Virus Research 355 (2025) 199569

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Evolution of enteric viruses in the progression of colorectal cancer via the adenoma-carcinoma sequence pathway

Ying Yang [a,†], Dan Wang [a,b,†], Longlin Li [a], Jieyu Song [a], Xianglan Yang [c], Jun Li [a,*]

[a] Department of Gastroenterology, The First Affiliated Hospital of Chengdu Medical College, Chengdu, Sichuan, China
[b] Department of Gastroenterology, Pidu District People's Hospital, Chengdu, Sichuan, China
[c] Pengzhou Branch of the First Affiliated Hospital of Chengdu Medical College, Pengzhou Second People's Hospital, Chengdu, China

ARTICLE INFO

Keywords:
Enteric viruses
Adenomatous polyps
Colorectal adenocarcinoma
Adenoma-carcinoma sequence pathway
Metagenomics

ABSTRACT

The global incidence of colorectal cancer (CRC) is increasing. In the majority of CRC cases, colon cancer develops from alterations in the adenoma-carcinoma sequence pathway. Currently, there are few studies regarding the effects of enteric viruses on the adenoma-carcinoma sequence pathway, and subsequently, the progression and development of the CRC. Here, fecal and tissue samples from a normal control group, an adenomatous polyp group, and a colorectal adenocarcinoma group were collected to gain a deeper understanding of the variations in enteric viruses in CRC patients and to analyze their significance. With the progression of CRC from adenoma to adenocarcinoma, the number of DNA viruses in the virus-like particles (VLPs) of fecal and tissue samples gradually increased, and there were distinct differences in the composition of enteric viruses among the different groups. *Multiple species* correlation analysis revealed extensive interactions among viruses, bacteria, and fungi in fecal and tissue samples. Functional analysis also revealed that the functional pathways in fecal and tissue samples also underwent significant changes. In conclusion, the changes in the composition and function of enteric viruses in the progression of CRC via adenoma-carcinoma sequence pathway were analyzed in this study, and these changes hold certain importance for exploring the role of enteric viruses in the occurrence of this disease; however, their mode of action and specific mechanisms require further investigation.

1. Introduction

Colorectal cancer (CRC) is one of the most common cancers worldwide, accounting for 10 % of the global cancer incidence (F et al., 2018). It is the second most common cancer in women and the third most common cancer in men (Dekker, Tanis, Vleugels, Kasi, and Wallace, 2019).

There are three pathways through which CRC develops, namely, the adenoma-carcinoma sequence pathway, the inflammatory pathway, and the de novo pathway, and 70–90 % of CRC cases develop through the adenoma–adenocarcinoma pathway (Dekker et al., 2019). Its multistage development involves a variety of factors, including genetic or environmental factors (C et al., 2022), yet the majority of CRC cases are sporadic (Carethers and Jung, 2015). Dietary factors, smoking, alcohol consumption, and obesity are all known factors for this disease (Moskal et al., 2016). In addition, a growing body of research suggests that alterations in the gut microbiota of CRC may play a role in its occurrence and progression. The majority of the gut microbiome is bacteria (Z et al., 2022), and the most common phylum is Firmicutes. The most researched genus is Bacteroides (Shapira, 2016). Disturbances in the gut microbiota have been associated with a variety of human diseases, with changes in the gut microbiota found in diseases such as inflammatory bowel disease (IBD) (Nishino et al., 2018), metabolic diseases such as obesity and diabetes (Karlsson, Tremaroli, Nielsen, and Bäckhed, 2013), and autoimmune diseases (Gong et al., 2021).

In addition to bacteria, many viruses also colonize the human gut. The number of enteric viruses in adults is estimated to be approximately the same as the number of gut bacteria, with $>10^{12}$ virus-like particles (VLPs) in each person (Shkoporov and Hill, 2019). The human enteric virome includes bacteriophages that only infect bacteria, eukaryotic

* Corresponding author at: Department of Gastroenterology, Clinical Medical College and the First Affiliated Hospital of Chengdu Medical College, Road Baoguang 278#, Region Xindu, Chengdu 610500, China.
E-mail address: 84183967@qq.com (J. Li).
† These authors contributed equally to this work.

https://doi.org/10.1016/j.virusres.2025.199569
Received 13 September 2024; Received in revised form 22 March 2025; Accepted 31 March 2025
Available online 1 April 2025

viruses that only infect eukaryotic cells, and viruses that only infect human cells (G. Liang and Bushman, 2021). With advances in sequencing technology, screening, and culture omics, research on the population and genetic composition of human enteric viruses has increased (Breitbart et al., 2003). In a study of IBD in children, the abundances of Caudovirales and *Microvirdae* were found to differ significantly between IBD patients and controls (Fernandes et al., 2019). Jiang et al. also reported that specific viral taxa, such as Staphylococcus phages and *Herpesviridae*, were associated with increased disease severity in patients with the Alcoholic liver disease (ALD) (Jiang et al., 2020). Differences in enteric viruses have also been found in COVID-19 infection (Zuo et al., 2020), autoimmune diseases (Tomofuji et al., 2022), and malnutrition (M et al., 2020). However, there is insufficient research on enteric viruses in the context of the CRC development via the adenoma-carcinoma sequence pathway.

In this study, we analyzed the changes in fecal and tissue VLPs and used metagenomic sequencing technology to analyze the changes in the composition and function of enteric viruses in the development and progression of CRC via the adenoma-carcinoma sequence pathway and to perform multiboundary species correlation analysis to provide a reference for further exploration of the role of enteric viruses in the CRC adenoma-carcinoma sequence pathway.

2. Materials and methods

2.1. Selection of study participants and collection of samples

This study was approved by the Medical Ethics Committee (Opinion No 2022CYFYIRB-BA-Jul20–01), and all participants signed an informed consent form. From January 2022 to October 2022, stool and tissue samples from healthy individuals ($n = 6$), patients with adenomatous polyps ($n = 10$) and patients with colorectal adenocarcinoma ($n = 12$) were collected via endoscopic biopsy and examined by pathologists at the First Affiliated Hospital of Chengdu Medical College. The inclusion and exclusion criteria for specimen collection in this experiment are as follows.

Inclusion criteria:

(1) Age 18 to 80 years old, with no gender restrictions;
(2) Individuals with no intestinal lesions diagnosed by endoscopy are included in the normal control group;
(3) Those diagnosed with colorectal adenoma through endoscopic biopsy and histopathology are included in the adenomatous polyp group;
(4) Those diagnosed with colorectal adenocarcinoma through endoscopic biopsy and histopathology are included in the colorectal adenocarcinoma group.

Exclusion criteria:

(1) Having undergone bowel preparation within 2 weeks;
(2) Having used acid-suppressing drugs, antibiotics, probiotics, prebiotics, synbiotics, etc. within 3 months;
(3) Having obvious mental disorders that significantly affect the conduct of the study;
(4) Having severe heart disease, lung disease, and digestive system diseases, etc.

The general demographic and clinical information of the study participants was collected (Table 1).There were no significant differences among the groups in terms of gender, BMI, sampling site, abdominal pain, abdominal distension, diarrhea, constipation, and Helicobacter pylori infection. Although the incidence of hematochezia in the AF/AT group was not significantly different from that in the NF/NT group and the CF/CT group, the incidence of hematochezia in the NF/NT group was significantly lower than that in the CF/CT group. The age of the NF/NT group was significantly younger than that of the AF/AT group and the CF/CT group, and the age of the AF/AT group was significantly younger than that of the CF/CT group. Regression analysis was conducted on confounding factors such as age, and the results indicated that age did not significantly interfere with the research outcome.

Table 1
Basic information of the subjects.

Clinical features	NF/NT ($n = 6$)	AF/AT ($n = 10$)	CF/CT ($n = 12$)
Gender (male/female)	3/3	7/3	8/4
Age (years)	32.33±7.31	55.40±9.37	22.76±1.57
BMI (kg/m2)	22.15±1.27	22.76±1.57	22.76±1.57
Sampling site (sigmoid colon/ rectum)	4/2	8/2	4/8
Abdominal pain (%)	0	20	33.3
Abdominal distension (%)	0	30	41.7
Diarrhea (%)	0	30	41.7
Constipation (%)	0	20	25
Hematochezia (%)	0	30	66.7
Prevalence of *Helicobacter pylori* infection (%)	16.7	20	33.3

Abbreviations: NF/NT: feces/tissue of normal control group, AF/AT: feces/tissue of adenomatous polyp group, CF/CT: feces/tissue of colorectal adenocarcinoma group, BMI: Body Mass Index.

2.2. VLP staining

Stool and tissue samples (0.5 g each) were crushed in a mortar for 30–60 s, suspended and mixed with 10 ml of 0.02 μm filtered sterile magnesium salt buffer, and then centrifuged for 10 min (4 °C, 2500 r/min), after which the supernatant was filtered through a syringe filter with pore sizes of 0.45 μm and 0.22 μm, respectively, to remove cells. The filtrate was diluted 10-fold, filtered with a 0.02 μm filter, stained with 10X SYBR Gold (for DNA viruses) and 10X SYBR Green II (for RNA viruses) for 15 min, and then washed. Finally, fluorescence visualization was performed under a microscope.

2.3. Metagenomic sequencing and data processing

Metagenomic sequencing is a high-throughput sequencing method that is used to analyze the genomic material of environmental or clinical samples that contain microbial communities, with a focus on microbial population structure, gene functional activity, the cooperative relationship between microorganisms, and the relationship between microorganisms and the environment.The specific experimental details are as follows: First, genomic DNA was extracted and quality checked: (1) DNA from tissue samples was extracted using the Magen Hipure Soil DNA Kit (Magen, China). Approximately 30 mg of fresh frozen intestinal tissue samples were taken, ground in liquid nitrogen, and transferred to a 2 mL lysis tube. Then, 500 μL Buffer SL and 20 μL proteinase K (20 mg/mL) were added, mixed by vortexing, and incubated with shaking at 65 °C for 1 h, with vortexing every 15 min. Next, 100 μL Buffer SC was added, incubated on ice for 5 min, and centrifuged at 12,000 x g for 5 min. The supernatant was transferred to a new tube and 1.5 times the volume of Buffer SB was added, mixed, and transferred to a Hipure Soil DNA Column, centrifuged at 12,000 x g for 1 min. Then, 500 μL Buffer SW1 and Buffer SW2 were added successively and centrifuged to remove residues. Finally, DNA was eluted with 50 μL Elution Buffer (prewarmed at 65 °C), and the concentration was determined (A260/A280 ratio 1.8–2.0, Qubit 4.0). (2) For fecal sample DNA extraction, the Hipure Stool DNA Mini Kit (Magen, China) was used. First, 200 mg of fecal samples were taken and 1 mL Buffer STL was added, vortexed to homogenize, followed by the addition of 20 μL proteinase K (20 mg/mL) and incubated at 70 °C for 10 min. Then, 200 μL Buffer SP2 was added, mixed by vortexing, and incubated at 95 °C for 5 min. After an ice bath for 2 min, 1 vol of Buffer SI was added, mixed by vortexing, and

centrifuged at 12,000 x g for 3 min. The supernatant was transferred to a new tube and 1.5 times the volume of Buffer SB was added, transferred to a Hipure Stool DNA Column, and DNA was bound by centrifugation. Then, 500 μL Buffer SW1 and SW2 were used for washing twice, and DNA was eluted with 50 μL Elution Buffer, and the concentration and purity were determined (Nanodrop detection A260/A230 $>$ 1.8). Second, library construction: We used the VAHTS Universal Pro DNA Library Prep Kit for Illumina (Vazyme, China) for library construction. The specific steps are as follows: (1) DNA fragmentation: 1 μg of DNA was fragmented to the target length (300–500 bp) using the Covaris M220 ultrasonic instrument. The parameters were set as follows: peak power 50 W, duty cycle 10 %, cycle number 200, and processing time 60 s. (2) End repair and 3′ end A-tailing: Fragmented DNA was mixed with 5 × VAHTS DNA Damage Repair Buffer and VAHTS DNA Damage Repair Enzyme, and incubated at 37 °C for 15 min. Then, dATP and VAHTS A-Tailing Enzyme were added, and the reaction was carried out at 37 °C for 30 min, followed by inactivation at 75 °C for 5 min. (3) Adapter ligation: Mix the end-repaired products with VAHTS Adapter for Illumina (pre-diluted to a final concentration of 0.6 μM) and T4 DNA Ligase, and incubate at 20 °C for 15 min. (4) Magnetic bead purification and fragment selection: Purify using VAFSTM DNA Clean Beads (Vazyme, China). Mix the DNA and magnetic beads at a volume ratio of 1:0.8 and incubate at room temperature for 5 min to remove short fragments ($<$ 300 bp). After removing the supernatant, add magnetic beads at a volume ratio of 1:1 with the DNA and incubate at room temperature for 5 min to elute the target fragments (300 - 700 bp). Finally, elute the DNA with 20 μL EB buffer and measure the concentration (Qubit 4.0 Fluorometer). (5) PCR amplification: Amplify using VAHTS Pro DNA Library PCR Mix: Use VAHTS Universal Primer and Index Primer for 8 cycles of amplification. (6) Library quality control and quantification: First, perform fragment distribution detection: Use Agilent 2100 Bioanalyzer (High Sensitivity DNA chip) to confirm that the main peak of the library is within the range of 400 - 500 bp. Then, measure the concentration: Use Qubit dsDNA HS Assay Kit (Thermo Fisher) to ensure that the library concentration is $\geq$ 2 ng/μL. Finally, sequence on the Illumina Novaseq 6000 (PE150) platform.

The raw image data of the sequencing results are processed by the software Bcl2fastq (v2.17.1.14) for base recognition and preliminary quality analysis to obtain the raw sequencing data (Pass Filter Data). The raw sequencing data are then processed by the second-generation sequencing data quality statistics software cutadapt (v1.9.1) to remove adapters and low-quality sequences, etc. (removing primer and adapter sequences; removing bases with quality values lower than 20 at both ends; removing sequences with N base content greater than 10 % and retaining the minimum read length of 75 bp), to obtain the Clean Data for subsequent information analysis.

Based on the optimized Clean Data, assembly analysis is performed using the MEGAHIT (v1.1.3) software. MEGAHIT is suitable for the assembly of large or complex metagenomic data and achieves low-memory assembly by constructing a compact de Bruijn graph. For each sample, de Bruijn graphs and pre-assemblies are constructed with different K-mers (59, 79, 99, 119, 141), and the results are combined to obtain the optimal assembly result for each sample.

The coding genes are predicted using the Prodigal (v3.02) software, and the gene sequences of all samples are integrated. Further redundancy removal is performed using the sequence clustering software MMseq2 with default parameters of identity 95 % and coverage 95 % to obtain the non-redundant gene set unigene sequences. The preprocessed Clean reads are aligned to the constructed non-redundant gene set unigene sequences using the SoapAligner (version 2.21) alignment software to obtain the number of reads aligned to each unigene in each sample. Then, based on the number of reads aligned to each unigene and the gene length, the abundance information of each unigene in each sample is calculated.

2.4. Species and gene functional annotation

The NR database (Non-Redundant Protein Database, NR) is a non-redundant protein database. Its characteristic lies in its comprehensive coverage, and the annotation results include species information, which can be used for species classification.The Diamond program was used to compare the unigene sequences obtained by metagene sequencing with the NR database, and species annotation of each sequence was performed by obtaining the taxonomic annotation information corresponding to each sequence in the NR database to explore the microbial composition of the samples. Draw the distribution of the top 30 species at the genus and family levels for each sample as bar charts.

The KEGG database (Kyoto Encyclopedia of Genes and Genomes, KEGG) is a database for systematically analyzing gene functions and genomic information. It integrates information from genomics, biochemistry, and systems functional genomics, and maps corresponding biological pathways according to different types of biological processes. Through annotation analysis of genes, one can view the biological pathways and corresponding biological functions that the genes are involved in. On the basis of the predicted protein sequence of the coding gene, BLAST software (version 2.2.31+) was used to compare the protein sequences in the KEGG database. The E value of the sequence alignment was set as 1e-5, and the optimal matching result was selected as the annotation result of the gene.

2.5. Composition and differential analysis of species and functions and multispecies coabundance analysis

The species and functions of the samples were compared, and the unique and shared species and functions among the groups are presented in the form of bar graphs and Venn diagrams. Analysis of similarities (ANOSIM) was used to test whether the differences among groups were significantly greater than the differences within groups to determine whether the grouping was meaningful. The microbial communities whose abundance significantly differed between two groups were detected at the genus level via MetaStat software. Principal component analysis (PCA) was used to analyze the species community compositions of different samples, reflecting the similarities and differences between samples. The species found within in the samples were visualized in multidimensional space in the form of points through nonparametric multidimensional scaling (NMDS), and the degree of difference between different samples was reflected through the distance between points. KEGG metabolic pathway enrichment analysis was used to identify the enriched pathways in the differentially expressed genes (DEGs) between two groups to determine the functions of these genes in the body. Multispecies coabundance analysis was conducted via Spearman correlation to calculate species correlations at the genus level. The results were screened according to a cor value $>$ 0.6 and a P value $<$ 0.0001, and Cytoscape (v3.6.1) was used for network mapping.

2.6. Statistical analysis

SPSS 26.0 software was used for statistical analysis. Qualitative data were compared between groups using the chi-square test or Fisher's exact probability test.The quantitative data that conformed to a normal distribution are presented as the mean ±standard deviation (SD), and the overall differences among the three groups were compared via ANOVA. For quantitative data with a nonnormal distribution, the Kruskal–Wallis rank sum test was used to analyze the overall differences among the three groups. In the sophisticated differential analyses of the enteric virus composition and the gene functional annotation analysis, as the majority of the data presented a skewed distribution, the Kruskal–Wallis rank sum test was employed to analyze the differences among the three groups, followed by the utilization of the Bonferroni correction method to analyze the differences between two groups. A statistically significant difference was defined as $P < 0.05$ ($^{*}p < 0.05$, $^{**}p < 0.001$).

3. Results

3.1. The number of DNA viruses in the intestine was higher in both of the CRC groups than the healthy control group

VLPs were isolated from the feces and tissues of different groups, and DNA viruses and RNA viruses were stained and observed (Fig. 1A & 1B). The number of DNA viruses in feces and tissues was significantly higher in the disease groups (adenoma group and adenocarcinoma group) than in the normal group ($P < 0.05$), whereas there was no significant difference in the number of RNA viruses among the three groups ($P > 0.05$) (Fig. 1C & 1D). Therefore, during the development and progression of CRC, the number of DNA viruses in fecal and tissue VLPs gradually increase, while the number of RNA viruses were relatively stable. Thus, metagenomic sequencing was selected for subsequent analysis.

3.2. The composition of enteric viruses in feces and tissues significantly changed at various species levels during the development and progression of CRC via the adenoma-carcinoma pathway

The viruses in the three fecal groups mainly belonged to the Uroviricota phylum (Fig. 2A & 2C), and there was no significant change in viral abundance among the three fecal groups ($P > 0.05$). In tissues, viruses mainly belonged to the Artverviricota phylum (Fig. 3A & 3C). Compared with that in the tissue of adenomatous polyp group (AT), the relative abundance of Artverviricota viruses in that of the colorectal adenocarcinoma group (CT) was significantly lower ($P = 0.036$), but there was no significant difference compared with that in the healthy control group (NT).

Viruses belonging to the Limestonevirus genus had the highest abundance in the feces among the three groups (Fig. 2B & 2D). The abundances of Coetzeevirus and Tequatrovirus were detected only in the fecal samples of the healthy control group (NF). During the progression of CRC via the adenoma-carcinoma pathway, the proportions of Hendrixvirus and Ravinvirus gradually decreased, and the proportion of Roufvirus gradually increased. Gammaretrovirus was the most abundant virus in the tissue of each group at the genus level (Fig. 3B & 3D). The proportion of Roufvirus gradually increased during the progression of CRC. At the genus level, the composition and number of common and different enteric viruses among the groups were determined. Thirteen viral genera were common in the feces among the three groups, and the AF had the greatest number of endemic genera (24) (Fig. 2E). There were 33 common virus genera in the tissue samples among the three groups, with no unique viruses in the NT and only 1 genus of viruses unique to the AT and CT (Fig. 3E).

At the species level, the abundance of Lactobacillus virus phiJL1 was detected only in the NF. Compared with those in the AF, the relative abundances of Lactobacillus phage LF1 ($P = 0.027$) and Lactobacillus phage LfeSau ($P = 0.027$) in the CF were significantly lower (Supplementary Figure 1A). The abundance of human betaherpesvirus 6A in the AT ($P = 0.000$) and CT ($P = 0.014$) was significantly lower than that in the NT (Supplementary Figure 1B).

3.3. Pronounced variations in the abundance of several enterovirus species in the feces and tissues of different groups were detected

ANOSIM revealed that there was no significant difference in the abundance of enteric viruses in feces ($R = 0.03$, $P = 0.282$) or tissues ($R=-0.008$, $P = 0.472$) among the three groups (Supplementary Figure 2A & 2C). MetaStat analysis suggested that there were five species with significant differences in abundance between the NF and AF at the genus level, including Gammaretrovirus and Jedunavirus (Fig. 4A); four species with significant differences in abundance between the NF and CF, including Hendrixvirus and Ravinvirus (Fig. 4B); and six species

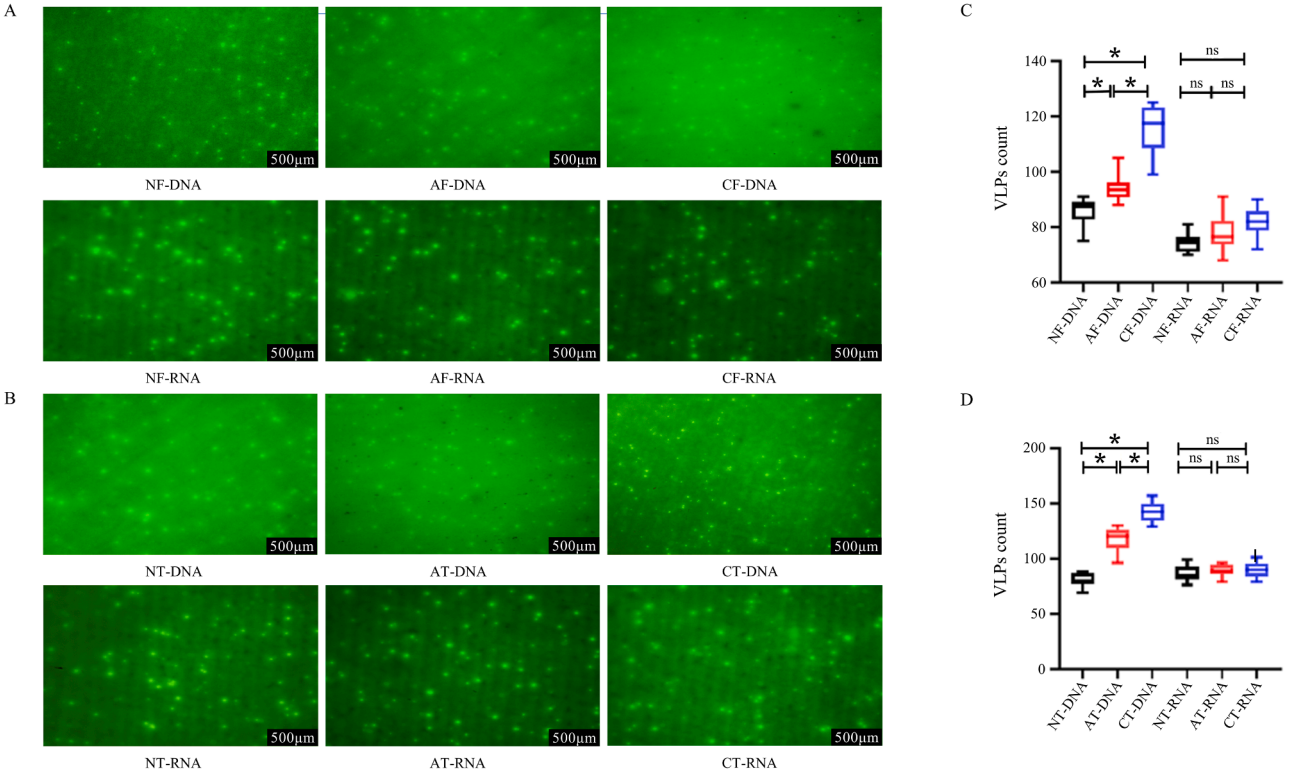

Fig. 1. VLPs in feces and tissues of different groups. (A), (B) VLPs were observed under fluorescence microscope after staining of feces and tissue filtrates of different groups with SYBR Gold and SYBR Green II. (C), (D) The number of VLPs in different groups and its variance analysis. * $P < 0.05$, ** $P < 0.01$,ns: non-significant.1. VLPs: virus-like particles,NF/NT: feces/tissue of normal control group, AF/AT: feces/tissue of adenomatous polyp group,CF/CT: feces/tissue of colorectal adenocarcinoma group.

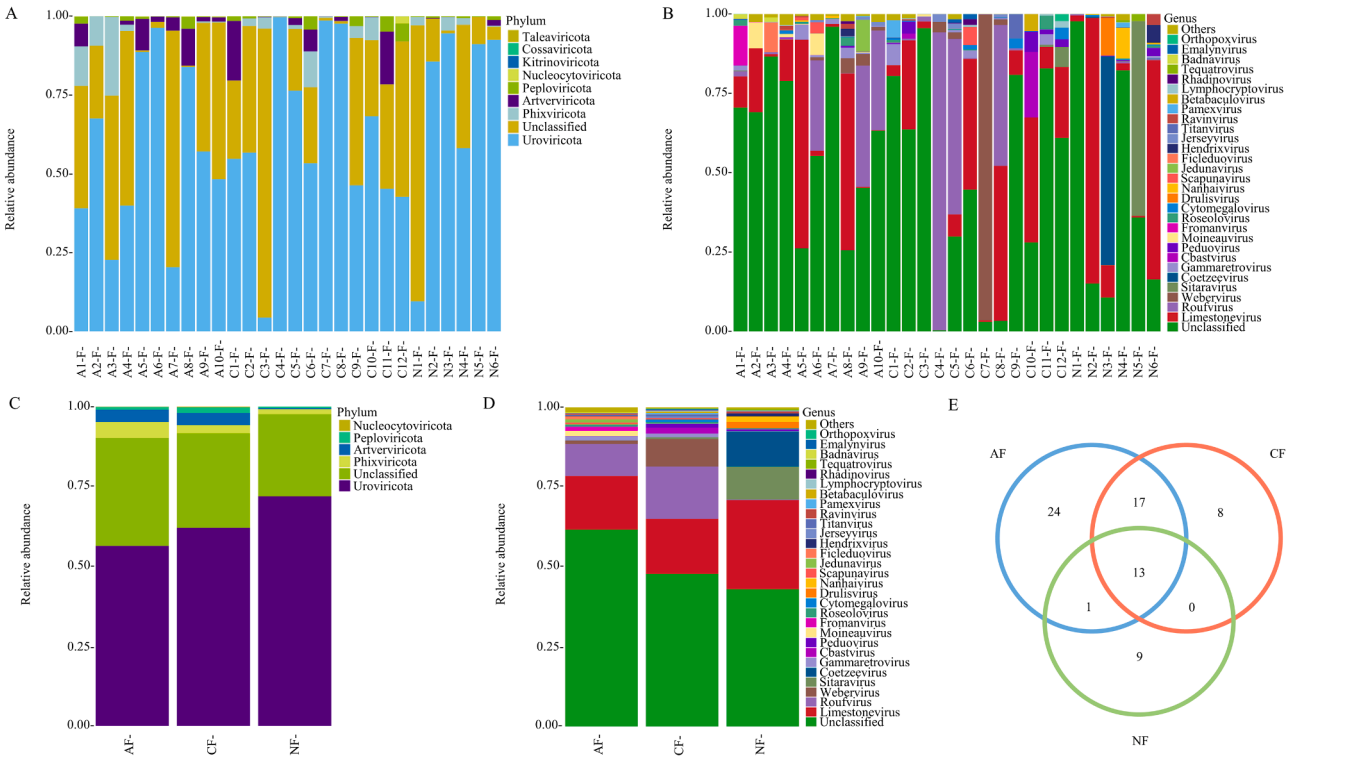

Fig. 2. Distribution of fecal enteric viruses. (A), (B) Distribution of enteric viruses in each fecal sample at the phylum and genus levels. (C), (D) Distribution of enteric viruses among different fecal groups at the phylum and genus levels. (E) The Venn diagram of enteric viruses among fecal groups at genus level.NF/NT: feces/tissue of normal control group, AF/AT: feces/tissue of adenomatous polyp group,CF/CT: feces/tissue of colorectal adenocarcinoma group.

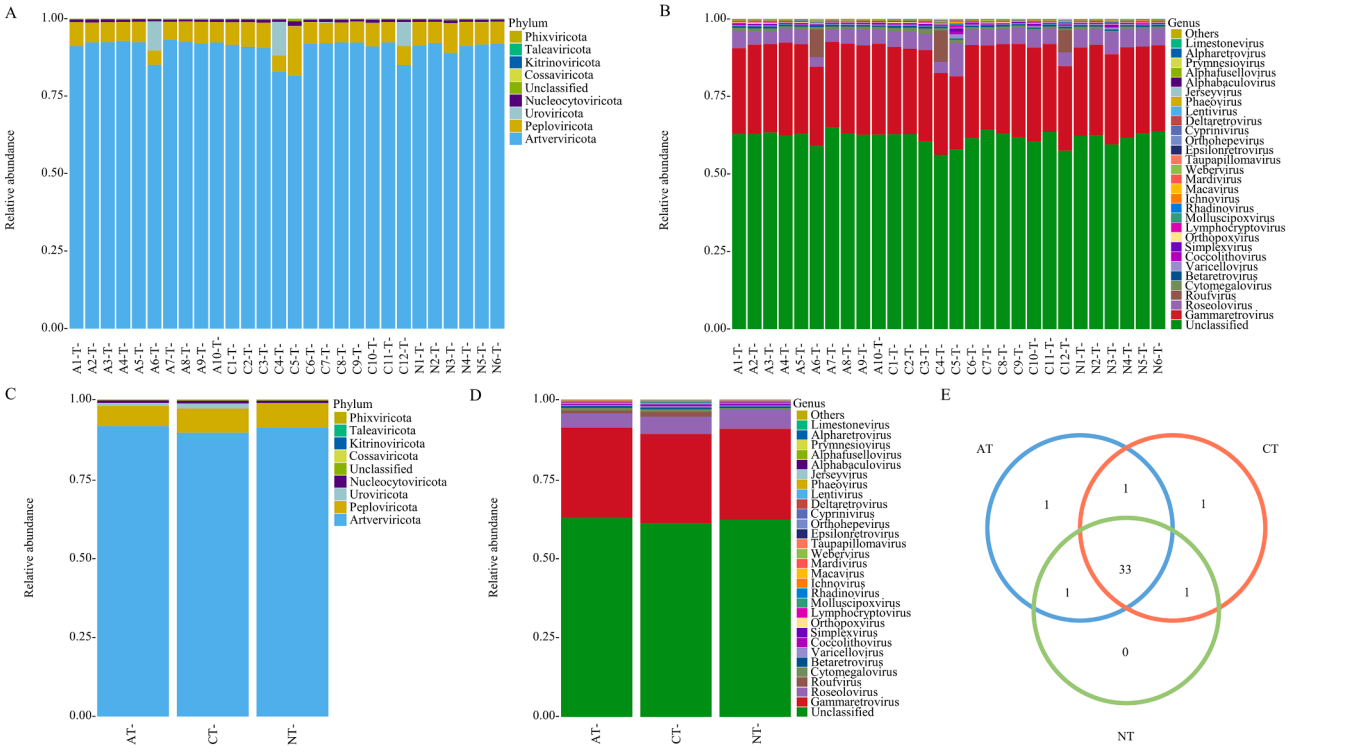

Fig. 3. Distribution of enteric viruses within tissues. (A), (B) Distribution of enteric viruses in each tissue samples at the phylum and genus levels. (C), (D) Distribution of enteric viruses among tissue groups at the phylum and genus levels. (E) The Venn diagram of enteric viruses among tissue groups at the genus level.NF/NT: feces/tissue of normal control group, AF/AT: feces/tissue of adenomatous polyp group,CF/CT: feces/tissue of colorectal adenocarcinoma group.

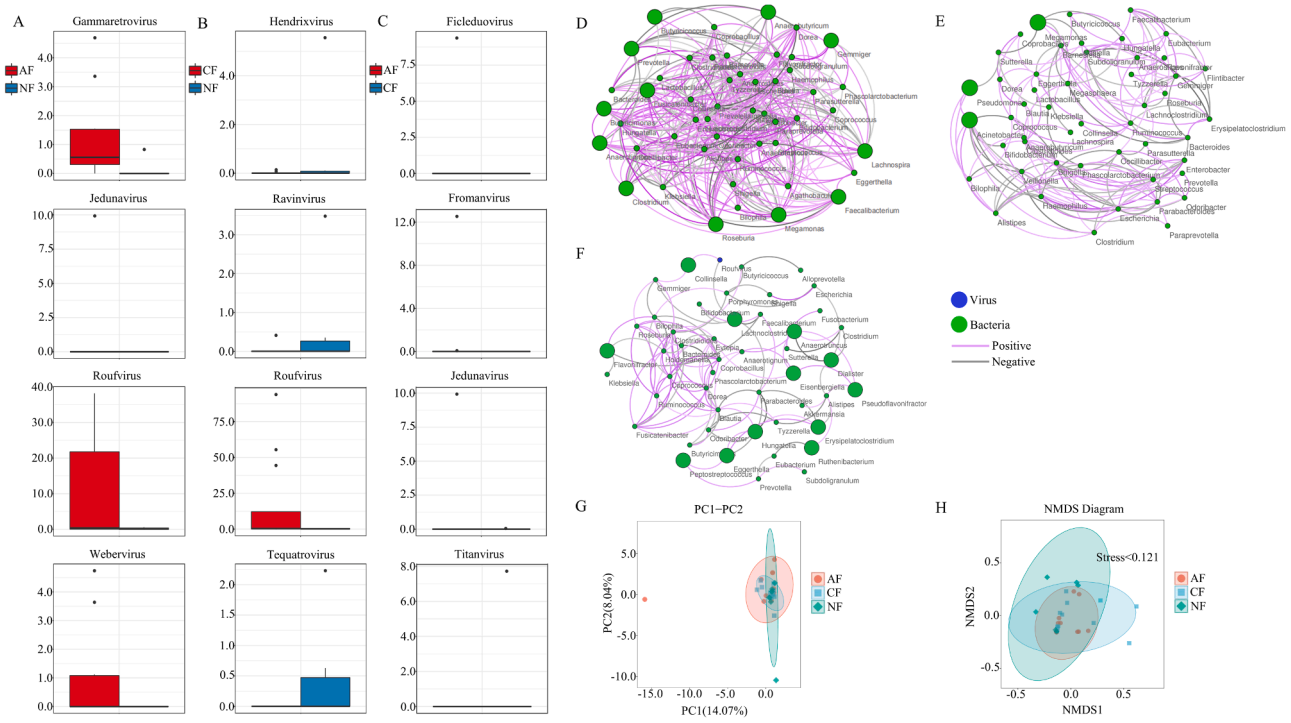

Fig. 4. Enteric viruses difference analysis and Co-abundance correlations among multi-kingdom species analysis in different feces groups. (A), (B), (C) Metastat analysis of inter-group differences of fecal enteric viruses in the genus level. The top 4 distinct species with relative abundance were chosen to draw the box map. The abscissa was the species taxonomy name, and the ordinate was the relative abundance of the species. Co-abundance correlations among multi-kingdom species of the (D)NF group, the (E)AF group, and the (F)CF group.)The color of the node represents the species from bacteria (green) and viruses (blue). The purple line represents positive species interactions, and the gray line represents negative species interactions. (G) Comparison of fecal enteric viruses by PCA analysis. PC1 and PC2 respectively represent the first and second principal components, and the percentage subsequent to the principal component indicates the contribution rate of this component to sample dissimilarities. The closer the sample points are, the greater the similarity of species distribution within the sample. (H) Comparison of fecal enteric viruses by NMDS analysis. The distance between the points represents the degree of difference. When the stress was <0.2, it indicated that NMDS could accurately reflect the degree of difference between samples.

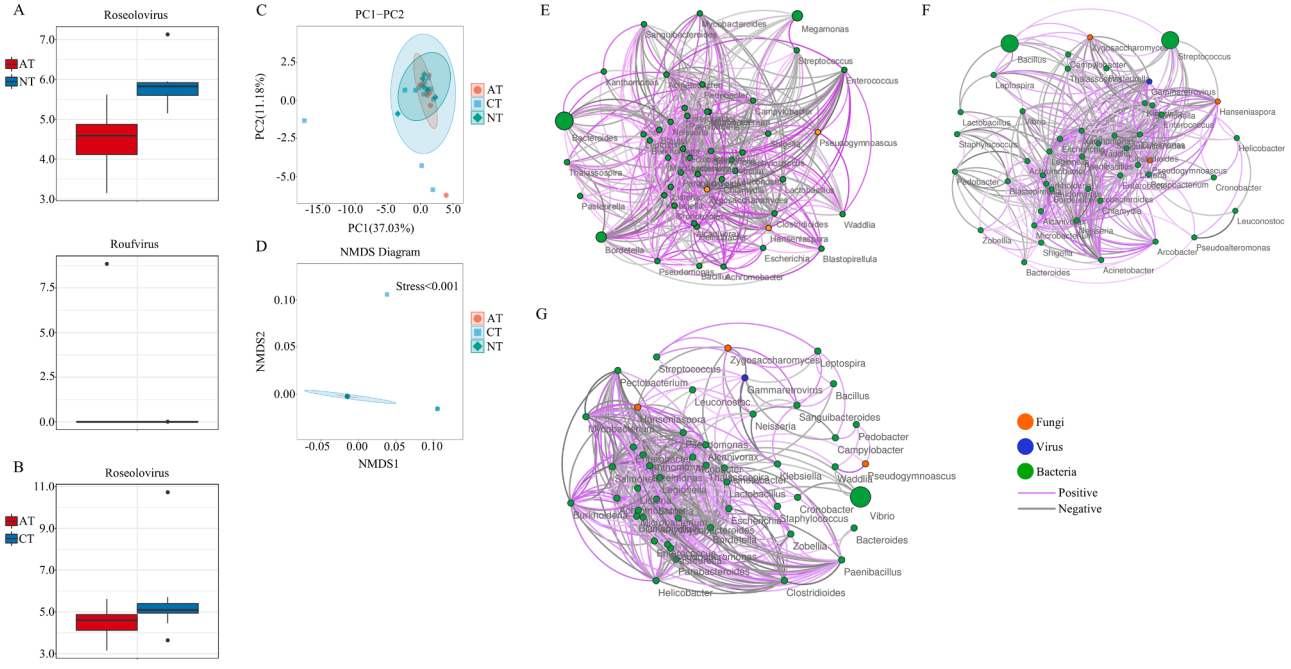

Fig. 5. Analysis of the enteric viruses differences and Co-abundance correlations among multi-kingdom species among tissue groups. (A), (B) Metastat analysis of the inter-group differences of tissue enteric viruses in the genus level, and disparate species were selected to draw the box-type plots. (C), (D) Comparison of enteric viruses in tissues by PCA/NMDS. Co-abundance correlations among multi-kingdom species of the (E)NT group, the (F)AT group, and the (G)CT group. The color of the nodes represents species from bacteria (green), fungi (orange), and viruses (blue).

with significant differences in abundance between the AF and CF, including Jedunvirus and Fromanvirus (Fig. 4C). MetaStat analysis revealed that there were two species whose abundance significantly differed between the NT and AT: Roseolovirus and Roufvirus (Fig. 5A); however, no species differed between the NT and CT, and only Roseolovirus differed in abundance between the AT and CT: (Fig. 5B). The abundance of Roseolovirus decreased first but then increased in during the progression of CRC via the adenoma-carcinoma pathway. Next, PCA and NMDS were used to compare fecal (Fig. 4G & 4H) and tissue (Fig. 5C & 5D) enterovirus spcies among the three groups, and the analysis results revealed that there was no significant difference in the samples among the three groups.

3.4. Variations in microbial interactions were identified during the progression of CRC via the adenoma-carcinoma sequence pathway

The virus–bacteria–fungi coabundance correlation analysis was performed at the genus level. Only bacterial interactions were screened in the NF (Fig. 4D) and AF (Fig. 4E), and bacterial interactions were significantly reduced in the feces of patients with adenomatous polyps and adenocarcinoma. Moreover, interactions between Roufvirus, Collinsella, and Lachnoclostridium were found in the CF (Fig. 4F):.

Only bacterial interactions were screened in the NT (Fig. 5E), while associations between viruses, bacteria and fungi were also found in the AT (Fig. 5F) and CT (Fig. 5G). There were interactions between Gammaretrovirus and different bacteria (such as Enterobacter, Enterococcus, Clostridioides, Escherichia, Klebsiella and other 14 species) and between Gammaretrovirus and different fungi (Hanseniaspora, Pseudogymnoascus, Zygosaccharomyces) in the AT. In the CT, there was a correlation between Gammaretrovirus and different bacteria (eight species, such as Klebsiella, Neisser, etc.) and different fungi (Hanseniaspora, Zygosaccharomyces).

3.5. Significant differences in the KEGG pathway enrichment analysis results were identified among the groups

At KEGG level 2, the functional pathways associated with the DEGs in the NF were mainly related to cancer, whereas the functional pathways associated with the DEGs in the AF and CF were associated with DNA replication and repair, with no significant differences among the three groups (Fig. 6A & 6C). Cancer-related pathways were mainly enriched in the DEGs identified the tissue samples among the groups (Fig. 7A & 7C). Lipid metabolism was significantly enriched in the downregulated DEGs between the NT and the AT ($P = 0.003$) and in the upregulated DEGs between the NT and CT ($P = 0.041$). Moreover, the upregulated DEGs between the CT and AT were significantly enriched in DNA replication and repair ($P = 0.022$).

At KEGG level 3, cysteine and methionine metabolism was the main enriched pathway in the NF genes, whereas the genes in the AF and CF mainly were involved mainly in DNA replication, with no significant differences among the groups (Fig. 6B & 6D). The major functional pathways involved in all the tissue groups were pathways associated with cancer (Fig. 7B & 7D). Tryptophan metabolism was significantly enriched in the downregulated DEGs between the CT and AT ($P = 0.037$). Then, the number of shared and unique enriched pathways among groups was determined, and a Venn diagram was produced. There were 31 common pathways among the DEGs identified from the fecal samples of the three groups, and the AF had the largest number of unique pathways (70) (Fig. 6E). There were 173 common pathways among the DEGs identified from the tissue samples of the three groups, and no unique pathways were found in the NT, AT or CT (Fig. 7E).

3.6. Significant changes in the functional pathways associated with enteric viruses were present among the groups

ANOSIM analysis revealed significant differences among the fecal samples of the three groups ($R = 0.115$, $P = 0.045$) but no significant differences among the tissue samples of the three groups ($R = -0.026$, P

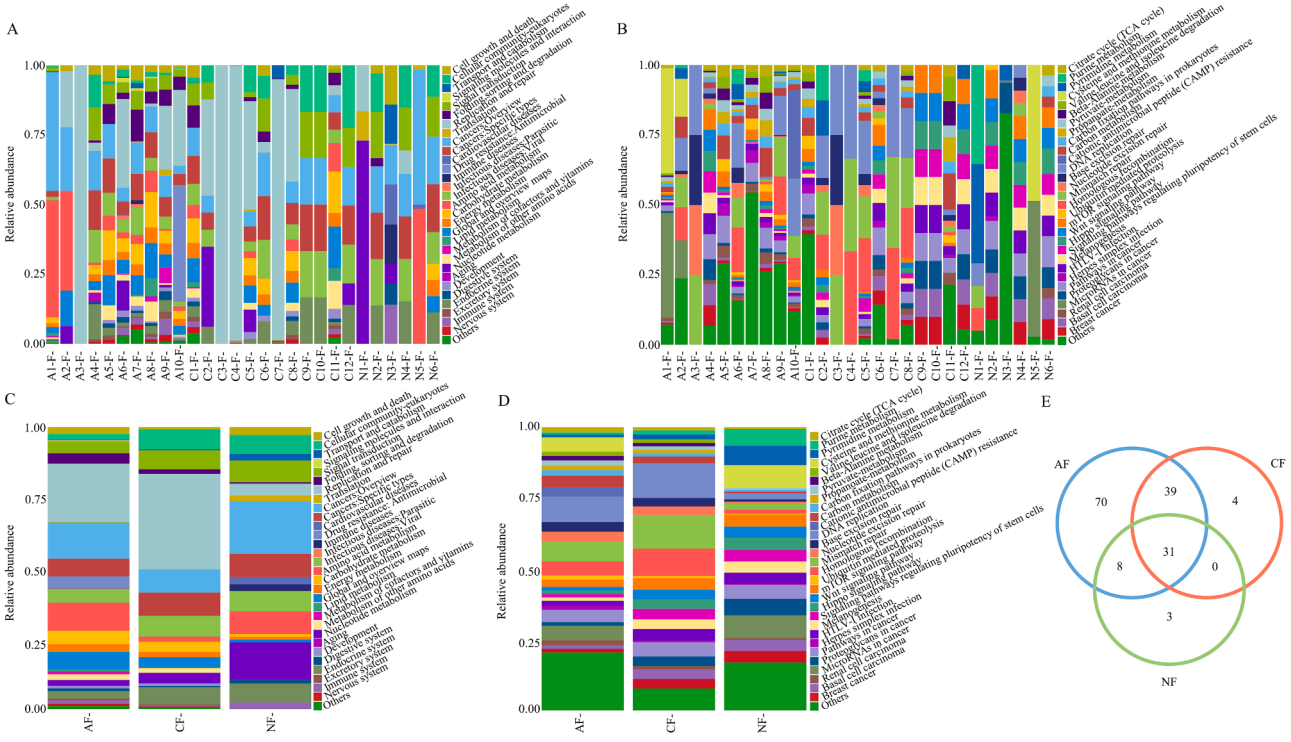

Fig. 6. Relative abundance of functional annotations in the KEGG database (feces). (A), (C) Functional distribution of each fecal sample and different fecal groups at level 2. (B), (D) Functional distribution of each fecal sample and group at level 3. (E) Veen diagram of KEGG database function annotation at level 3.

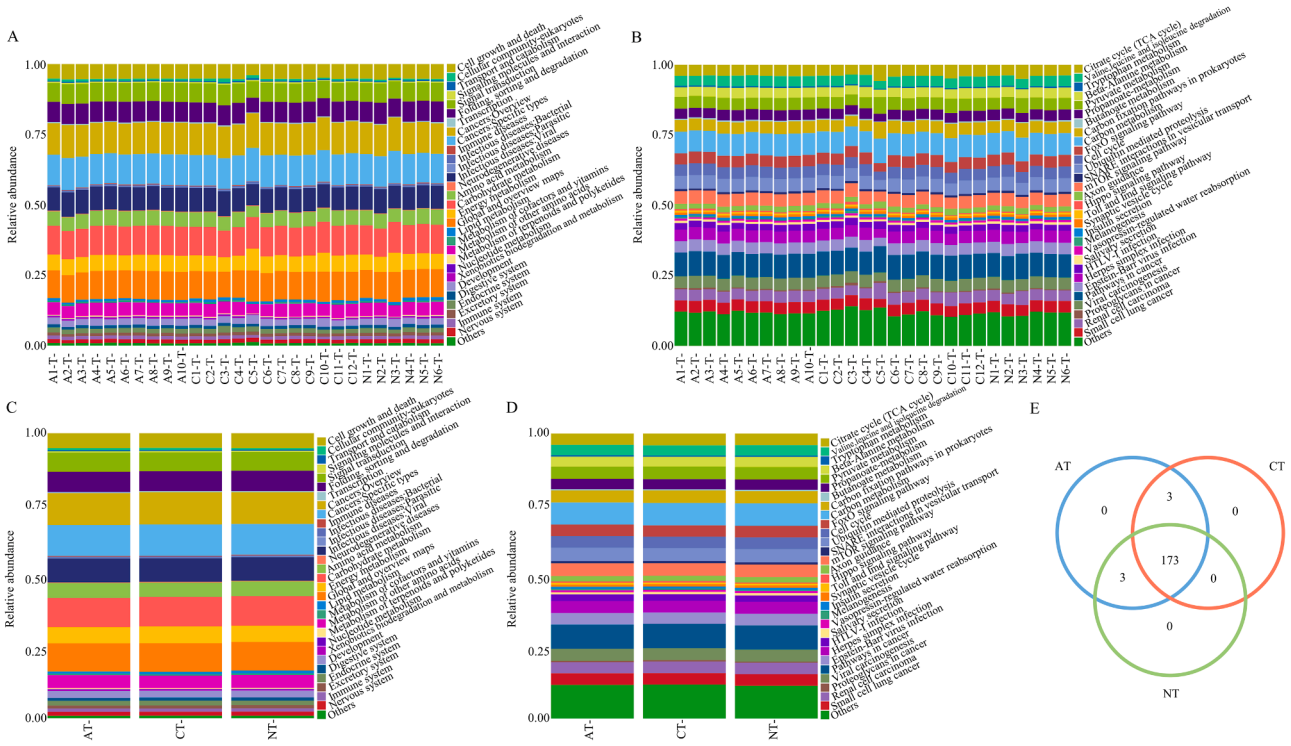

Fig. 7. Relative abundance of functional annotations in the KEGG database (tissues). (A), (C) Functional distribution of each sample and different groups of tissues at level 2. (B), (D) Functional distribution of each tissue sample and group at level 3. (E) Veen diagram of KEGG database function annotation at level 3.

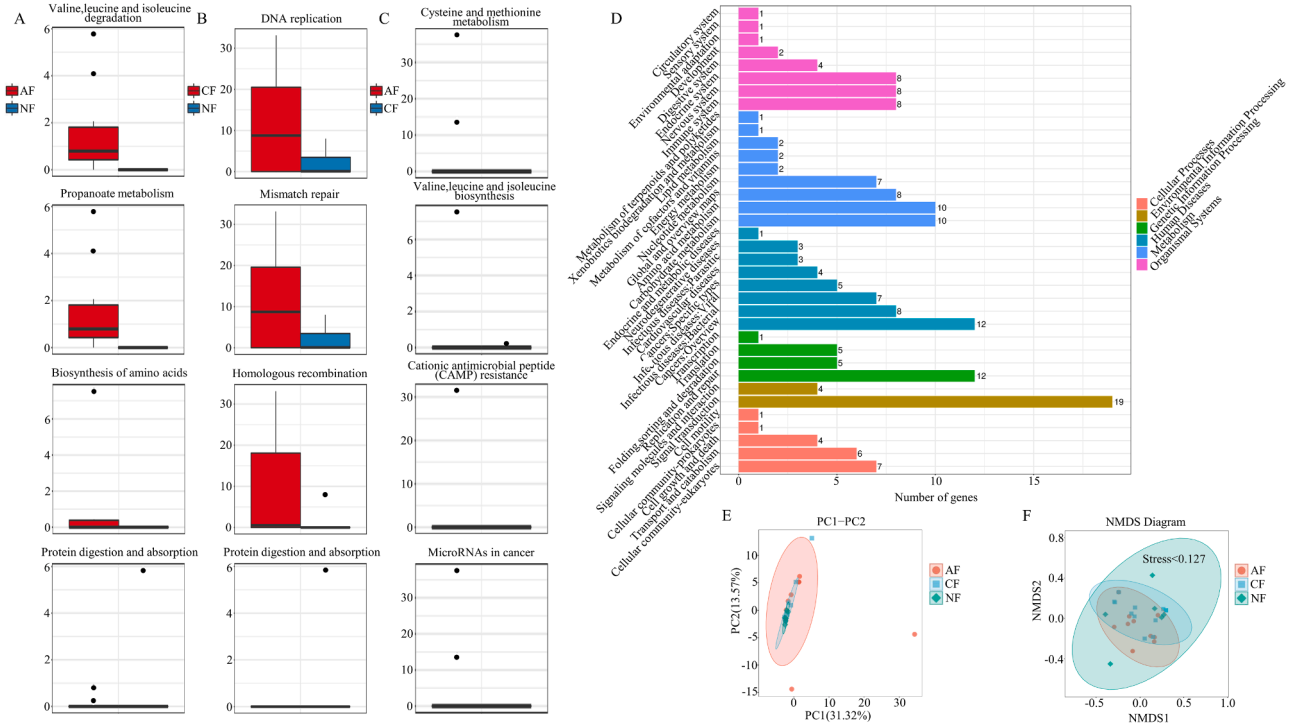

Fig. 8. Functional disparities among fecal groups. (A),(B),(C)Metastats analysis among groups for comparing functional differences, and the top 4 differential pathways with the highest abundance were chosen to draw box diagrams. The horizontal coordinate represents the path name, and the vertical coordinate indicates the relative abundance of the path. (D) KEGG-enrich analysis between AF group and CF group. The vertical coordinate indicates the pathway name, the horizontal coordinate represents the number of significantly different genes, and distinct colors are used to distinguish the first-level classification of biological pathways. (E), (F) PCA/NMDS analyzes differences among fecal groups based on KEGG database.

= 0.616) (Supplementary Figure 2B & 2D). MetaStat analysis revealed nine significantly enriched pathways, such as valine, leucine and isoleucine degradation, in the DEGs between the NF and AF (Fig. 8A), four significantly enriched pathways in the DEGs between the NF and CF, including DNA replication (Fig. 8B), and nine significantly enriched pathways, including MicroRNAs in cancer in the DEGs between the AF and CF (Fig. 8C). In the fecal samples among groups, with the development and progression of CRC, the protein digestion and absorption pathway was gradually suppressed, whereas DNA replication and the mismatch repair pathway were gradually activated. Amino acid biosynthesis and the propanoate metabolic pathway were activated in the disease groups compared with the healthy group, but the differences between the AF and CF were not significant. MetaStat analysis revealed that there was only one significantly different pathway between the NT and AT, HTLV-1 infection, which was activated in the AT (Fig. 9A). There was no significant difference between the NT and CT or between the AT and CT.

Pairwise comparison of the KEGG enrichment analysis results. The most enriched pathway of the DEGs between the NF and CF was replication and repair (Supplementary Figure 3A). There were 36 enriched pathways in the 184 DEGs between the AF and CF, which were mainly associated with signal transduction, and replication and repair (Fig. 8D). Signal transduction was the most enriched pathway in the DEGs between the NT and CT (Supplementary Fig. 3B), whereas 43 pathways were enriched in the 309 DEGs between the AT and CT, which included signal transduction, cancer-related pathways and other pathways (Fig. 9E).

PCA and NMDS analysis revealed that there were no significant differences in terms of the functions of the DEGs in the fecal (Fig. 8E & 8F) and tissue samples among the three groups (Fig. 9B & 9C).

4. Discussion

Viruses are the most abundant organisms on Earth and can infect various life forms. Over the past few decades, microorganisms that are difficult to isolate and grow in the laboratory have been able to be studied with the development of metagenomics (Sudarikov, Tyakht, and Alexeev, 2017). However, few metagenomic studies on the relationship between enteric viruses and CRC exist. In this study, stool and tissue samples were collected simultaneously from 12 patients with CRC, 10 patients with adenomatous polyps, and 6 healthy people. The research findings indicate that the abundance of DNA viruses in feces and tissue VLPs was increased in the disease groups compared to the healthy group, whereas the abundance of RNA viruses was relatively stable, which preliminarily confirmed the existence of viruses in feces and tissues and their changes in abundance in disease states. The objective of VLP staining is to compare the differences in viral content along the CRC adenoma pathway rather than to quantitatively determine the absolute viral content. By making relative comparisons among the three groups, the trend of viral dynamic changes can be indirectly reflected. Therefore, traditional positive and negative controls were not set up. However, the absence of controls may lead to potential non-specific signal interference (such as residual host free nucleic acids or bacterial DNA/RNA), and the experimental results still need to be comprehensively interpreted in combination with molecular biological methods (such as sequencing).Therefore, metagenomic sequencing analysis was subsequently carried out to reveal additional changes in enteric viruses and related functions in CRC patients to provide a theoretical basis for the prevention and treatment of CRC.

At the genus level, tequatrovirus was only abundant in the NF. MetaStat analysis revealed that the viruses that differed in abundance between the NF and AF included Gammaretrovirus,; the viruses that significantly differed in abundance between the NF and CF included Hendrixvirus, and Ravinvirus; and the viruses that differed in abundance between the AF and CF were Jedunavirus and Fromvirus. Moreover, the viruses that significantly differed in abundance between the NT and AT were Roseolovirus and Roufvirus, whereas the virus that differed in abundance between the AT and CT was Roseolovirus. The results suggested that the changes in enteric viruses in the fecal samples were more significant than those in the tissue samples, possibly due to the contamination of microbial samples by host tissue DNA (Nakatsu et al., 2018). However, there is no technique that completely prevents

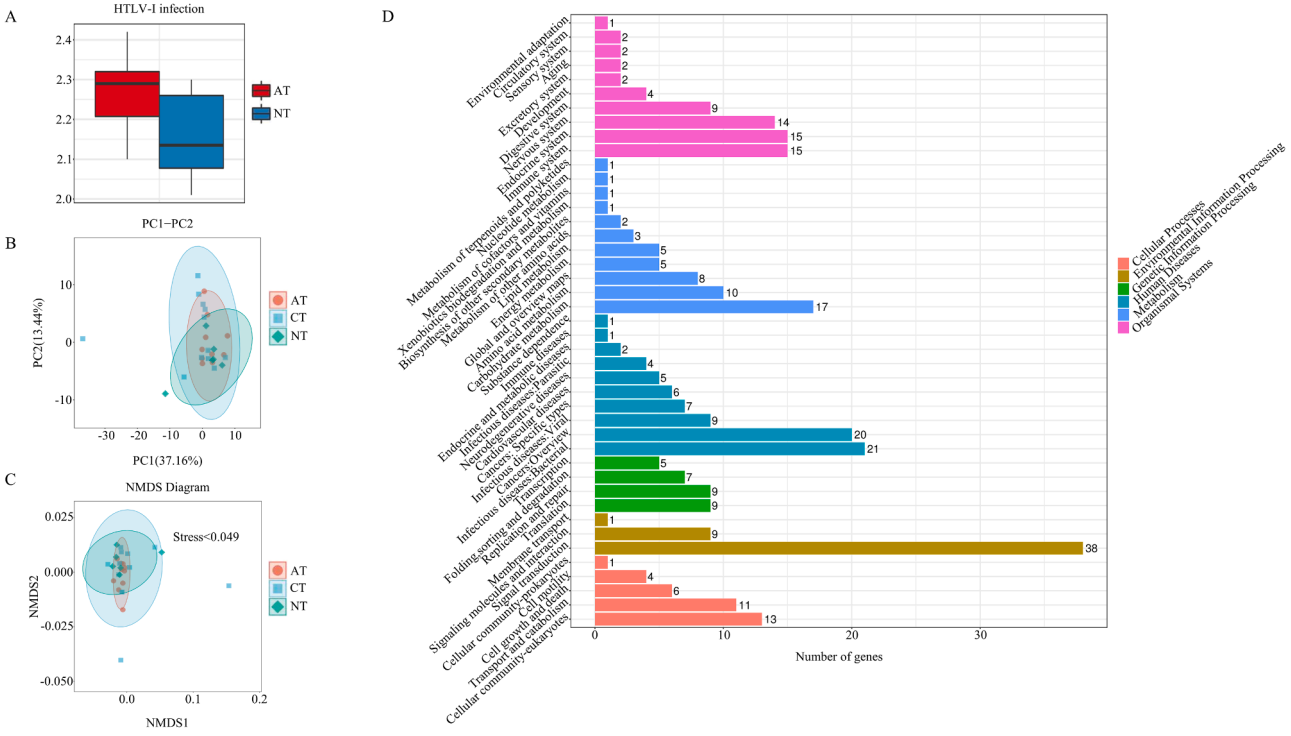

Fig. 9. Functional differences among tissue groups. (A) Metastats analysis to compare the inter-group differences, and draw the box diagram for the different path. (B), (C)PCA/NMDS analyzes differences among tissue groups based on KEGG database. (D) KEGG-enrich analysis between the AT group and the CT group.

contamination, and it is inconclusive whether the changes in the fecal viral community reflect the changes in the tissue microbial community.

According to the latest international classification and nomenclature principles of viruses (Lefkowitz et al., 2018), the above viruses belong to the Caudovirales order, with the exceptions of Gammaretrovirus and Roseolovirus. Among them, Hendrixvirus, Ravinvirus, Roufvirus, and Fromanvirus belong to the *Siphoviridae* family, whereas Tequatrovirus and Jedunavirus belong to the *Myoviridae* family.

Benler et al. reported that 78 % of intestinal phages belong to the Caudovirales order (Benler et al., 2021). Hannigan et al. reported that the viruses with altered microbial compositions in the fecal samples of CRC patients were mainly phages from the Siphoviridae and Myoviridae families (Hannigan, Duhaime, Ruffin, Koumpouras, and Schloss, 2018). This finding is consistent with our results. In addition, the depolymerase (Dep_kpv79) encoded by Klebsiella phage kpv79 in the genus Jedunavirus is a specific β-galactosidase that can cut *Klebsiella pneumoniae* K57-type CPS through a hydrolytic mechanism, which can combat sepsis and hip joint infection caused by *Klebsiella pneumoniae* in a lethal mouse model (M. Gao et al., 2022), and is a promising alternative antibacterial agent (V Volozhantsev et al., 2020). However, this virus has not yet been studied in the context of cancer.

Gammaretrovirus is a single-stranded RNA virus that is classified into the *Retroviridae* family. Currently, gammaretroviruses such as murine leukemia viruses (MLVs) and feline leukemia viruses (FLVs), have been identified. MLVs can cause leukemia in mice through early bone marrow progenitor cells (Tumas-Brundage, Garret, Blank, and Prystowsky, 1996), and FLVs can cause cancer, bone marrow diseases and immunosuppression in cats (Hartmann and Hofmann-Lehmann, 2020).

Roseoloviruses belong to the *Herpesviridae* family, and the typical representative of this genus is human herpesvirus 6 (HHV-6), which includes HHV-6A and HHV-6B (Krug and Pellett, 2014). Roseoloviruses are cytotropic and can establish a lifelong incubation period in the human body (Agut, Bonnafous, and Gautheret-Dejean, 2016). The virus can downregulate host cell-derived MHC Ia and MHC Ib molecules to trigger immune evasion by blocking MHC Ia in the endoplasmic reticulum and directing it to lysosomes for degradation (Krug and Pellett, 2014). In immunocompromised individuals, Roseolovirus can cause serious complications, such as encephalitis and pneumonia. In patients with organ and hematopoietic stem cell transplantation, infection or reactivation of Roseolovirus can lead to organ rejection and graft-versus-host disease (Dulery et al., 2012). In addition, Roseolovirus is closely related to Alzheimer's disease, hepatitis, colitis and other diseases (De Bolle, Naesens, and De Clercq, 2005). However, whether Roseolovirus plays a role in the occurrence and development of CRC needs further study. Furthermore, the remaining viruses listed above have not yet been characterized in the context of disease and intestinal microbiota.

In this study, the abundance and composition of these enteric viruses in feces and tissues changed during the development and progression of CRC, suggesting that enteric viruses may potentially contribute to the occurrence and development of CRC. However, whether enteric viruses affect the progression of CRC, which specific viruses are expected to be potential biomarkers for CRC, and whether the mechanisms of action of viruses in feces and tissues differ need further study.

Fungi, bacteria, and viruses coexist in the gut, and their patterns of interaction in the colon may change during disease states and may reflect their potential role in CRC. Therefore, we conducted a multi-species coabundance analysis and revealed that there is a complex ecological network among intestinal microorganisms, which may be damaged by the occurrence of CRC. In the CF, Roufvirus interacted with Collinsella and Lachnoclostridium. Studies have shown that Lachnoclostridium can be used as a microbial marker for early screening of CRC (J. Q. Liang et al., 2020). Moreover, although the changes in Collinsella in IBD (Joossens et al., 2011), COVID-19 (Hirayama et al., 2021), rheumatoid arthritis (Ruiz-Limón et al., 2022) and other related diseases have been confirmed, this bacterium has not been studied in the context of CRC.

Both the AT and CT were screened for interactions between Gammaretrovirus and different bacteria and fungi (as described previously). Studies have shown that, compared with those in healthy individuals, the abundances of certain bacteria that interact with Gammaretrovirus, such as Enterobacter (Chen et al., 2022), Escherichia (Z. Gao, Guo, Gao, Zhu, and Qin, 2015), and Klebsiella (Wang et al., 2012), are significantly greater in CRC patients, whereas the abundances of Enterococcus (Wang et al., 2012), Clostridioides (Drewes et al., 2022), Neisseria (Zhang et al., 2022), and other genera are significantly lower in CRC patients. Professor Yu Jun's team isolated and identified a new gene marker "m3″ from Lachnoclostridium sp., *Fusobacterium nucleatum* and *Clostridium hathewayi* and reported that *Fusobacterium nucleatum* abundance gradually increased during the development of CRC via the adenoma-carcinoma sequence pathway, indicating the potential of these bacteria in distinguishing CRC patients from healthy individuals (J. Q. Liang et al., 2020). *Enterococcus faecalis* can lead to DNA breakage, point mutations, and protein−DNA cross-linking, leading to chromosome instability and CRC risk (de Almeida, Taddei, and Amedei, 2018). The Pks *Escherichia coli* genome has a special operon (pks), which induces DNA interstrand cross-linking and double-strand breakage, promoting the occurrence of CRC (Pleguezuelos-Manzano et al., 2020). Other studies have shown that the abundance of Pseudogymnoascus, a fungus closely associated with Gammaretrovirus, is significantly greater in CRC patients than in control individuals (Liu et al., 2022), and Pseudogymnoascus sp. VKM F-4518 has been found to have a more than twofold change in the abundance of fungi enriched in CRC (Coker et al., 2019). However, Hanseniaspora and Zygosaccharomyces have not been found to be linked to disease or other intestinal microorganisms. The above analysis revealed that the interactions between different species in feces and tissues are not necessarily consistent. Changes in related bacteria or fungi in the disease state may cause changes in viruses and mediate their effects on the occurrence of CRC, but the specific effects of interspecies interactions need further study.

In the KEGG analysis of the DEGs from the tissue and fecal samples from the three groups, MetaStat analysis showed that the protein digestion and absorption pathways were gradually suppressed as CRC progressed, whereas the biosynthesis of amino acids was activated during the progression of CRC. Compared with that in the NT, the HTLV-1 infection pathway was activated in the AT. This pathway represents a series of immune responses caused by HTLV-1 infection in the body, such as HTLV-1-infected cells upregulating HLA class II molecules and presenting antigens to induce anergy of antigen-specific T cells, and subsequently inducing the occurrence of some tumors (Tan et al., 2021). The above results suggest that changes in pathway activity in feces may affect the development of CRC, but the specific mechanism needs to be further verified. The evidence of functional changes in tissues was not sufficient, but it cannot be completely denied; the reason for the insignificant difference may be the small sample size and the heterogeneity of the population in this study.

Changes in intestinal microorganisms are closely related to the development and progression of CRC. Fecal-derived communities can provide insights into pathological changes in the lumen, but it is inconclusive whether changes in the fecal microbial community reflect changes in the tissue microbiota. In this study, fecal and tissue samples were collected from the same individual, and the community composition of enteric viruses in the feces and tissues of CRC adenoma patients were explored via metagenomic technology, along with the possible metabolic pathways these enteric viruses are associated with. The causal relationship between the observed differences in this study and CRC remains undetermined. These differences might be attributed to the alterations in the microecology during tumor progression. Currently, there is insufficient research evidence to confirm that these differences in viruses can lead to the progression of CRC. However, these differences could potentially be one of the identifying changes of CRC.The limitations of this study are the small sample size, and the samples were not

grouped according to disease status and clinical indicators. However, this study objectively provides a reference for further understanding the role of enteric viruses in CRC.

5. Conclusion

During the development of the CRC via the adenoma-carcinoma sequence pathway, VLPs, the composition of enteric viruses, the interactions between virus-bacteria-fungi and the associated metabolic pathways in both fecal and tissue samples change, which provides a reference for exploring the role of enteric viruses in the occurrence of this disease. However, the mechanisms of action still need further study.

Funding sources

This study is funded by the Health Commission of Sichuan Province Medical Science and Technology Program (Grant No 24QNMP087), the Central Government-Directed Project for Local Science and Technology Development (Grant No 2024ZYD0146) and the Clinical Science Research Fund Project at Chengdu Medical College (24LHXHZL-02).

CRediT authorship contribution statement

Ying Yang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Dan Wang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Longlin Li:** Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jieyu Song:** Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xianglan Yang:** Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jun Li:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

We declare that we have no known competing financial interests or personal relationships that could affect the work reported in this article.

Acknowledgements

We thank all the people who helped us with this manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.virusres.2025.199569.

Data availability

Data will be made available on request.

References

Agut, H., Bonnafous, P., Gautheret-Dejean, A., 2016. Human herpesviruses 6A, 6B, and 7. Microbiol. Spectr. 4 (3). https://doi.org/10.1128/microbiolspec.DMIH2-0007-2015.

Benler, S., Yutin, N., Antipov, D., Rayko, M., Shmakov, S., Gussow, A.B., Koonin, E.V., 2021. Thousands of previously unknown phages discovered in whole-community human gut metagenomes. Microbiome 9 (1), 78. https://doi.org/10.1186/s40168-021-01017-w.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P., Rohwer, F., 2003. Metagenomic analyses of an uncultured viral community from human feces. J. Bacteriol. 185 (20), 6220–6223. https://doi.org/10.1128/JB.185.20.6220-6223.2003.

C, E., Aa, J., R, A., Mh, H., Mx, B., An, H., Ch, L., 2022. A comprehensive framework for early-onset colorectal cancer research. Lancet Oncol. 23 (3). https://doi.org/10.1016/S1470-2045(21)00588-X.

Carethers, J.M., Jung, B.H., 2015. Genetics and genetic biomarkers in sporadic colorectal cancer. Gastroenterology 149 (5), 1177–1190. https://doi.org/10.1053/j.gastro.2015.06.047 e3.

Chen, F., Dai, X., Zhou, C.-C., Li, K.-X., Zhang, Y.-J., Lou, X.-Y., Cui, W., 2022. Integrated analysis of the faecal metagenome and serum metabolome reveals the role of gut microbiome-associated metabolites in the detection of colorectal cancer and adenoma. Gut 71 (7), 1315–1325. https://doi.org/10.1136/gutjnl-2020-323476.

Coker, O.O., Nakatsu, G., Dai, R.Z., Wu, W.K.K., Wong, S.H., Ng, S.C., Yu, J., 2019. Enteric fungal microbiota dysbiosis and ecological alterations in colorectal cancer. Gut 68 (4), 654–662. https://doi.org/10.1136/gutjnl-2018-317178.

de Almeida, C.V., Taddei, A., Amedei, A., 2018. The controversial role of Enterococcus faecalis in colorectal cancer. Therap. Adv. Gastroenterol. 11, 1756284818783606. https://doi.org/10.1177/1756284818783606.

De Bolle, L., Naesens, L., De Clercq, E., 2005. Update on human herpesvirus 6 biology, clinical features, and therapy. Clin. Microbiol. Rev. 18 (1), 217–245. https://doi.org/10.1128/CMR.18.1.217-245.2005.

Dekker, E., Tanis, P.J., Vleugels, J.L.A., Kasi, P.M., Wallace, M.B., 2019. Colorectal cancer. Lancet 394 (10207), 1467–1480. https://doi.org/10.1016/S0140-6736(19)32319-0.

Drewes, J.L., Chen, J., Markham, N.O., Knippel, R.J., Domingue, J.C., Tam, A.J., Sears, C. L., 2022. Human colon cancer-derived clostridioides difficile strains drive colonic tumorigenesis in mice. Cancer Discov. 12 (8), 1873–1885. https://doi.org/10.1158/2159-8290.CD-21-1273.

Dulery, R., Salleron, J., Dewilde, A., Rossignol, J., Boyle, E.M., Gay, J., Yakoub-Agha, I., 2012. Early human herpesvirus type 6 reactivation after allogeneic stem cell transplantation: a large-scale clinical study. Biol. Blood Marrow Transplant. J. Am. Soc. Blood Marrow Transplant. 18 (7), 1080–1089. https://doi.org/10.1016/j.bbmt.2011.12.579.

F, B., J, F., I, S., Rl, S., La, T., A, J., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 68 (6). https://doi.org/10.3322/caac.21492.

Fernandes, M.A., Verstraete, S.G., Phan, T.G., Deng, X., Stekol, E., LaMere, B., Delwart, E., 2019. Enteric virome and bacterial microbiota in children with ulcerative colitis and Crohn Disease. J. Pediatr. Gastroenterol. Nutr. 68 (1), 30–36. https://doi.org/10.1097/MPG.0000000000002140.

Gao, M., Yi, L., Wang, Y., Gao, J., Liu, H., Zhang, X., Bai, C., 2022. Characterization and Genomic analysis of bacteriophage vB_KpnM_IME346 targeting clinical Klebsiella pneumoniae strain of the K63 capsular type. Curr. Microbiol. 79 (6), 160. https://doi.org/10.1007/s00284-022-02834-4.

Gao, Z., Guo, B., Gao, R., Zhu, Q., Qin, H., 2015. Microbiota disbiosis is associated with colorectal cancer. Front. Microbiol. 6, 20. https://doi.org/10.3389/fmicb.2015.00020.

Gong, B., Wang, C., Meng, F., Wang, H., Song, B., Yang, Y., Shan, Z., 2021. Association between gut microbiota and autoimmune thyroid disease: a systematic review and meta-analysis. Front. Endocrinol. (Lausanne) 12, 774362. https://doi.org/10.3389/fendo.2021.774362.

Hannigan, G.D., Duhaime, M.B., Ruffin, M.T., Koumpouras, C.C., Schloss, P.D., 2018. Diagnostic potential and interactive dynamics of the colorectal cancer virome. mBio 9 (6). https://doi.org/10.1128/mBio.02248-18 e02248-18.

Hartmann, K., Hofmann-Lehmann, R., 2020. What's new in feline leukemia virus infection. *the veterinary clinics of North America*. Small Animal Pract. 50 (5), 1013–1036. https://doi.org/10.1016/j.cvsm.2020.05.006.

Hirayama, M., Nishiwaki, H., Hamaguchi, T., Ito, M., Ueyama, J., Maeda, T., Ohno, K., 2021. Intestinal Collinsella may mitigate infection and exacerbation of COVID-19 by producing ursodeoxycholate. PloS One 16 (11), e0260451. https://doi.org/10.1371/journal.pone.0260451.

Jiang, L., Lang, S., Duan, Y., Zhang, X., Gao, B., Chopyk, J., Schnabl, B., 2020. Intestinal virome in patients with alcoholic Hepatitis. Hepatology 72 (6), 2182–2196. https://doi.org/10.1002/hep.31459.

Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., Vermeire, S., 2011. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. Gut 60 (5), 631–637. https://doi.org/10.1136/gut.2010.223263.

Karlsson, F., Tremaroli, V., Nielsen, J., Bäckhed, F., 2013. Assessing the human gut microbiota in metabolic diseases. Diabetes 62 (10), 3341–3349. https://doi.org/10.2337/db13-0844.

Krug, L.T., Pellett, P.E., 2014. Roseolovirus molecular biology: recent advances. Curr. Opin. Virol. 9, 170–177. https://doi.org/10.1016/j.coviro.2014.10.004.

Lefkowitz, E.J., Dempsey, D.M., Hendrickson, R.C., Orton, R.J., Siddell, S.G., Smith, D.B., 2018. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Res. 46 (D1), D708–D717. https://doi.org/10.1093/nar/gkx932.

Liang, G., Bushman, F.D., 2021. The human virome: assembly, composition and host interactions. Nat. Rev. Microbiol. 19 (8), 514–527. https://doi.org/10.1038/s41579-021-00536-5.

Liang, J.Q., Li, T., Nakatsu, G., Chen, Y.-X., Yau, T.O., Chu, E., Yu, J., 2020. A novel faecal lachnoclostridium marker for the non-invasive diagnosis of colorectal adenoma and cancer. Gut 69 (7), 1248–1257. https://doi.org/10.1136/gutjnl-2019-318532.

Liu, N.-N., Jiao, N., Tan, J.-C., Wang, Z., Wu, D., Wang, A.-J., Wang, H., 2022. Multi-kingdom microbiota analyses identify bacterial-fungal interactions and biomarkers of colorectal cancer across cohorts. Nat. Microbiol. 7 (2), 238–250. https://doi.org/10.1038/s41564-021-01030-7.

M, K.M., Maa, K., P, G., Ze, T., M, T., J, R., Cf, M., 2020. Bacteriophages isolated from stunted children can regulate gut bacterial communities in an age-specific manner. Cell Host Microbe 27 (2). https://doi.org/10.1016/j.chom.2020.01.004.

Moskal, A., Freisling, H., Byrnes, G., Assi, N., Fahey, M.T., Jenab, M., Slimani, N., 2016. Main nutrient patterns and colorectal cancer risk in the European Prospective Investigation into Cancer and Nutrition study. Br. J. Cancer 115 (11), 1430–1440. https://doi.org/10.1038/bjc.2016.334.

Nakatsu, G., Zhou, H., Wu, W.K.K., Wong, S.H., Coker, O.O., Dai, Z., Yu, J., 2018. Alterations in enteric virome are associated with colorectal cancer and survival outcomes. Gastroenterology 155 (2), 529–541. https://doi.org/10.1053/j.gastro.2018.04.018 e5.

Nishino, K., Nishida, A., Inoue, R., Kawada, Y., Ohno, M., Sakai, S., Andoh, A., 2018. Analysis of endoscopic brush samples identified mucosa-associated dysbiosis in inflammatory bowel disease. J. Gastroenterol. 53 (1), 95–106. https://doi.org/10.1007/s00535-017-1384-4.

Pleguezuelos-Manzano, C., Puschhof, J., Rosendahl Huber, A., van Hoeck, A., Wood, H. M., Nomburg, J., Clevers, H., 2020. Mutational signature in colorectal cancer caused by genotoxic pks+ *E. coli*. Nature 580 (7802), 269–273. https://doi.org/10.1038/s41586-020-2080-8.

Ruiz-Limón, P., Mena-Vázquez, N., Moreno-Indias, I., Manrique-Arija, S., Lisbona-Montañez, J.M., Cano-García, L., Fernández-Nebro, A., 2022. Collinsella is associated with cumulative inflammatory burden in an established rheumatoid arthritis cohort. Biomed. Pharmacother. 153, 113518. https://doi.org/10.1016/j.biopha.2022.113518.

Shapira, M., 2016. Gut microbiotas and host evolution: scaling up symbiosis. Trends Ecol. Evol. (Amst.) 31 (7), 539–549. https://doi.org/10.1016/j.tree.2016.03.006.

Shkoporov, A.N., Hill, C., 2019. Bacteriophages of the Human gut: the 《known unknown》 of the microbiome. Cell Host Microbe 25 (2), 195–209. https://doi.org/10.1016/j.chom.2019.01.017.

Sudarikov, K., Tyakht, A., Alexeev, D., 2017. Methods for the metagenomic data visualization and analysis. Curr. Issues Mol. Biol. 24, 37–58. https://doi.org/10.21775/cimb.024.037.

Tan, B.J., Sugata, K., Reda, O., Matsuo, M., Uchiyama, K., Miyazato, P., Satou, Y., 2021. HTLV-1 infection promotes excessive T cell activation and transformation into adult T cell leukemia/lymphoma. J. Clin. Invest. 131 (24), e150472. https://doi.org/10.1172/JCI150472.

Tomofuji, Y., Kishikawa, T., Maeda, Y., Ogawa, K., Nii, T., Okuno, T., Okada, Y., 2022. Whole gut virome analysis of 476 Japanese revealed a link between phage and autoimmune disease. Ann. Rheum. Dis. 81 (2), 278–288. https://doi.org/10.1136/annrheumdis-2021-221267.

Tumas-Brundage, K.M., Garret, W., Blank, K., Prystowsky, M.B., 1996. Murine leukemia virus infects early bone marrow progenitors in immunocompetent mice. Virology 224 (2), 573–575. https://doi.org/10.1006/viro.1996.0567.

V Volozhantsev, N., M Shpirt, A., I Borzilov, A., V Komisarova, E., M Krasilnikova, V., S Shashkov, A., A Knirel, Y., 2020. Characterization and therapeutic potential of bacteriophage-encoded polysaccharide depolymerases with β galactosidase activity against Klebsiella pneumoniae K57 capsular type. Antibiotics (Basel) 9 (11), 732. https://doi.org/10.3390/antibiotics9110732.

Wang, T., Cai, G., Qiu, Y., Fei, N., Zhang, M., Pang, X., Zhao, L., 2012. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. ISME J. 6 (2), 320–329. https://doi.org/10.1038/ismej.2011.109.

Z, C., N, S., E, B., Mp, E., T, Z., P, L., 2022. The gut virome: a new microbiome component in health and disease. EBioMedicine 81. https://doi.org/10.1016/j.ebiom.2022.104113.

Zhang, C., Hu, A., Li, J., Zhang, F., Zhong, P., Li, Y., Li, Y., 2022. Combined non-invasive prediction and new biomarkers of oral and fecal microbiota in patients with gastric and colorectal cancer. Front. Cell Infect. Microbiol. 12, 830684. https://doi.org/10.3389/fcimb.2022.830684.

Zuo, T., Zhan, H., Zhang, F., Liu, Q., Tso, E.Y.K., Lui, G.C.Y., Ng, S.C., 2020. Alterations in fecal fungal microbiome of patients with COVID-19 during time of hospitalization until discharge. Gastroenterology 159 (4), 1302–1310. https://doi.org/10.1053/j.gastro.2020.06.048 e5.