pISSN 1598-2998, eISSN 2005-9256

https://doi.org/10.4143/crt.2024.382 Cancer Res Treat. 2025;57(1):198-211

Original Article

Fecal Microbial Dysbiosis Is Associated with Colorectal Cancer Risk in a Korean Population

Jeongseon Kim[1], **Madhawa Gunathilake**[1], **Hyun Yang Yeo**[2], **Jae Hwan Oh**[3], **Byung Chang Kim**[3], **Nayoung Han**[4], **Bun Kim**[2,3], **Hyojin Pyun**[1], **Mi Young Lim**[5], **Young-Do Nam**[5], **Hee Jin Chang**[2,3,4]

[1]Department of Cancer Biomedical Science, Graduate School of Cancer Science and Policy, National Cancer Center, Goyang, [2]Department of Cancer Diagnostics, Research Institute, National Cancer Center, Goyang, [3]Center for Colorectal Cancer, National Cancer Center Hospital, National Cancer Center, Goyang, [4]Department of Pathology, National Cancer Center Hospital, National Cancer Center, Goyang, [5]Personalized Diet Research Group, Food Functionality Research Division, Korea Food Research Institute, Wanju, Korea

Purpose The association between the fecal microbiota and colorectal cancer (CRC) risk has been suggested in epidemiologic studies. However, data from large-scale population-based studies are lacking.

Materials and Methods In this case-control study, we recruited 283 CRC patients from the Center for Colorectal Cancer, National Cancer Center Hospital, Korea to perform 16S rRNA gene sequencing of fecal samples. A total of 283 age- and sex-matched healthy participants were selected from 890 cohort of healthy Koreans that are publicly available (PRJEB33905). The microbial dysbiosis index (MDI) was calculated based on the differentially abundant species. The association between MDI and CRC risk was observed using conditional logistic regression. Sparse Canonical Correlation Analysis was performed to integrate species data with microbial pathways obtained by PICRUSt2.

Results There is a significant divergence of the microbial composition between CRC patients and controls (permutational multivariate analysis of variance p=0.001). Those who were in third tertile of the MDI showed a significantly increased risk of CRC in the total population (odds ratio [OR], 6.93; 95% confidence interval [CI], 3.98 to 12.06; p-trend < 0.001) compared to those in the lowest tertile. Similar results were found for men (OR, 6.28; 95% CI, 3.04 to 12.98; p-trend < 0.001) and women (OR, 7.39; 95% CI, 3.10 to 17.63; p-trend < 0.001). *Bacteroides coprocola* and *Bacteroides plebeius* species and 12 metabolic pathways were interrelated in healthy controls that explain 91% covariation across samples.

Conclusion Dysbiosis in the fecal microbiota may be associated with an increased risk of CRC. Due to the potentially modifiable nature of the gut microbiota, our findings may have implications for CRC prevention among Koreans.

Key words Fecal microbiome, 16S rRNA gene sequencing, Microbial dysbiosis, Microbial metabolic pathways, Colorectal neoplasms

Introduction

Colorectal cancer (CRC) ranks as the third most common malignancy and second in terms of mortality [1]. In 2020, it was estimated that more than 1.9 million new CRC cases and 935,000 deaths due to CRC occurred worldwide, representing approximately one in 10 cancer cases and deaths [1].

The incidence patterns of CRC are broadly recognized as a marker of a country's socioeconomic condition since the incidence varies according to the economic development of the country [1]. Of the known risk factors for CRC, lifestyle-related factors are important for its prevention since most of them are modifiable [2]. There is convincing evidence that physical activity has a protective effect against CRC, while obesity and frequent consumption of red and processed meat and alcohol drinks increase the risk [3]. Lifestyle, body fat percentage, and dietary patterns can also influence CRC-related morbidity [4]. Recently, several studies have indicated that the gut microbiome is also a key factor that regulates the development of CRC [5,6].

The human gut hosts a diverse community of bacteria that play essential roles in modifying the host's immunity and metabolic functions and help to digest and convert dietary constituents into their active forms [7,8]. The gut microbial population, known as the gut microbiota, has more than 100-fold more genes than the human genome, through which it regulates numerous processes, such as energy harvesting, dietary component metabolism, immunity, and host- or microbial-derived chemical activities [9]. Alterations in the gut microbial communities are closely associated with several diseases including inflammatory bowel disease, metabolic syndrome, and cancer such as CRC [10,11].

Correspondence: Hee Jin Chang
Department of Cancer Diagnostics, Research Institute, National Cancer Center, 323 Ilsan-ro, Ilsandong-gu, Goyang 10408, Korea
Tel: 82-31-920-1741 Fax: 82-31-920-1368 E-mail: heejincmd@ncc.re.kr
Received April 18, 2024 **Accepted** July 25, 2024 **Published Online** July 26, 2024

In an epidemiological context, some studies have compared tumor and adjacent normal tissues to observe microbial compositional differences and have identified specific bacterial species associated with CRC risk [12,13]. The application of high-throughput sequencing technologies to investigate the gut microbiota revealed differences in the gut microbial composition between CRC patients and healthy individuals [14]. Although numerous studies have already examined the association between the human microbiome and CRC risk, data from large-scale population-based studies are lacking. Moreover, there is a paucity of evidence for the application of novel compositional analysis techniques that help to derive definitive and high-quality evidence [15]. Therefore, in the current study, we applied novel analytical approaches to characterize the composition of gut microbial communities in a population of Korean CRC patients and healthy controls.

This study aimed to derive a microbial dysbiosis index (MDI) by characterizing the fecal microbial composition using 1:1 matched 283 CRC cases and 283 healthy controls. Moreover, we observed the association between the MDI and the risk of CRC.

Materials and Methods

1. Study population

From 2009 to 2012, 197 patients and from 2018 to 2020, 98 patients were recruited from Center for Colorectal Cancer, National Cancer Center (NCC) Hospital, Korea. Polyps and other cancers were excluded: 4 and 8 from each set. Finally, 283 patients were selected. Healthy controls were selected from a publicly available data set in the European Nucleotide Archive (ENA) repository under accession No. PRJEB33905. In total, 890 healthy participants were publicly accessible and were originally recruited from the healthcare center of Chung-Ang University Hospital (Seoul, Korea) considering following exclusion criteria: those who had been administered antibiotics within 3 months prior to the start of the study, those with a history of major gastrointestinal surgery or active uncontrolled gastrointestinal disorders, those diagnosed with cancer or chronic clinically significant cardiovascular pulmonary, renal or hepatic diseases, and women who were pregnant or lactating. The detailed recruitment procedure with exclusion criteria has been described elsewhere [16]. We performed propensity score matching based on age and sex to select 1:1 matched CRC cases and controls using the MatchIt R package [17]. The final sample of 566 participants was composed of 283 CRC patients and 283 healthy controls for the analysis. The general characteristics of the study population are shown in Table 1. Since the metadata on sex and age were not publicly accessible for the cohort of healthy controls, we contacted the principal investigator of the respective study to obtain sex and age variables for the 890 samples available in ENA. Their study approval information and information on written informed consent can be found in Lim et al. (2021) [16].

Table 1. General characteristics of the study participants

Characteristic	Control (n=283)	Case (n=283)	p-value[a]
Age (yr)	59.9±11.8	60.7±11.1	0.384
Sex			
Male	162 (57.2)	162 (57.2)	> 0.99
Female	121 (42.8)	121 (42.8)	
Tumor site			
Colon	-	164 (58.0)	-
Rectum	-	119 (42.1)	
Stage			
Stage 0-I	-	67 (23.8)	-
Stage II	-	73 (25.9)	
Stage III	-	105 (37.2)	
Stage IV	-	37 (13.1)	

Values are presented mean±standard deviation or number (%).
[a]Student's t test and chi-squared test.

2. Sample collection and gene sequencing

Fecal samples were collected from each participant. Following collection, the samples were immediately delivered to the laboratory for processing, DNA extraction, and 16S rRNA gene sequencing. The methods of sample collection and DNA extraction differed slightly between the two study populations. Details of the methods have been described elsewhere [16,18]. Briefly, in the NCC study, up to 10 mL of fresh fecal sample was collected in a sterile tube or OMNIgene GUT tube (DNA Genotek, Ottawa, Canada). The samples were brought to the laboratory within 12 hours, and the samples were immediately stored at –70°C DNA was extracted using a QIAamp DNA Stool Kit according to the manufacturer's protocol (Qiagen, Valencia, CA). The quality and concentration of the DNA were checked using a Nanodrop 2000 spectrophotometer (Nanodrop Technologies, Wilmington, DE).

In the cohort of healthy controls, the participants collected fecal samples at home within 48 hours of the study using OMNIgene GUT tubes (DNA Genotek) according to the manufacturer's instructions. DNA was extracted using a QIAamp DNA stool mini kit (Qiagen, Hilden, Germany). The extracted DNA samples were stored at –20°C until further analysis for 16S rRNA gene sequencing [16].

The 16S rRNA amplicons covering hypervariable regions V3-V4 of the bacterial 16S rRNA gene were sequenced using

the MiSeqTM platform (Illumina, San Diego, CA) in NCC study [18]. For the study related to healthy controls, library construction of the V3-V4 hypervariable regions was performed following the 16S metagenomic sequencing library preparation Illumina protocol. Amplicon libraries for each sample were pooled at equimolar quantities and sequenced using a MiSeq 2×300 instrument (Illumina) [16].

Reads from the sequencing were sorted using unique barcodes for each PCR product. The barcode, linker, and primer sequences were then removed from the original sequencing reads, and the removed reads were merged by paired-end reads using FLASH v 1.2.11. Raw FASTQ files composed of paired-end (forward and reverse) sequencing reads obtained from each study were merged and used for bioinformatics analysis using Quantitative Insights Into Microbial Ecology (QIIME2), a plugin-based platform for microbiome analysis.

The paired-end FASTQ files that had already been demultiplexed were imported to generate QIIME2 artifact files. The DADA2 pipeline was applied to perform the quality control steps for the raw sequences that helped to remove low-quality reads, ambiguous reads, chimeric reads, dereplicate sequences, cluster sequences, and chimeras using QIIME v2.2021.4 [19]. An amplicon sequence variant (i.e., 100% exact sequence match) table was generated with forward and reverse truncation lengths of 278 and 210, respectively. Finally, representative sequence file and table file were obtained as the end products. Sampling depth was identified based on the table file, and an alpha rarefaction curve was plotted. Taxonomic classification was assigned based on the naïve Bayes classifier using the classify-sklearn package against the Silva-138-99 reference sequences. Host mitochondria and chloroplasts, archaea, eukaryotes, and unassigned reads were filtered before calculating relative abundance. The microbial composition was normalized using the values calculated from the taxonomic abundance count divided by the number of preprocessed reads for each sample to obtain the relative abundance. Six taxonomy levels were used for the analysis namely, phylum, class, order, family, genus, and species.

3. Diversity analyses

Microbial diversity analyses were carried out to measure within-sample diversity (alpha diversity) and between-sample diversity (beta diversity). For the alpha diversity, six indices were calculated, namely, ACE, Chao1, Observe, Pielou, Shannon, and Simpson. For the beta diversity, distance matrix was obtained using weighted UniFrac distance measure, and principal coordinate analysis (PCoA) was performed to visualize the microbial composition between CRC cases and controls. Permutational multivariate analysis of variance (PERMANOVA) was used to observe the significance of the beta diversity differences. The diversity analyses were performed based on the anatomical subsites and the TNM stages of CRC.

4. Differential abundance analysis and cladogram

Differential abundance analysis of the taxa was performed for overall, and subgroups based on anatomical location and TNM stages using statistical analysis of taxonomic and functional profiles (STAMP v2) software, a tool that provides extensive hypothesis testing, exploratory plots, effect size measures, and confidence intervals (CIs) for facilitating the identification of biologically relevant differences [20]. The feature table was prepared based on the Silva database as an input file to plot the cladogram which shows the phylogenetic relationship of microbial taxa between CRC cases and controls.

5. Identification of the enterotypes

Microbial enterotypes (ET) in the human gut microbiome have been introduced in the Arumugam et al. (2011) [21]. We selected only CRC cases to identify the enterotypes. Clustering of the samples based on the relative abundance of genera among CRC cases was performed with the use of Jensen-Shannon divergence distance and the partitioning around medoids clustering algorithm. The optimal number of clusters was obtained based on the Calinski-Harabasz (CH) index. ET were obtained for overall, subgroups based on anatomical sites and TNM stage.

6. Deriving the MDI

The concept of the MDI is a unique way of quantifying the imbalance of the microbial community, although there is no gold standard to determine the presence or extent of a given imbalance or disturbance. Several indexes have been defined and applied. These indexes can help characterize diseases and adverse conditions, predict treatment outcomes, and provide information beyond the commonly used alpha and beta diversity assessments [22,23]. Thus, a positive MDI indicates a shift towards dysbiosis, where harmful microbial populations are predominant, and high MDI values have been associated with an increased risk of CRC. In contrast, a negative MDI suggests a balanced or healthy microbial community with a predominance of beneficial microbes that support gut health and lower CRC risk. Based on the STAMP, we identified differentially abundant taxa in CRC cases and controls at the species level. The MDI was calculated as the log of [total abundance among species increased in CRC] over [total abundance among species decreased in CRC] [24]. The mean MDI between CRC cases and controls was compared using Student's t test, and the distribution of MDI between two groups was visualized in Violin plots using the

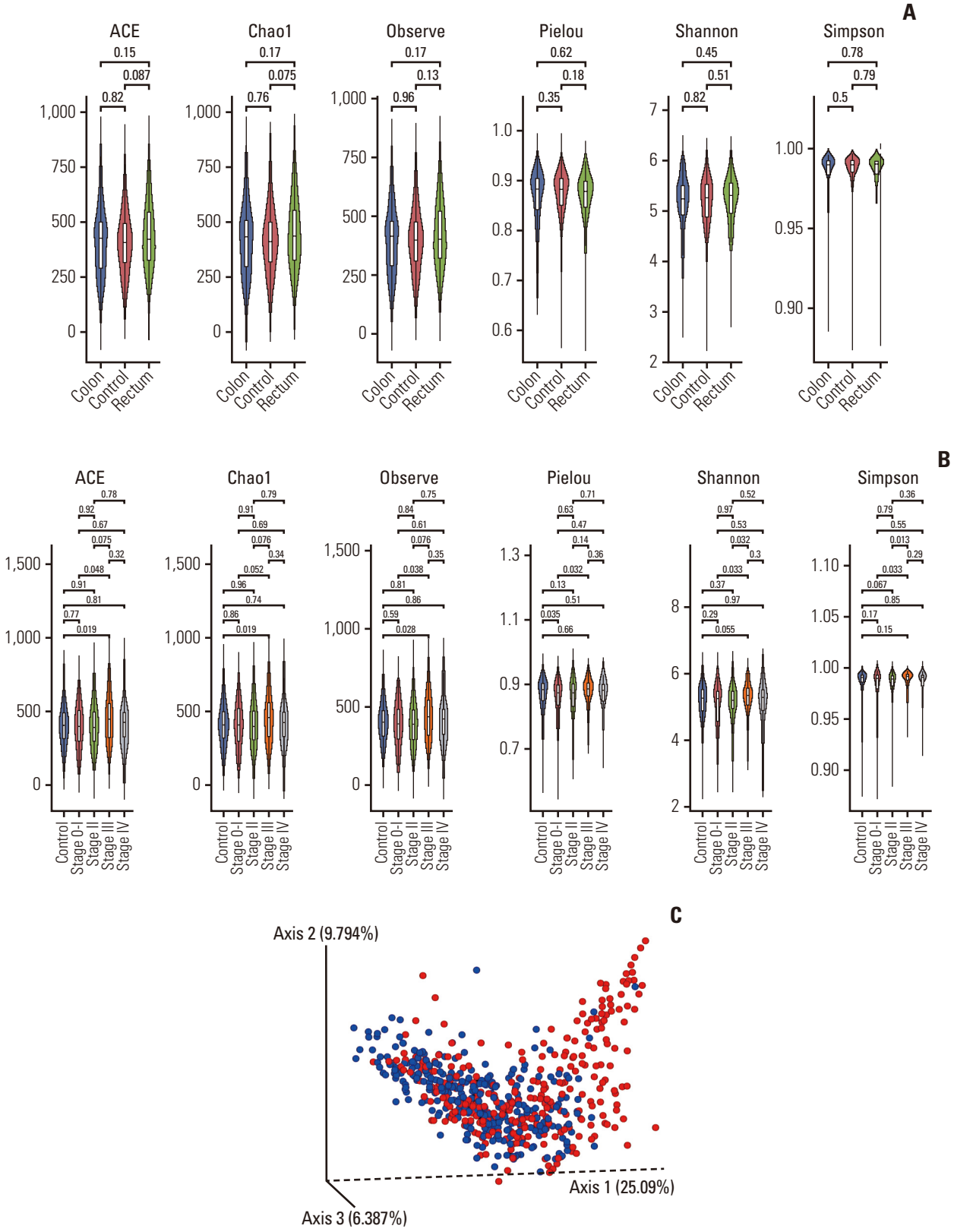

Fig. 1. (A) Comparison of alpha diversity indices based on anatomical sites. (B) Based on cancer stage. (C) Principal coordinate analysis plot of the weighted UniFrac distance. Red indicates colorectal cancer cases and blue indicates healthy controls.

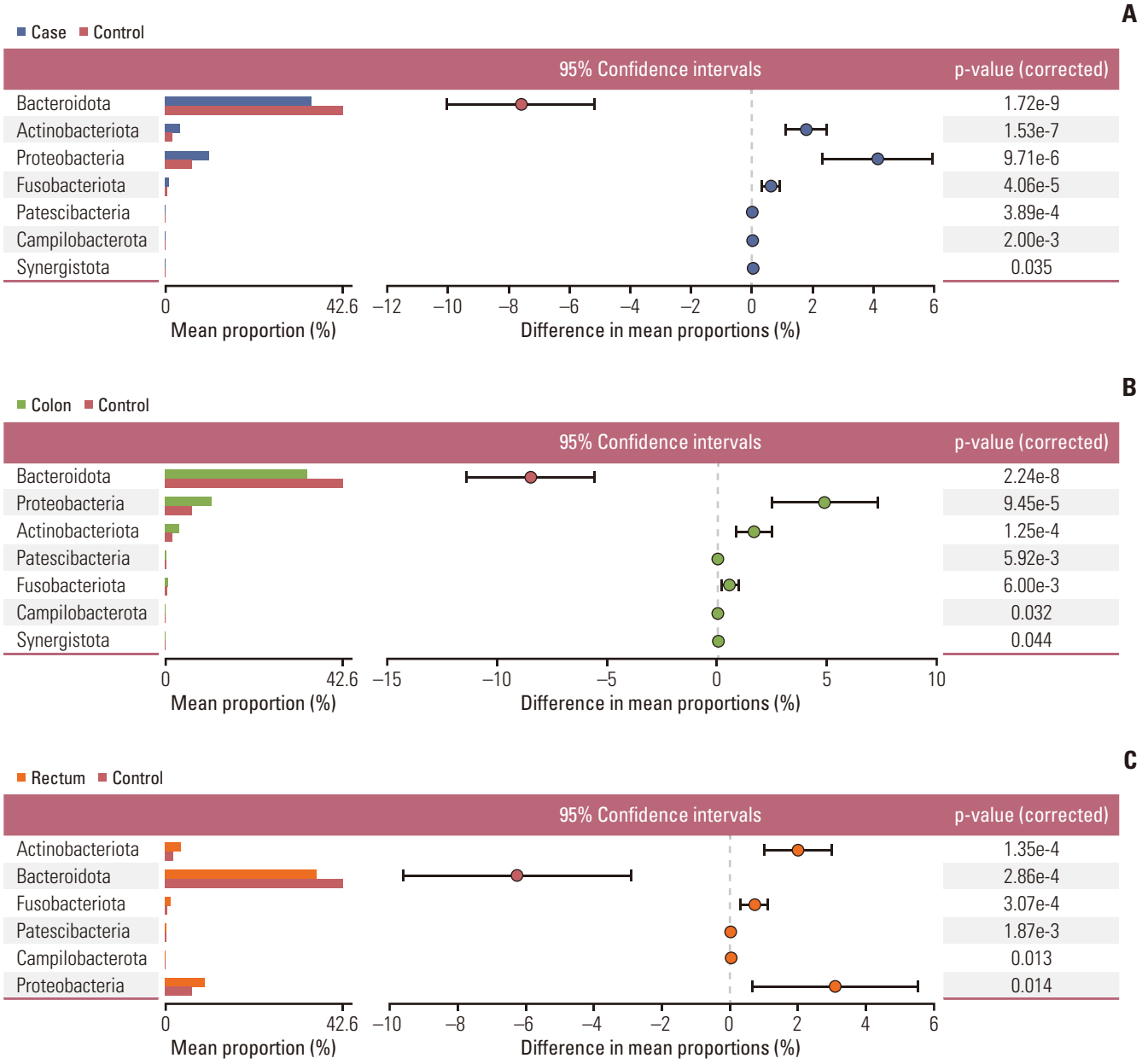

Fig. 2. (A) Statistical analysis of taxonomic and functional profiles (STAMP) results plot for taxonomic phylum STAMP results plots for colon cancer (B) and rectal cancer (C) at phylum level.

R package "ggplot2." The comparison of the MDI between CRC cases and controls were visualized depending on ETs, anatomical sites, and TNM stage.

7. Association between fecal microbial dysbiosis and CRC risk

The MDI was categorized into tertiles based on the distribution in the control group. The group with the lowest MDI was used as the reference group. Odds ratios (ORs) and 95% CIs were estimated using conditional logistic regression models. The median values of the MDI in each tertile category were used as continuous variables to test for trends. The associations for each enterotype-specific CRC and for anatomical sites were performed using multinomial logistic regression models. All statistical analyses were carried out using SAS v9.4 software (SAS Inc., Cary, NC), the R platform v3.5.1 (R Foundation for Statistical Computing, Vienna, Austria), and QIIME2-2021.4 [19].

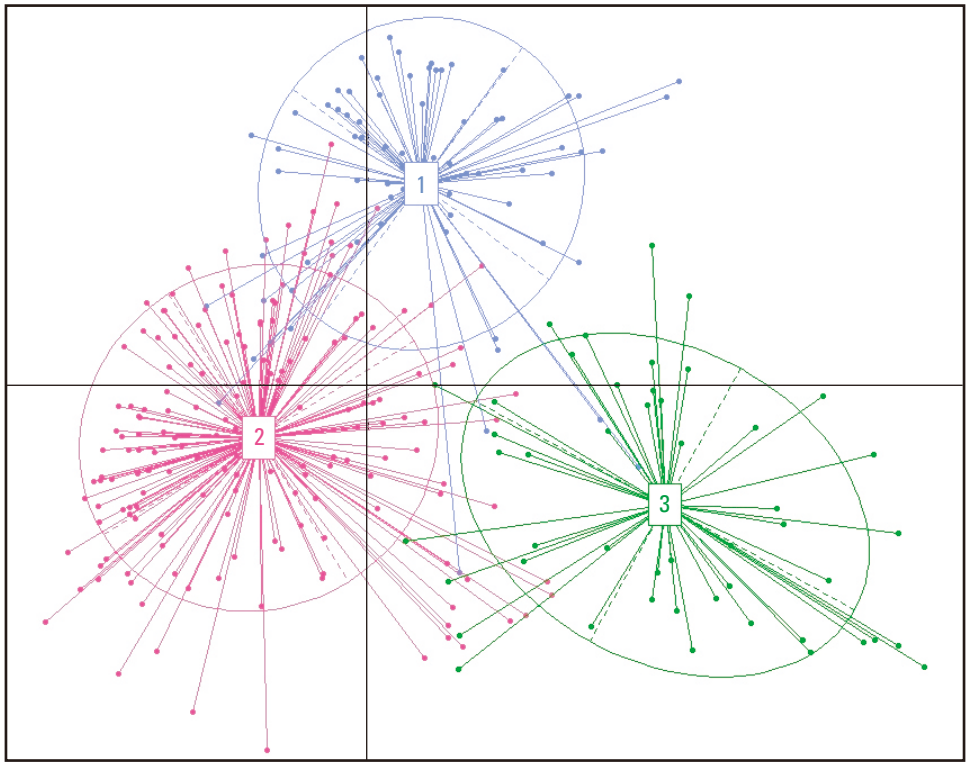

Fig. 3. Enterotypes identified from the microbial composition of colorectal cancer cases.

8. Metagenomic functional analysis

The fecal microbial functional gene contents were predicted using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt v2). The input sequence abundance table was normalized by the predicted number of marker genes and then the predicted functional profiles per sample were determined. Finally, Enzyme Commission (EC) number and Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO) metagenome and MetaCyc pathway abundances were predicted based on the predicted EC number abundances. Differential abundance pathways were determined using STAMP. Sparse Canonical Correlation Analysis (sCCA) was performed to identify group-level correlations between microbial species and their metabolic pathways as two sets of variables to be correlated using R package 'PMA.' Correlation network was constructed for selected interrelated species and their pathways by detecting the communities within the network based on greedy optimization of modularity of the network using R package 'igraph.' The integration of species and pathways was performed based on the anatomical sites of CRC.

Table 2. Comparison of MDI between CRC cases and controls

	Case	Control	p-value
Total (n=287)			
MDI	0.29±1.85	−1.28±1.64	< 0.001
Male (n=164)			
MDI	0.26±1.91	−1.31±1.55	< 0.001
Female (n=123)			
MDI	−0.04±1.25	−1.44±0.71	< 0.001

Values are presented mean±standard deviation. CRC, colorectal cancer; MDI, microbial dysbiosis index.

Results

1. Comparison of diversity measures between CRC patients and controls

The comparison of alpha diversity measures between CRC patients and controls showed no significant difference between two groups (S1 Fig.). Also, no significant difference was observed for each alpha diversity index based the anatomical sites; colon and rectum (Fig. 1A). Compared to controls, ACE (p=0.019), Chao1 (p=0.019), and Observe (p=0.028) indices were significantly higher in stage III can-

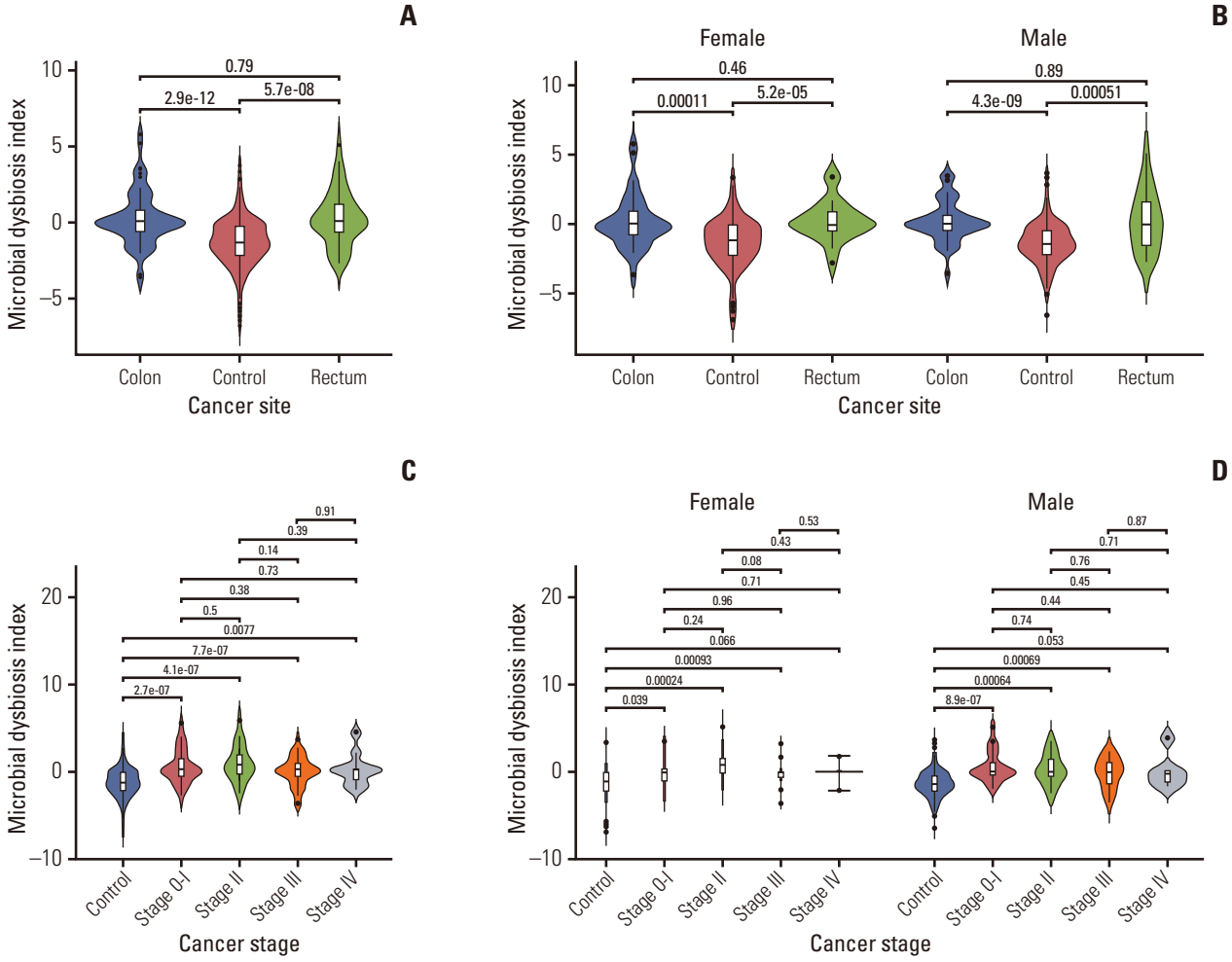

Fig. 4. Comparison of microbial dysbiosis index between each anatomical sites of colorectal cancer cases and controls. (A) Total population. (B) Each sex. (C) Across cancer stages. (D) Across cancer stages by sex.

cers. There was a significantly higher alpha diversity in stage III compared to stages 0-I cancers in terms of Observe index (p=0.038). There was a significantly higher Shannon index (p=0.032) and Simpson index (p=0.013) in stage III compared to stage II cancers (Fig. 1B).

We performed beta diversity analysis based on the weighted UniFrac distance measure, and the sampling depth was identified as 6,194 from the table file. The PERMANOVA test results indicated a significant divergence in the microbial composition between CRC patients and controls. We further found that the first three principal coordinates of the PCoA based on weighted UniFrac distance explained 41.27% of the overall microbial diversity (PERMANOVA p=0.001), which was the highest among the four distance measures (Fig. 1C). Similar difference was observed among colon, rectal cancers and controls (PERMANOVA p=0.001) (S2 Fig.).

2. Differential abundance analysis of taxa and cladogram

Differentially abundant taxa between CRC cases and controls were identified using STAMP analysis for six taxonomy levels. The effect sizes with their 95% CIs and the corrected p values are shown in the figure for each identified taxon. At the phylum level, *Bacteroidota* was enriched in controls whereas *Actinoabcteriota*, *Proteobacteria*, and *Fusobacteriota* phyla were enriched in CRC cases (Fig. 2A). In colon and rectal cancers, *Proteobacteria* and *Actinobacteriota* phyla were enriched compared to controls (Fig. 2B and C) *Proteobacteria* phyla was highly enriched in stages 0-I and stage II cancers while *Actinobacteriota* and *Fusobacteriota* phyla were enriched in stage III and IV cancers compared to controls (S3A-D Fig.).

At the genus level, *Fecalibacterium*, *Prevotella*, and *Bacteroides* were the major differentially abundant genera in controls while *Escherichia-Shigella*, *Alistipes*, *Bifidobacterium*, *Veillonella*, and *Enterococcus* were identified as the major dif-

ferentially enriched genera in CRC cases (S4A Fig.). In colon and rectal cancers, *Escherichia-Shigella* genus was differentially enriched (S4B and S4C Fig.). In stages 0-I, II, and III cancers, *Escherichia-Shigella* genus was highly enriched whereas *Phascolarctobacterium*, *Parabacteroides* genera were highly enriched in stage IV cancers (S5A-S5D Fig.).

At the species level, *Ruminococcus bicirculans*, *Bacteroides plebeius*, *Lachnospiraceae* bacterium, *Prevotella copri*, and *Bacteroides coprocola* were enriched in controls whereas *Parabacteroides merdae*, *Lactobacillus ruminis*, *Fusobacterium necrophorum*, *Odoribacter splanchnicus*, and *Alistipes shahii* were enriched in CRC cases (S6 Fig.).

Based on the Silva database, we plotted the cladogram up to the species level. C_*Bacteroidia*, F_*Bacteroidaceae*, F_*Ruminococcaceae*, F_*Sutterellaceae*, and S_*Bacteroides plebeius* were phylogenetically related in healthy controls, while several taxa, including O_*Enterobacterales*, O_*Fusobacteriales*, F_*Eggerthellaceae*, F_*Erysipelatoclostridiaceae*, F_*Erysipelotrichaceae*, and G_*Streptococcus* were phylogenetically enriched in CRC patients (S7 Fig.).

3. Enterotypes among CRC cases

Three optimal clusters were identified based on the CH index (S8 Fig.). As a result, three ETs were obtained with respect to the CRC case microbiome at the genus level. ET1 was characterized by *Ruminococcus*, while ET2 and ET3 were predominantly characterized by *Bacteroides* and *Prevotella*, respectively (Fig. 3). Consistent ETs were found for the colon cancers while in rectal cancer, ET1 and ET2 were consistently characterized by *Ruminococcus*, and *Bacteroides*, and ET3 was characterized by *Fecalibacterium* (S9A and S9B Fig.). Based on cancer stages, we found that the ET1 was characterized by *Ruminococcus* in early cancers while in advanced cancers, ET2 and ET3 were consistently characterized by *Bacteroides* and *Prevotella* (S9C and S9D Fig.).

4. Microbial dysbiosis index

Table 2 shows the comparison of the MDI between CRC patients and controls. Interestingly, we found that the MDI was significantly higher in CRC patients than in healthy controls in the total population ($p < 0.001$), males ($p < 0.001$), and females ($p < 0.001$) (S10A and S10B Fig.). According to the enterotype-specific CRC, the MDI significantly differed among ET1, ET2, and ET3 CRC patients compared with healthy controls in overall population ($p < 0.001$) and each sex (S10C and S10D Fig.).

Based on the anatomical sites, MDI was significantly higher in both colon and rectal cancers compared to controls in overall ($p < 0.05$) and in each sex ($p < 0.001$) (Fig. 4A and B). Based on cancer stage, MDI was significantly higher in each cancer stage (0-IV) compared to controls in overall ($p < 0.05$). However, in males and females, significantly higher MDI was observed only in stages 0-I, II, and III compared to controls ($p < 0.05$) (Fig. 4C and D).

Table 3 shows the association between MDI and CRC risk. In the total population, those in the third tertile of the MDI showed a significantly increased risk of CRC (OR, 6.93; 95% CI, 3.98 to 12.06; p-trend < 0.001) compared to those in the lowest tertile. Males who had a higher MDI showed an increased risk of CRC compared to those with a low MDI (OR, 6.28; 95% CI, 3.04 to 12.98; p-trend < 0.001). In females, those in the third tertile of the MDI showed a significantly increased risk of CRC (OR, 7.39; 95% CI, 3.10 to 17.63; p-trend < 0.001) compared to those in the lowest tertile. Regarding anatomical sites, those who were in the highest tertile of MDI showed increased risks for colon cancers (OR, 7.63; 95% CI, 4.21 to 18.23; p-trend < 0.001) and rectal cancer (OR, 6.21; 95% CI, 3.30 to 11.69; p-trend < 0.001). Regarding enterotype-specific CRC, those in the third tertile of MDI showed increased risks of ET1-CRC (OR, 9.31; 95% CI, 4.91 to 17.63; p-trend < 0.001), ET2-CRC (OR, 3.54; 95% CI, 1.80 to 6.97; p-trend < 0.001), and ET3-CRC (OR, 10.50; 95% CI, 3.62 to 30.43; p-trend < 0.001) (S11 Table).

5. Integrate microbiome and metabolic pathways

To perform sCCA, 48 species and 202 pathways identified by STAMP were selected. As a result, two bacterial species (*B. coprocola* and *B. plebeius*) and 12 microbial pathways (mannan degradation, L-rhamnose degradation I, preQ0 biosynthesis, queuosine biosynthesis, superpathway of GDP mannose derived O antigen building blocks biosynthesis, GDP mannose biosynthesis, polyisoprenoid biosynthesis *E. coli*, 4-deoxy-L-threo-hex-4-enopyranuronate degradation, NAD salvage pathway I, pyrimidine deoxyribonucleosides salvage, S-adenosyl-L-methionine cycle I, urate biosynthesis inosine 5 phosphate degradation) were selected based on their ability to explain covariance. These 14 features result in a correlation of 0.68 between two tables (species and pathways). Also, selected species and pathways are closely associated with healthy controls compared to cases. The association of those bacterial species and the 12 pathways could explain the 91% variation across samples (S12A Fig.). According to the anatomical sites, similar results were observed in colon cancer having two bacterial species and 12 pathways which are closely associated with healthy controls than colon cases (S12B Fig.). For the rectal cancer, two bacterial species (*B. coprocola* and *B. plebeius*) and 10 pathways were integrated to explain the 90% variation across samples (S12C Fig.).

According to the correlation network analysis, two major modules were detected as communities within the whole network. One module was comprised with *B. coprocola* species whereas the other module was comprised with *B. ple-*

Table 3. Association between MDI and CRC risk

MDI	Case, n (%)	Control, n (%)	Model: OR (95% CI)
Colorectal cancer			
Total			
T1 (< –1.87)	30 (10.6)	95 (33.6)	1.00
T2 (–1.87 to –0.64)	44 (15.6)	93 (32.9)	1.35 (0.72-2.52)
T3 (> –0.64)	209 (73.9)	95 (33.6)	6.93 (3.98-12.06)
p-trend			< 0.001
Male			
T1 (< –1.95)	19 (11.7)	54 (33.3)	1.00
T2 (–1.95 to –0.79)	22 (13.6)	54 (33.3)	1.04 (0.44-2.50)
T3 (> –0.79)	121 (74.7)	54 (33.3)	6.28 (3.04-12.98)
p-trend			< 0.001
Female			
T1 (< –1.84)	11 (9.1)	41 (33.9)	1.00
T2 (–1.84 to –0.38)	26 (21.5)	39 (32.2)	2.18 (0.83-5.74)
T3 (> –0.38)	84 (69.4)	41 (33.9)	7.39 (3.10-17.63)
p-trend			< 0.001
Colon cancer			
Total			
T1 (< –1.87)	16 (9.8)	95 (33.6)	1.00
T2 (–1.87 to –0.64)	26 (15.9)	93 (32.9)	1.66 (0.84-3.29)
T3 (> –0.64)	122 (74.4)	95 (33.6)	7.63 (4.21-13.81)
p-trend			< 0.001
Male			
T1 (< –1.95)	9 (10.1)	54 (33.3)	1.00
T2 (–1.95 to –0.79)	11 (12.4)	54 (33.3)	1.22 (0.47-3.19)
T3 (> –0.79)	69 (77.5)	54 (33.3)	7.77 (3.48-16.90)
p-trend			< 0.001
Female			
T1 (< –1.84)	8 (10.7)	41 (33.9)	1.00
T2 (–1.84 to –0.38)	17 (22.7)	39 (32.2)	2.23 (0.87-5.76)
T3 (> –0.38)	50 (66.7)	41 (33.9)	6.25 (2.64-14.81)
p-trend			< 0.001
Rectal cancer			
Total			
T1 (< –1.87)	14 (11.8)	95 (33.6)	1.00
T2 (–1.87 to –0.64)	18 (15.1)	93 (32.9)	1.31 (0.62-2.79)
T3 (> –0.64)	87 (73.1)	95 (33.6)	6.21 (3.30-11.69)
p-trend			< 0.001
Male			
T1 (< –1.95)	10 (13.7)	54 (33.3)	1.00
T2 (–1.95 to –0.79)	11 (15.1)	54 (33.3)	1.10 (0.43-2.80)
T3 (> –0.79)	52 (71.2)	54 (33.3)	5.20 (2.40-11.30)
p-trend			< 0.001
Female			
T1 (< –1.84)	3 (6.5)	41 (33.9)	1.00
T2 (–1.84 to –0.38)	9 (19.6)	39 (32.2)	3.15 (0.80-12.51)
T3 (> –0.38)	34 (73.9)	41 (33.9)	11.33 (3.22-39.85)
p-trend			< 0.001

CI, confidence interval; CRC, colorectal cancer; MDI, microbial dysbiosis index; OR, odds ratio.

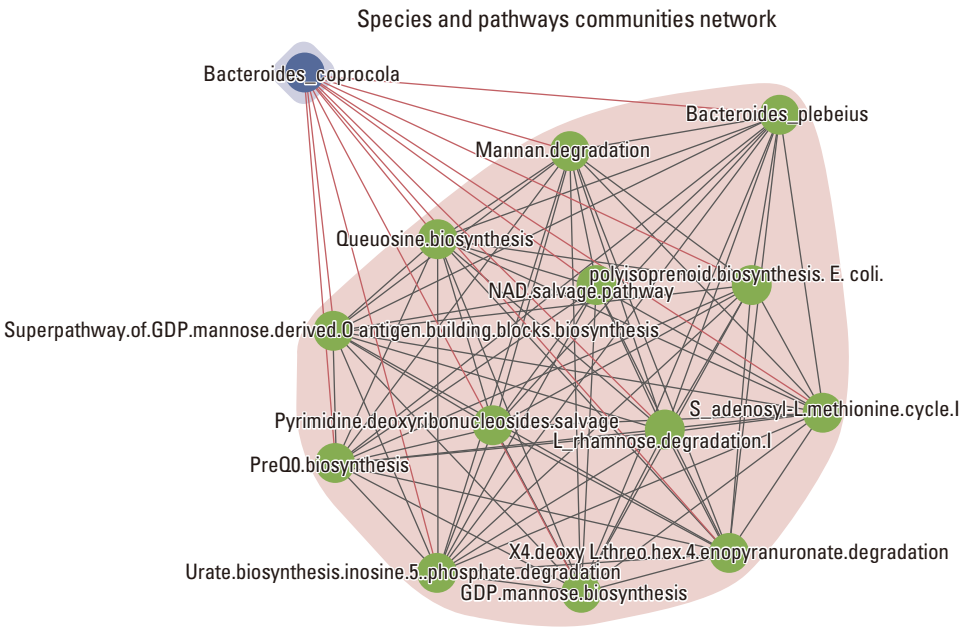

Fig. 5. Community detection based on greedy optimization of modularity of the network.

beius species and other 12 pathways. *B. coprocola* community was independently correlated with other species and pathways (Fig. 5).

Discussion

This study derived MDI based on differentially abundant species to observe the association between fecal microbial dysbiosis and CRC risk using 1:1 matched CRC cases and healthy controls. The MDI was significantly higher in CRC patients than in controls. The difference of MDI was consistently observed for each anatomical site and cancer stage. A higher MDI was significantly associated with a higher risk of CRC. The correlation of *B. coprocola* and *B. plebeius* with 12 pathways was closely associated with healthy controls compared to CRC.

In our study, we found that there was no significant difference in alpha diversity indices between CRC cases and controls. Previous evidence showed inconsistent findings related to alpha diversity [25-27]. However, ACE, Chao1, and Observe indices were significantly higher in stage III cancers compared to controls. In a Chinese study, no significant differences in Chao1, Shannon, and Simpson indices were observed between CRC patients of stages I-IV and healthy controls [28]. Notably, beta diversity showed a significant divergent between CRC cases and controls consistent with our results [25,29].

In the current study, the *Bacteroidota* phylum was enriched in controls, while the *Actinobacteriota*, *Proteobacteria*, and *Fusobacterium* phyla were enriched in CRC patients. Specifically, at the genus level, *Escherichia-Shigella* and *Fusobacterium* were highly enriched in CRC patients. Consistent with our results, *Escherichia-Shigella* and *Fusobacterium* were highly abundant in CRC cases than controls [25,26]. Previous studies also showed that *Porphyromonas* and *Lactobacillus* were significantly enriched in fecal samples of patients with CRC [26,30]. In contrast, a nonsignificant higher abundance of the *Lactobacillus* genus was identified in controls [31]. Moreover, *Bacteroides fragilis*, *Enterococcus*, *Escherichia-Shigella*, *Klebsiella*, *Streptococcus*, *Peptostreptococcus*, *Enterobacteriaceae*, and *Porphyromonas* displayed a higher relative abundance in CRC patients, while *Roseburia*- and *Lachnospiraceae*-related species dominated in healthy controls, which are consistent with our study [32,33]. As we observed for cancer stages, previous studies reported that *Peptostreptococcus*, *Collinsella*, and *Ruminococcus* were enriched in stage I patients while *Alistipes*, and *Parabacteroides* were highly enriched in stage III-IV patients [28,34]. We found that *Escherichia-Shigella* was highly enriched in stage 0-III patients. In contrast, a study reported that *Haemophilus* in the CRC stage 0 group was significantly higher than in the CRC stages I-II groups [35] while another study reported that there was no significantly different fecal microbiota at stages I-IV [36].

We found that *B. plebeius* was highly enriched in healthy controls in our study similar to previous evidence [37]. It has

been reported that *B. plebeius* reduces claudin-2 expression while increasing intestinal mucosal tightness and restoring the intestinal mucosal barrier which can directly affect for enhancing intestinal barrier integrity and thereby prevent CRC risk [38]. The *Fusobacterium* genus was highly enriched in CRC cases in our study. Moreover, we found that *F. necrophorum* was a major species abundant in CRC cases. Plausible biological mechanisms of *Fusobacterium* species are ability to inhibit T-cell proliferation and induce T-cell apoptosis which may impair the host immunity in eliminating transformed cancer cells and ability of FadA adhesin molecule to bind to E-cadherin on the colonic epithelium to activate the Wnt/β-catenin oncogenic pathway [39]. *P. merdae* was highly enriched in CRC cases in a Thai population which is similar to our results [40]. The bacterial species of the *Parabacteroides* genus could degrade the beneficial dietary flavonoids found in plant-based diets by lowering their bioavailability. Thus, the beneficial anticancer effect of flavonoids could deteriorate while increasing CRC carcinogenesis [40].

A prior study reported that the enrichment of potential pathogens and the depletion of butyrate-producing members represent a specific microbial signature of CRC [41]. Wu et al. [41] identified 16 genera that were significantly more abundant in CRC samples than in control samples, including potentially pathogenic *Fusobacterium* and *Campylobacter*. A recent study classified 12 genera, and four of which were enriched in controls namely *Bacteroides*, *Parabacteroides*, *Roseburia*, and *Faecalibacterium*, similar to our findings [42]. A possible mechanism of microbiota dysbiosis in CRC is the reduction in the abundance of butyrate-producing bacterial species which eventually lack the mucosal barrier function [43]. Moreover, dysbiosis related to *Alistipes* species has been linked to colon cancer which is consistent with the highly enriched *Alistipes shahii* speices and *Alistipes* genus in CRC in our study [44]. In overall, due to dysbiosis, normal physiological functions of gut microbiota are altered leading to pathological damage to intestinal lining and intestinal barrier. This will eventually induce immune system disorders which can cause inflammatory diseases such as cancers [45,46].

We found that the integration of two species namely *B. coprocola* and *B. plebeius* and 12 identified pathways could explain the covariance across the samples and those 14 markers are closely associated with healthy controls compared to CRC cases. One key metabolic pathway identified by sCCA is mannan degradation which is a hemicellulose component found in plant cell walls and microbes are rich source of mannan [47]. Fermentation of mannan to short-chain fatty acids has been reported to exhibit biological activity such as anti-inflammatory, anticancer, immunomodulatory and gastroprotective properties [47]. Another key pathway found was PreQ0 biosynthesis and PreQ0 base which is a biosynthetic precursor of queuosine-tRNA derived from *Streptomyces qinglanensis* has been identified as an unusual metabolite with anticancer activity [48]. Other two important pathways are NAD salvage and pyrimidine deoxyribonucleosides salvage enriched in healthy controls compared to CRC cases. It has been reported that both NAD and pyrimidine ribonucleotides are involved in many cellular functions such as DNA repair, cell growth, and cell death. Many of those pathways are typically dysregulated in cancer cells where DNA repair and cell growth is not under control that leads to cancer cell proliferation [49]. S-adenosyl-L-methionine is a naturally occurring sulfur-containing nucleoside synthesized from adenosine triphosphate and methionine has been identified to have novel therapeutic potential in overcoming drug resistance in colon cancer cells that would provide better clinical outcomes [50]. Furthermore, we found that Calvin-Benson-Bassham cycle, stearate biosynthesis II, palmitoleate biosynthesis I, oleate biosynthesis IV pathways which are beneficial pathways were enriched in healthy controls. Consistently, a recent Korean prospective study using CRC cases has identified those pathways as beneficial pathways to lower CRC risk [51].

In modules identified by the correlation network derived from those 14 bacterial and pathway markers that were enriched in healthy controls, *B. coprocola* species was independently correlated with *B. plebeius* and other pathways. Notably, both *B. coprocola* and *B. plebeius* were identified to be highly correlated in short-chain fatty acids and mucosal barrier integrity, respectively [38,52].

There are several strengths of our study. Firstly, we have relatively larger sample size compared to previous studies that may improve the statistical power. Secondly, healthy microbiome data delineates the gut microbial composition in healthy Koreans and suggests utilization as the control group for case-control studies [16]. Both data sets used similar 16S rRNA gene sequencing methods by covering the V3-V4 hypervariable region, which could minimize the errors associated with the combination of two separate data sets. However, there are also potential limitations associated with the current study. While our study provides valuable insights into the microbial differences associated with colorectal cancer, the noted limitations regarding sample collection and DNA extraction methods in terms of collecting container and time between collection and analysis time must be acknowledged. These methodological discrepancies may introduce biases that affect our findings' validity and reproducibility. Lack of information on colonoscopic screening of polyps or tumor status among controls could introduce selection bias. Since this study is a combination of two studies, the metadata associated with the study participants are limited such as

smoking, alcohol consumption, red meat consumption, and physical activity. However, we tried to perform propensity score matching based on sex and age to select a 1:1 matched control group for CRC cases, which may minimize selection bias. Second, the current study was not a prospective design. Thus, associations between the fecal microbiome and CRC risk could have occurred without having a causal relation.

In conclusion, we found that the fecal microbiota of CRC patients differs from that of healthy controls. Microbial dysbiosis is significant associated with higher risk of CRC. The microbial dysbiosis was significantly differed based on colon and rectal cancers and across cancer stages. Integrating *B. coprocola* and *B. plebeius* species and 12 metabolic pathways would be closely correlated with healthy participants and could explain much covariation across samples. Characterizing the fecal microbial profile associated with CRC would be beneficial to identify CRC-related bacterial and metabolic pathway signatures to develop novel CRC preventive and therapeutic guidelines based on the individual microbiome profiles of Koreans in the future.

Electronic Supplementary Material

Supplementary materials are available at Cancer Research and Treatment website (https://www.e-crt.org).

Ethical Statement

This study was conducted in accordance with the Declaration of Helsinki. All procedures and protocols of the study were approved by the Institutional Review Board of the NCC (approval no: NCC2018-0070 and NCC2021-0181). All participants gave their written informed consent.

Author Contributions

Conceived and designed the analysis: Kim J, Gunathilake M, Chang HJ.
Collected the data: Kim J, Yeo HY, Oh JH, Kim BC, Han N, Kim B, Chang HJ.
Contributed data or analysis tools: Kim J, Gunathilake M, Yeo HY, Oh JH, Kim BC, Han N, Kim B, Lim MY, Nam YD, Chang HJ.
Performed the analysis: Kim J, Gunathilake M, Pyun H, Chang HJ.
Wrote the paper: Kim J, Gunathilake M.
Critically reviewed the manuscript: Kim J, Gunathilake M, Yeo HY, Oh JH, Kim BC, Han N, Kim B, Lim MY, Nam YD, Chang HJ.

ORCID iDs

Jeongseon Kim : https://orcid.org/0000-0002-0889-2686
Hee Jin Chang : https://orcid.org/0000-0003-2263-2247

Conflicts of Interest

Conflict of interest relevant to this article was not reported.

Acknowledgments

The NCC study was funded by grants from the National Cancer Center of Korea (NCC-1710880 and NCC-2310490). The study related to publicly available microbiome data was supported by the Main Research Program [E0170600-07] of the Korea Food Research Institute (KFRI), funded by the Ministry of Science and ICT.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71:209-49.
2. Murphy N, Moreno V, Hughes DJ, Vodicka L, Vodicka P, Aglago EK, et al. Lifestyle and dietary environmental factors in colorectal cancer susceptibility. Mol Aspects Med. 2019;69:2-9.
3. World Cancer Research Fund, American Institute for Cancer Research. Continuous Update Project Report: diet, nutrition, physical activity and colorectal cancer. London: World Cancer Research Fund International; 2018.
4. Sawicki T, Ruszkowska M, Danielewicz A, Niedzwiedzka E, Arlukowicz T, Przybylowicz KE. A review of colorectal cancer in terms of epidemiology, risk factors, development, symptoms and diagnosis. Cancers (Basel). 2021;13:2025.
5. Cheng Y, Ling Z, Li L. The intestinal microbiota and colorectal cancer. Front Immunol. 2020;11:615056.
6. Wong SH, Yu J. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. Nat Rev Gastroenterol Hepatol. 2019;16:690-704.
7. O'Keefe SJ. Nutrition and colonic health: the critical role of the microbiota. Curr Opin Gastroenterol. 2008;24:51-8.
8. Koropatkin NM, Cameron EA, Martens EC. How glycan metabolism shapes the human gut microbiota. Nat Rev Microbiol. 2012;10:323-35.
9. Sinha R, Ahn J, Sampson JN, Shi J, Yu G, Xiong X, et al. Fecal microbiota, fecal metabolome, and colorectal cancer interrelations. PLoS One. 2016;11:e0152126.
10. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. Nature. 2009;457:480-4.
11. McHardy IH, Goudarzi M, Tong M, Ruegger PM, Schwager E, Weger JR, et al. Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. Microbiome. 2013;1:17.
12. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, et al. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res. 2012;22:292-8.

13. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res. 2012; 22:299-306.
14. Young C, Wood HM, Seshadri RA, Van Nang P, Vaccaro C, Melendez LC, et al. The colorectal cancer-associated faecal microbiome of developing countries resembles that of developed countries. Genome Med. 2021;13:27.
15. Berg G, Rybakova D, Fischer D, Cernava T, Verges MC, Charles T, et al. Microbiome definition re-visited: old concepts and new challenges. Microbiome. 2020;8:103.
16. Lim MY, Hong S, Bang SJ, Chung WH, Shin JH, Kim JH, et al. Gut microbiome structure and association with host factors in a Korean population. mSystems. 2021;6:e0017921.
17. Package 'MatchIt' October 12, 2023, version 4.5.5 [Internet]. Vienna: R Foundation for Statistical Computing; 2023 [cited 2023 Sep 20]. Available from: https://cran.r-project.org/web/packages/MatchIt/MatchIt.pdf.
18. Park SS, Kim B, Kim MJ, Roh SJ, Park SC, Kim BC, et al. The effect of curative resection on fecal microbiota in patients with colorectal cancer: a prospective pilot study. Ann Surg Treat Res. 2020;99:44-51.
19. QIIME [Internet]. QIIME; 2023 [cited 2023 Sep 20]. Available from: http://qiime.org/.
20. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics. 2014;30:3123-4.
21. 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.
22. Wei S, Bahl MI, Baunwall SM, Hvas CL, Licht TR. Determining gut microbial dysbiosis: a review of applied indexes for assessment of intestinal microbiota imbalances. Appl Environ Microbiol. 2021;87:e00395-21.
23. Kim MJ, Jung DR, Lee JM, Kim I, Son H, Kim ES, et al. Microbial dysbiosis index for assessing colitis status in mouse models: a systematic review and meta-analysis. iScience. 2024;27: 108657.
24. Gevers D, Kugathasan S, Denson LA, Vazquez-Baeza Y, Van Treuren W, Ren B, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe. 2014;15:382-92.
25. Liu W, Zhang R, Shu R, Yu J, Li H, Long H, et al. Study of the relationship between microbiome and colorectal cancer susceptibility using 16SrRNA sequencing. Biomed Res Int. 2020; 2020:7828392.
26. Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, et al. Human gut microbiome and risk for colorectal cancer. J Natl Cancer Inst. 2013;105:1907-11.
27. Carson TL, Byrd DA, Smith KS, Carter D, Abaskaron M, Little RB, et al. A case-control study of the association between the gut microbiota and colorectal cancer: exploring the roles of diet, stress, and race. Gut Pathog. 2024;16:13.
28. Sheng Q, Du H, Cheng X, Cheng X, Tang Y, Pan L, et al. Characteristics of fecal gut microbiota in patients with colorectal cancer at different stages and different sites. Oncol Lett. 2019; 18:4834-44.
29. Gao R, Kong C, Li H, Huang L, Qu X, Qin N, et al. Dysbiosis signature of mycobiota in colon polyp and colorectal cancer. Eur J Clin Microbiol Infect Dis. 2017;36:2457-68.
30. Amiot A, Dona AC, Wijeyesekera A, Tournigand C, Baumgaertner I, Lebaleur Y, et al. (1)H NMR spectroscopy of fecal extracts enables detection of advanced colorectal neoplasia. J Proteome Res. 2015;14:3871-81.
31. Mira-Pascual L, Cabrera-Rubio R, Ocon S, Costales P, Parra A, Suarez A, et al. Microbial mucosal colonic shifts associated with the development of colorectal cancer reveal the presence of different bacterial and archaeal biomarkers. J Gastroenterol. 2015;50:167-79.
32. Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, et al. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. ISME J. 2012;6:320-9.
33. Yang J, McDowell A, Kim EK, Seo H, Lee WH, Moon CM, et al. Development of a colorectal cancer diagnostic model and dietary risk assessment through gut microbiome analysis. Exp Mol Med. 2019;51:1-15.
34. Liu J, Huang X, Chen C, Wang Z, Huang Z, Qin M, et al. Identification of colorectal cancer progression-associated intestinal microbiome and predictive signature construction. J Transl Med. 2023;21:373.
35. Yang TW, Lee WH, Tu SJ, Huang WC, Chen HM, Sun TH, et al. Enterotype-based analysis of gut microbiota along the conventional adenoma-carcinoma colorectal cancer pathway. Sci Rep. 2019;9:10923.
36. Du X, Li Q, Tang Z, Yan L, Zhang L, Zheng Q, et al. Alterations of the gut microbiome and fecal metabolome in colorectal cancer: implication of intestinal metabolism for tumorigenesis. Front Physiol. 2022;13:854545.
37. He T, Cheng X, Xing C. The gut microbial diversity of colon cancer patients and the clinical significance. Bioengineered. 2021;12:7046-60.
38. Pei T, Zhu D, Yang S, Hu R, Wang F, Zhang J, et al. Bacteroides plebeius improves muscle wasting in chronic kidney disease by modulating the gut-renal muscle axis. J Cell Mol Med. 2022;26:6066-78.
39. King M, Hurley H, Davidson KR, Dempsey EC, Barron MA, Chan ED, et al. The link between fusobacteria and colon cancer: a fulminant example and review of the evidence. Immune Netw. 2020;20:e30.
40. Iadsee N, Chuaypen N, Techawiwattanaboon T, Jinato T, Patcharatrakul T, Malakorn S, et al. Identification of a novel gut microbiota signature associated with colorectal cancer in Thai population. Sci Rep. 2023;13:6702.
41. Wu N, Yang X, Zhang R, Li J, Xiao X, Hu Y, et al. Dysbiosis signature of fecal microbiota in colorectal cancer patients. Microb Ecol. 2013;66:462-70.
42. Yuan B, Ma B, Yu J, Meng Q, Du T, Li H, et al. Fecal bacteria as non-invasive biomarkers for colorectal adenocarcinoma. Front Oncol. 2021;11:664321.
43. Sokol H, Seksik P, Furet JP, Firmesse O, Nion-Larmurier I, Beaugerie L, et al. Low counts of Faecalibacterium prausnitzii in colitis microbiota. Inflamm Bowel Dis. 2009;15:1183-9.
44. Khattab RH, Abo-Hammam RH, Salah M, Hanora AM,

Shabayek S, Zakeer S. Multi-omics analysis of fecal samples in colorectal cancer Egyptians patients: a pilot study. BMC Microbiol. 2023;23:238.
45. Zhao M, Chu J, Feng S, Guo C, Xue B, He K, et al. Immunological mechanisms of inflammatory diseases caused by gut microbiota dysbiosis: a review. Biomed Pharmacother. 2023;164:114985.
46. Artemev A, Naik S, Pougno A, Honnavar P, Shanbhag NM. The association of microbiome dysbiosis with colorectal cancer. Cureus. 2022;14:e22156.
47. Tester RF, Al-Ghazzewi FH. Mannans and health, with a special focus on glucomannans. Food Res Int. 2013;50:384-91.
48. Xu D, Ma M, Liu Y, Zhou T, Wang K, Deng Z, et al. PreQ0 base, an unusual metabolite with anti-cancer activity from Streptomyces qinglanensis 172205. Anticancer Agents Med Chem. 2015;15:285-90.
49. Mollick T, Lain S. Modulating pyrimidine ribonucleotide levels for the treatment of cancer. Cancer Metab. 2020;8:12.
50. Mosca L, Pagano M, Pecoraro A, Borzacchiello L, Mele L, Cacciapuoti G, et al. S-Adenosyl-l-methionine overcomes uL3-mediated drug resistance in p53 deleted colon cancer cells. Int J Mol Sci. 2020;22:103.
51. Huh JW, Kim MJ, Kim J, Lee HG, Ryoo SB, Ku JL, et al. Enterotypical Prevotella and three novel bacterial biomarkers in preoperative stool predict the clinical outcome of colorectal cancer. Microbiome. 2022;10:203.
52. Calderon-Perez L, Llaurado E, Companys J, Pla-Paga L, Pedret A, Rubio L, et al. Interplay between dietary phenolic compound intake and the human gut microbiome in hypertension: a cross-sectional study. Food Chem. 2021;344:128567.