancer: A retrospective cohort study. *CMAR* 11, 9951–9959. doi: 10.2147/CMAR.S229376

Liu, J., Wang, X., Zheng, M., and Luan, Q. (2018). Lipopolysaccharide from Porphyromonas gingivalis promotes autophagy of human gingival fibroblasts through the PI3K/Akt/mTOR signaling pathway. *Life Sci.* 211, 133–139. doi: 10.1016/j.lfs.2018.09.023

Liu, W., Zhang, R., Shu, R., Yu, J., Li, H., Long, H., et al. (2020). Study of the relationship between microbiome and colorectal cancer susceptibility using 16SrRNA sequencing. *BioMed. Res. Int.* 2020, 1–17. doi: 10.1155/2020/7828392

Liu, Y., Han, R., Zhou, L., Luo, M., Zeng, L., Zhao, X., et al. (2021). Comparative performance of the GenoLab M and NovaSeq 6000 sequencing platforms for transcriptome and LncRNA analysis. *BMC Genomics* 22, 829. doi: 10.1186/s12864-021-08150-8

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Lundmark, A., Hu, Y. O. O., Huss, M., Johannsen, G., Andersson, A. F., and Yucel-Lindberg, T. (2019). Identification of salivary microbiota and its association with host inflammatory mediators in periodontitis. *Front. Cell. Infect. Microbiol.* 9. doi: 10.3389/fcimb.2019.00216

Mao, L., Chen, J., Lu, X., Yang, C., Ding, Y., Wang, M., et al. (2021). Proteomic analysis of lung cancer cells reveals a critical role of BCAT1 in cancer cell metastasis. *Theranostics* 11, 9705–9720. doi: 10.7150/thno.61731

Mohebali, N., Ekat, K., Kreikemeyer, B., and Breitrück, A. (2020). Barrier protection and recovery effects of gut commensal bacteria on differentiated intestinal epithelial cells *in vitro*. *Nutrients* 12, 2251. doi: 10.3390/nu12082251

Osman, M. A., Neoh, H., Ab Mutalib, N.-S., Chin, S.-F., Mazlan, L., Raja Ali, R. A., et al. (2021). Parvimonas micra, Peptostreptococcus stomatis, Fusobacterium nucleatum and Akkermansia muciniphila as a four-bacteria biomarker panel of colorectal cancer. *Sci. Rep.* 11, 2925. doi: 10.1038/s41598-021-82465-0

Papathanassiu, A. E., Ko, J.-H., Imprialou, M., Bagnati, M., Srivastava, P. K., Vu, H. A., et al. (2017). BCAT1 controls metabolic reprogramming in activated human macrophages and is associated with inflammatory diseases. *Nat. Commun.* 8, 16040. doi: 10.1038/ncomms16040

Piñol, V., Andreu, M., Castells, A., Payá, A., Bessa, X., and Jover, R. (2004). Synchronous colorectal neoplasms in patients with colorectal cancer: predisposing individual and familial factors. *Dis. Colon Rectum* 47, 1192–1200. doi: 10.1007/s10350-004-0562-7

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4, e2584. doi: 10.7717/peerj.2584

Rubinstein, M. R., Baik, J. E., Lagana, S. M., Han, R. P., Raab, W. J., Sahoo, D., et al. (2019). *Fusobacterium nucleatum* promotes colorectal cancer by inducing Wnt/β-catenin modulator Annexin A1. *EMBO Rep.* 20, e47638. doi: 10.15252/embr.201847638

Shariati, A., Razavi, S., Ghaznavi-Rad, E., Jahanbin, B., Akbari, A., Norzaee, S., et al. (2021). Association between colorectal cancer and Fusobacterium nucleatum and Bacteroides fragilis bacteria in Iranian patients: a preliminary study. *Infect. Agents Cancer* 16, 41. doi: 10.1186/s13027-021-00381-4

Siegel, R. L., Giaquinto, A. N., and Jemal, A. (2024). Cancer statistics 2024. *CA A Cancer J. Clin.* 74, 12–49. doi: 10.3322/caac.21820

Silva, M., Brunner, V., and Tschurtschenthaler, M. (2021). Microbiota and colorectal cancer: from gut to bedside. *Front. Pharmacol.* 12. doi: 10.3389/fphar.2021.760280

Song, G., Xu, S., Zhang, H., Wang, Y., Xiao, C., Jiang, T., et al. (2016). TIMP1 is a prognostic marker for the progression and metastasis of colon cancer through FAK-PI3K/AKT and MAPK pathway. *J. Exp. Clin. Cancer Res.* 35, 148. doi: 10.1186/s13046-016-0427-7

Summanen, P. H., Durmaz, B., Väisänen, M.-L., Liu, C., Molitoris, D., Eerola, E., et al. (2005). *Porphyromonas somerae* sp. nov., a Pathogen Isolated from Humans and Distinct from *Porphyromonas levii*. *J. Clin. Microbiol.* 43, 4455–4459. doi: 10.1128/JCM.43.9.4455-4459.2005

Symonds, E. L., Pedersen, S. K., Yeo, B., Al Naji, H., Byrne, S. E., Roy, A., et al. (2022). Assessment of tumor burden and response to therapy in patients with colorectal cancer using a quantitative ctDNA test for methylated *BCAT1/IKZF1*. *Mol. Oncol.* 16, 2031–2041. doi: 10.1002/1878-0261.13178

Tilg, H., Zmora, N., Adolph, T. E., and Elinav, E. (2020). The intestinal microbiota fuelling metabolic inflammation. *Nat. Rev. Immunol.* 20, 40–54. doi: 10.1038/s41577-019-0198-4

Valdes, A. M., Walter, J., Segal, E., and Spector, T. D. (2018). Role of the gut microbiota in nutrition and health. *BMJ* 361, k2179. doi: 10.1136/bmj.k2179

Wang, X., Jia, Y., Wen, L., Mu, W., Wu, X., Liu, T., et al. (2021). *Porphyromonas gingivalis* promotes colorectal carcinoma by activating the hematopoietic *NLRP3* inflammasome. *Cancer Res.* 81, 2745–2759. doi: 10.1158/0008-5472.CAN-20-3827

Wirbel, J., Pyl, P. T., Kartal, E., Zych, K., Kashani, A., Milanese, A., et al. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat. Med.* 25, 679–689. doi: 10.1038/s41591-019-0406-6

Wong, S. H., and Yu, J. (2019). Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nat. Rev. Gastroenterol. Hepatol.* 16, 690–704. doi: 10.1038/s41575-019-0209-8

Xia, X., Ni, J., Yin, S., Yang, Z., Jiang, H., Wang, C., et al. (2021). Elevated systemic and intestinal inflammatory response are associated with gut microbiome disorder after cardiovascular surgery. *Front. Microbiol.* 12. doi: 10.3389/fmicb.2021.686648

Xu, C., Fan, L., Lin, Y., Shen, W., Qi, Y., Zhang, Y., et al. (2021a). *Fusobacterium nucleatum* promotes colorectal cancer metastasis through miR-1322/CCL20 axis and M2 polarization. *Gut Microbes* 13, 1980347. doi: 10.1080/19490976.2021.1980347

Xu, C., Yu, S., Han, J., Zong, M., Tan, Q., Zeng, X., et al. (2021b). Detection of circulating tumor DNA methylation in diagnosis of colorectal cancer. *Clin. Transl. Gastroenterol.* 12, e00386. doi: 10.14309/ctg.0000000000000386

Yachida, S., Mizutani, S., Shiroma, H., Shiba, S., Nakajima, T., Sakamoto, T., et al. (2019). Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat. Med.* 25, 968–976. doi: 10.1038/s41591-019-0458-2

Yeoh, Y. K., Chen, Z., Wong, M. C. S., Hui, M., Yu, J., Ng, S. C., et al. (2020). Southern Chinese populations harbour non-nucleatum Fusobacteria possessing homologues of the colorectal cancer-associated FadA virulence factor. *Gut* 69, 1998–2007. doi: 10.1136/gutjnl-2019-319635

Zhao, L., Cho, W. C., and Nicolls, M. R. (2021). Colorectal cancer-associated microbiome patterns and signatures. *Front. Genet.* 12. doi: 10.3389/fgene.2021.787176

Zhu, X., Xu, P., Zhu, R., Gao, W., Yin, W., Lan, P., et al. (2024). Multi-kingdom microbial signatures in excess body weight colorectal cancer based on global metagenomic analysis. *Commun. Biol.* 7, 1–14. doi: 10.1038/s42003-023-05714-0

Zou, S., Yang, C., Zhang, J., Zhong, D., Meng, M., Zhang, L., et al. (2024). Multi-omic profiling reveals associations between the gut microbiome, host genome and transcriptome in patients with colorectal cancer. *J. Transl. Med.* 22, 175. doi: 10.1186/s12967-024-04984-4