


RESEARCH

Open Access



Characteristics of the gut virome in patients with premalignant colorectal adenoma

Pan Zhang^{1,2,3,4*†}, Xiaofeng Tuo^{1,2,3,4†}, Jiong Jiang^{1,2,3,4†}, Yue Zhang^{5†}, Juhui Zhao^{1,2,3,4†}, Chengzhao Deng^{1,2,3,4†}, Gang Zhao^{1,2,3,4}, Yan Cheng^{1,2,3,4}, Lingqin Song⁶, Yan Yang⁷, Ruochun Guo⁵, Huan Zhang^{1,2,3,4}, Hongli Zhao^{1,2,3,4}, Shiyang Ma^{1,2,3,4*}, Lu Li^{1,2,3,4*} and Haitao Shi^{1,2,3,4*} 

Abstract

Background The multi-kingdom gut microbiota (e.g., bacteriome, mycobiome, and virome) characteristics of colorectal cancer have been extensively studied, yet there is still an insufficient description of the microbiota features in its early-stage, colorectal adenoma, particularly in the gut virome aspect.

Methods Based on the Metagenomic Gut Virus catalogue (MGV) containing 54,118 non-redundant gut viral genomes, this study characterized the virome composition and diversity using publicly available metagenomic sequencing data from 419 individuals with premalignant colorectal adenoma and 552 healthy controls. Furthermore, we identified and assessed the reliability and classification performance of adenoma-associated microbial signatures through comparative analysis and the random forest model.

Results Our results revealed a notable shift in the gut virome structure of patients compared to healthy controls, characterized by a significant increase in viral families such as *Microviridae*, *Podoviridae*, *crAss-like*, and *Quimbyviridae*. At the viral operational taxonomic unit (vOTU) level, we identified 479 vOTU signatures showing significant differences in relative abundances between patients and controls, including some patient-enriched vOTUs tending to infect *Bacteroidaceae* and *Lachnospiraceae*. Correlation network analysis revealed specific bacterial species correlated with adenoma-associated viruses, suggesting frequent interactions between them. Moreover, random forest models trained on gut viral and bacterial signatures demonstrated area under the curve (AUC) scores of 0.68, 0.82, and 0.76 for classifying healthy individuals versus patients with tubular adenomas, patients with sessile serrated adenomas, and patients with both conditions, respectively. In three independent validation cohorts, the classification performance achieved AUC scores ranging from 0.61 to 0.65.

[†]Pan Zhang, Xiaofeng Tuo, Jiong Jiang, Yue Zhang, Juhui Zhao and Chengzhao Deng have contributed equally.

*Correspondence:

Pan Zhang
zhangwangpan71@163.com
Shiyang Ma
shiyangma@163.com
Lu Li
lilu0505@aliyun.com
Haitao Shi
shihaitao7@163.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Conclusions Our study provides insights into the gut virome in premalignant colorectal adenoma, highlighting its potential role in disease development and diagnosis. Further investigations are warranted to elucidate the underlying mechanisms of gut virus-bacteria interactions and validate diagnostic models in larger populations.

Keywords Gut virome, Colorectal adenoma, Metagenomics, Viral signatures, Microbiota

Introduction

Colorectal cancer (CRC) is the third most common cancer worldwide with significant global health concerns, representing a major cause of morbidity and mortality, particularly in low- and middle-income countries [1, 2, 3]. In 2018 alone, a staggering 1.85 million individuals globally were diagnosed with CRC [4]. Projected on the trajectories of population aging and growth, it is anticipated that the incidence of new colorectal cancer cases will surge to 3.2 million by the year 2040 [5]. The genesis and progression of CRC commonly involve the evolution of colorectal polyps and adenomas, recognized as precursor lesions to colorectal cancer. Two primary premalignant lesions, tubular adenomas (TA) and sessile serrated adenomas (SSA), play integral roles in this progression [6, 7]. A comprehensive understanding of the background of colorectal adenomas is vital, given their pivotal role in the incremental journey toward malignancy. Factors contributing to the development of colorectal adenomas include a combination of genetic predisposition and environmental exposures such as smoking, alcohol consumption, obesity, and unhealthy dietary habits [8, 9, 10].

Research interest in the gut microbiome concerning CRC and adenoma has grown significantly. Numerous studies have extensively explored the association between the gut bacteriome and colorectal cancer and adenoma [11, 12, 13, 14]. Typical findings include a decrease in microbial diversity in cancer and adenoma patients, along with the enrichment of certain opportunistic pathogens, such as *Fusobacterium nucleatum* [15, 16]. This pathogen has been demonstrated to be causative in animal models of CRC carcinogenesis and progression [17, 18]. Importantly, recent meta-analysis studies have identified a high consistency of CRC bacterial signatures across populations and established a functional link with choline degradation [16, 19]. However, despite the extensive focus on the bacterial component, exploration of the viral fraction, specifically the virome, in CRC and adenomas is relatively limited [20]. A representative study by Cheng et al., through meta-analysis, found differences in the gut virome of cancer and adenoma patients compared to healthy controls, indicating that viral signatures contribute to distinguishing cancer cases [21]. Nonetheless, studies investigating the gut virome in these patients face challenges, including sample size limitations, diverse analysis methods, and the subdivision of disease subtypes

such as SSA and TA. Moreover, beyond its relevance in CRC and adenomas, the gut virome has been explored in various systemic diseases, including inflammatory bowel disease (IBD) [22], autoimmune disorders [23, 24, 25], and liver diseases [26]. Typically, alterations in the gut virome of individuals with these diseases manifest as a reduction in diversity and changes in specific viral taxonomies and functionalities. Considering that premalignant colorectal adenomas are often precursors to cancer, and predictive models based on gut bacterial composition often exhibit cross-sectional efficacy, we hypothesize that the gut virome might offer a potential avenue for predicting disease progression.

To these ends, herein, we conducted a reanalysis of a large-scale dataset comprising fecal metagenomes from 419 patients, including 62 patients with SSA, 321 patients with TA, and 36 patients with both conditions, along with 552 healthy controls. The dataset was sourced from a recently published study [14], providing a comprehensive platform for investigating the potential associations between the gut virome and colorectal adenomas.

Materials and methods

Dataset and processing of sequencing reads

The fecal metagenomic dataset used in this study, comprising samples from 419 adenoma patients and 522 controls, was obtained from the NCBI Sequence Read Archive (SRA) with accession code PRJNA784939 [14]. Initially, we filtered out low-quality sequencing reads in each metagenome using fastp with the following parameters “-l 90 -q 20 -u 30 -y --trim_poly_g” [27]. The human reads were then removed by aligning against the GRCh38 human reference genome using Bowtie 2 with default parameters [28]. The reads that did not map to the human genome were considered clean reads and used for subsequent analyses.

Analyses of the gut virome

After filtering out human-derived metagenomic reads, the remaining data from all samples were aligned to the Metagenomic Gut Virus catalogue (MGV) [29], which is constructed from over 11,000 publicly available fecal metagenomes, to generate gut viral compositions. The MGV, comprising nearly 190,000 viral genomes, was grouped into 54,118 viral operational taxonomic units (vOTUs) based on an average nucleotide similarity

threshold of 95%. Total clean reads of all samples were randomly subsampled to the same sequencing amount (20 million), and then were mapped to 54,118 vOTUs using Bowtie 2 with default parameters. The relative abundance of each vOTU was calculated as the number of reads mapped to this vOTU divided by the total mapping reads. Additionally, the relative abundance of each viral family was generated by summing up the relative abundances of vOTUs annotated with this family.

Protein-coding genes of the vOTUs were predicted using Prodigal with the option “-p meta” [30]. Taxonomic annotation of viral sequences was performed through protein sequence alignment against a comprehensive database, which combined resources from the Virus-Host DB [31] and viral proteins from *crAss-like* [32], *Flandersviridae*, *Quimbyviridae*, and *Gratiaviridae* [33]. A viral sequence was assigned a family-level classification when over a quarter of its proteins displayed match to the same viral family. Virus-host predictions were carried out using two bioinformatic methodologies: CRISPR-spacer matches and prophage blasts, following the previous methodologies [34, 35]. Functional annotation of viral proteins was performed using the DIAMOND tool with the parameters “--query-cover 50 --subject-cover 50 --id 30”, utilizing the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [36]. Each protein received a KO assignment based on the best-hit gene in the database.

Analyses of the gut bacteriome

We conducted bacterial taxonomic profiling of the fecal metagenomic dataset for patients with colorectal adenoma and healthy controls using MetaPhlAn 4 with default parameters, relying on the mpa_vJan21_CHOC-OPHlAnSGB_202103 database [37], which relies on clade-specific marker genes to unambiguously classify metagenomic reads to taxonomies and yield relative abundances of taxa identified in the sample. The analysis included taxonomic classification at the phylum, class, order, family, genus, and species levels. A uniform number of reads (20 million) were randomly selected from each fecal sample to calculate the relative abundance of each bacterial taxon.

Statistical analyses

Statistical analyses and visualization were conducted under the R platform and any functional modules not specifically mentioned were set to use the default parameters. Gut virome diversity indexes were estimated based on the profiles at the vOTU level. The observed number of vOTUs was calculated as the count of vOTUs with relative abundance greater than 0. The Shannon and Simpson diversity indexes were calculated using the *vegan*:

diversity function with the options `index="shannon"` and `index="simpson"`, respectively [38]. The Bray–Curtis distances between samples were calculated using the *vegdist* function. Distance-based redundancy analysis (dbRDA) of Bray–Curtis distances was implemented via the *capscale* function. Permutational multivariate analysis of variance (PERMANOVA) was carried out using the *adonis* function.

The identification of viral markers at the vOTU level was conducted between any two groups using the *wilcox.test* function from the R *stats* package. The resulting *P* values were adjusted using the *p.adjust* function with the option “method=BH.” A vOTU with a *P* value of <0.05 was recognized as a disease-associated viral signature. Additionally, the same statistical testing methods and filtering criteria were applied to identify disease-associated bacterial signatures. The correlation coefficient of gut viral and bacterial markers was measured by the *cor.test* function with the option “method=Spearman”. Data visualization was performed using Cytoscape software [39]. The random forest models based on adenomas-associated viral or bacterial markers were built using the *randomForest* package followed by five times of fivefold cross-validations, and their performances were evaluated based on the area under the receiver operating characteristic curve (AUC) that was calculated by the *roc* function. The importance ordering of markers was obtained via the *importance* function.

Results

Overview of the gut viral community

To profile the gut viral community in individuals with colorectal adenoma, we conducted a re-analysis of metagenomic sequencing data obtained from fecal samples of 419 patients, including 321 TA patients, 62 SSA patients, and 36 patients diagnosed with both TA and SSA, in addition to 552 healthy controls (Table S1) [14]. Gut viral compositions of all samples were generated based on the MGv catalogue [29] (see Materials and Methods). In total, 51,121 vOTUs were identified from metagenomic reads of colorectal adenoma patients and healthy controls, accounting for 10.6% ($\pm 2.2\%$) of their total reads. Of these vOTUs, 44.6% were taxonomically assigned into 22 viral families. Notably, *Siphoviridae* and *Myoviridae* emerged as dominant families, constituting 28.9 and 6.2% of all vOTUs, respectively (Fig. 1A). The remaining vOTUs belonged to families such as *Podoviridae*, *Quimbyviridae*, *Microviridae*, *Podoviridae_crAss-like*, and others. In terms of composition, we observed that viruses from the *Myoviridae*, *Podoviridae_crAss-like* families exhibited higher relative abundance across all samples, while *Microviridae* viruses demonstrated an average relative abundance lower than that of other

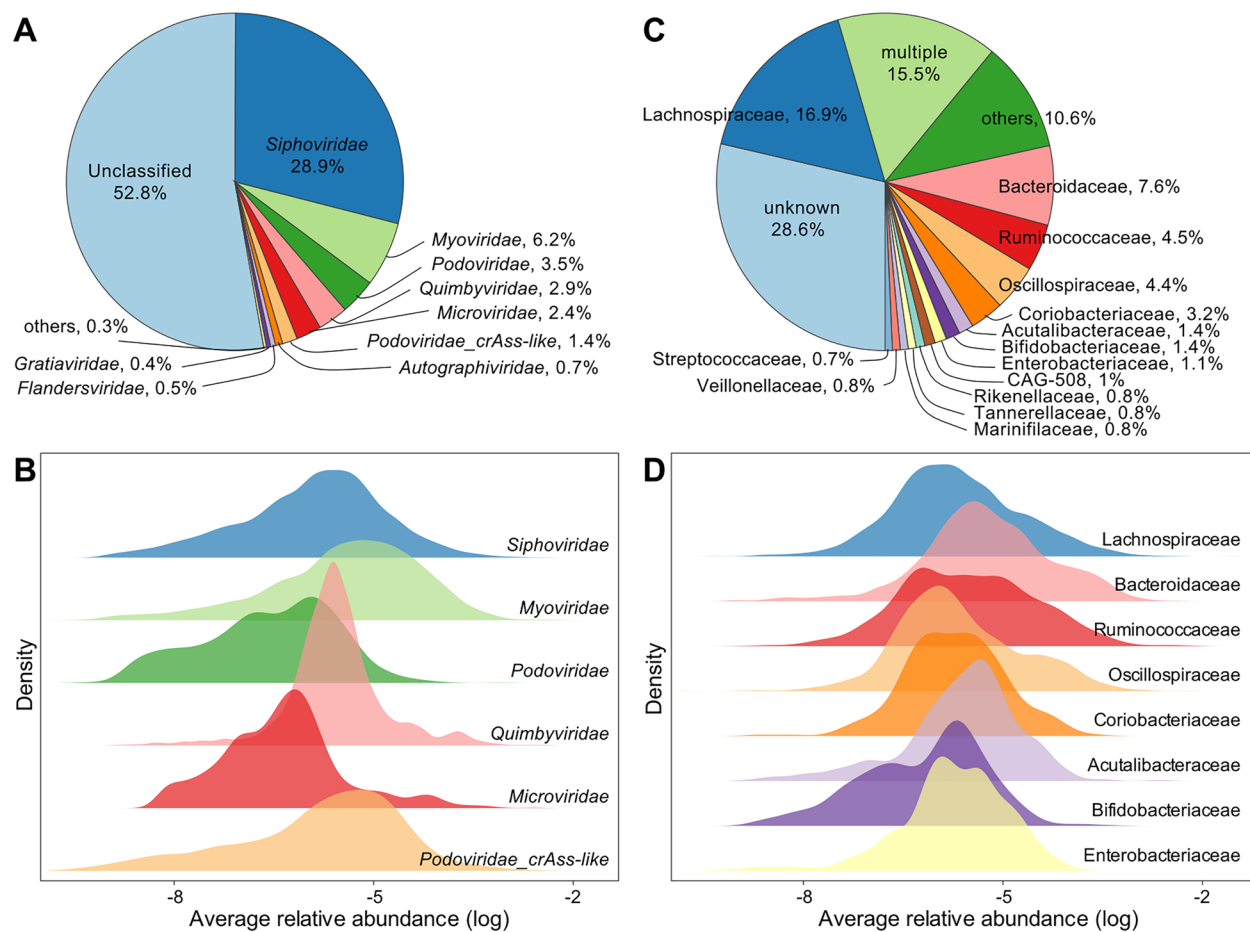


Fig. 1 Overview of the gut viral catalogue used in this study. **A** Pie plot showing the distribution of viral families of 51,121 viral operational taxonomic units (vOTUs). **B** Distribution of average relative abundances of vOTUs belonging to different families. **C** Pie plot showing the distribution of prokaryotic hosts (grouped at the bacterial family level) of 51,121 vOTUs. **D** Distribution of average relative abundances of vOTUs predicted to infect different bacterial families

families (Fig. 1B). Additionally, we predicted the prokaryotic hosts of 56.4% of vOTUs. The majority of predicted hosts were associated with *Firmicutes* (predominantly *Lachnospiraceae*, *Ruminococcaceae*, *Oscillospiraceae*, and *Acetivibacteraceae* at the family level), *Bacteroidaceae* (primarily *Bacteroidaceae*), *Actinobacteria* (mainly *Coriobacteriaceae* and *Bifidobacteriaceae*), and *Proteobacteria* (mainly *Enterobacteriaceae*) (Fig. 1C). In terms of abundance, viruses predicted to infect *Bacteroidaceae* exhibited a higher average relative abundance, whereas those infecting hosts in the *Bifidobacteriaceae* family generally displayed a lower relative abundance (Fig. 1D).

Diversity and structure of the gut virome in relation to colorectal adenoma

Rarefaction curve analysis revealed that the coverage of vOTUs in all samples tends to saturate at the current number of reads, indicating that the gut virome is well

represented (Fig. 2A). When comparing different groups, the vOTU richness of the gut virome showed no significant differences among the three patient groups and the control group (Fig. 2B). Similarly, at the vOTU level, both the Shannon diversity index and Simpson index were not significantly different between patients and healthy controls (Wilcoxon rank-sum test $p > 0.05$; Fig. 2C). To further explore differences between patients and healthy controls, we conducted dbRDA analysis based on the Bray–Curtis distance of the gut virome. The analysis revealed a noticeable shift in the dbRDA plot for patient groups, including TA patients, SSA patients, and patients with both conditions, compared to healthy controls (Fig. 2D).

At the family level, a substantial proportion of the total viral sequences were attributed to vOTUs belonging to unknown viral families (Fig. 2E). This was consistent with findings from previous studies [35, 40], indicating

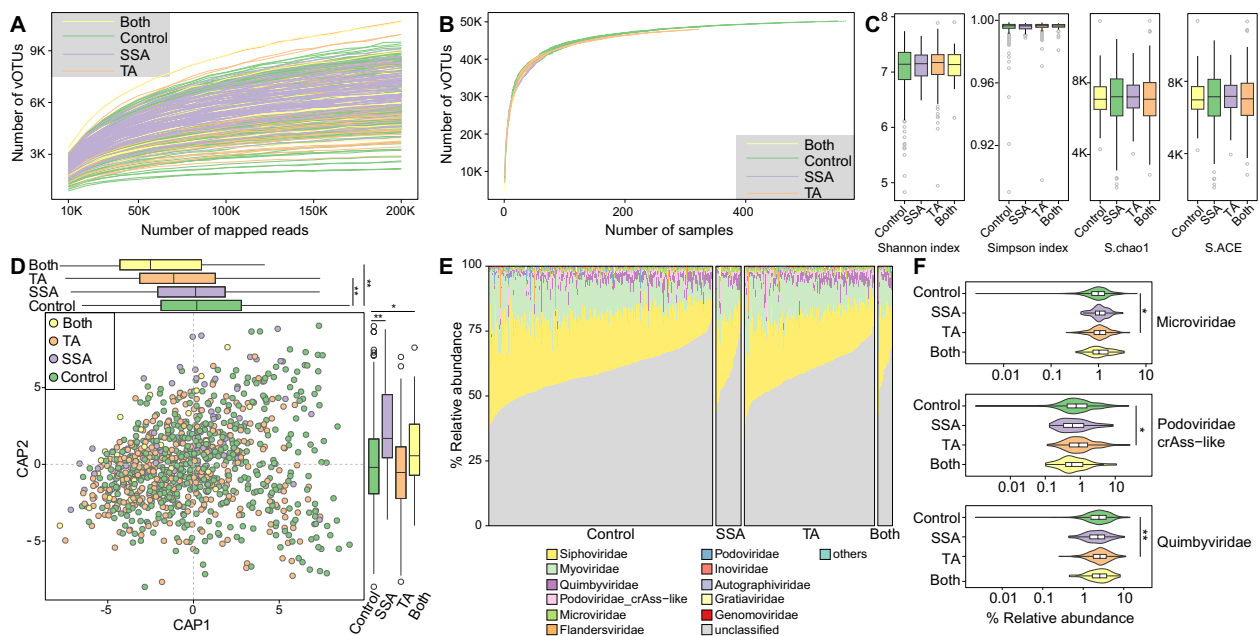


Fig. 2 Diversity and composition of the gut virome in adenoma patients and healthy controls. **A** Rarefaction curve analysis showing the coverage of viral operational taxonomic units (vOTUs) as a function of the number of reads for each sample. **B** Rarefaction analysis showing an increase in the number of vOTUs observed as the number of random samples increased. **C** Boxplot showing the comparison of alpha diversity indexes between patients and controls. Boxes show medians/quartiles; error bars extend to the most extreme values within 1.5 interquartile ranges. **D** Distance-based redundancy analysis (dbRDA) of the Bray–Curtis distance of the gut virome of all samples at the vOTU level. Samples are shown at the first and second principal coordinates (CAP1 and CAP2), and the ratio of variance contributed by these two principal coordinates is shown. Upper and left boxplots show the sample scores in CAP1 and CAP2, respectively. **E** Bar plot showing the gut viral composition of fecal metagenomes from patients and controls at the family level. Only the top 12 viral families with the highest abundance are shown. More details can be found in Table S2. **F** Boxplot showing the relative abundance of differentially abundant viral families between patients and controls. The statistical test results for all viral families can be found in Table S2. Wilcoxon rank-sum test: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

a significant underrepresentation of the gut virome. *Siphoviridae* and *Myoviridae* were identified as the most dominant families in all fecal samples, with no significant differences between patients and controls (Wilcoxon rank-sum test $p > 0.05$). Three families, including *Microviridae*, *Podoviridae_crAss-like*, and *Quimbyviridae*, exhibited significant differences between TA patients and healthy controls (Wilcoxon rank-sum test $p < 0.05$ and $q < 0.10$), with a notable increase in abundance observed in TA patients (Fig. 1F; Table S2).

Identification of viral signatures of colorectal adenoma

To explore the gut viral signatures associated with colorectal adenoma, we conducted a comparative analysis of the viral profiles between each patient group and the healthy controls at the vOTU level. This analysis identified a substantial number of vOTU signatures exhibiting significant differences in relative abundances between patients and controls, including 138 signatures in SSA patients (131 SSA-enriched and 7 SSA-depleted), 103 signatures in TA patients (72 TA-enriched and 31 TA-depleted), and 262 signatures in patients with both conditions (196 both-enriched and 66 both-depleted)

(Wilcoxon rank-sum test $p < 0.05$; Fig. 3A–C; Table S3). Taxonomically, the SSA-enriched vOTUs comprised 36 *Siphoviridae* (corresponding to 27.5% of the 131 SSA-enriched vOTUs), 10 *Myoviridae* (7.6%), 3 *Microviridae* (2.3%), and 1 *Podoviridae_crAss-like* (0.8%) members and 81 family-unclassified viruses (61.8%), whereas the SSA-depleted vOTUs included 3 *Siphoviridae*, 1 *Flandersviridae*, 1 *Podoviridae_crAss-like* member, and 2 unclassified viruses (Fig. 3D). The TA-enriched vOTUs were mostly unclassified viruses ($n = 57$; corresponding to 79.2% of the 72 TA-enriched vOTUs), along with several members of *Myoviridae* ($n = 7$), *Siphoviridae* ($n = 7$), and *Quimbyviridae* ($n = 1$). In contrast, the TA-depleted vOTUs included members of *Siphoviridae* ($n = 11$; corresponding to 35.5% of the 31 TA-depleted vOTUs), *Microviridae* ($n = 3$), *Myoviridae* ($n = 2$), and *Podoviridae_crAss-like* ($n = 2$) and some unclassified viruses ($n = 13$; corresponding to 41.9% of the 31 TA-depleted vOTUs) (Fig. 3E). vOTU signatures that enriched in patients with both SSA and TA included 29 members of *Siphoviridae* (corresponding to 14.8% of the 196 both-enriched vOTUs), 17 *Myoviridae* (9.7%), 9 *Quimbyviridae* (4.6%), 1 *Inoviridae* (0.5%), 1 *Microviridae* (0.5%), 1 *Podoviridae_crAss-like* (0.5%),

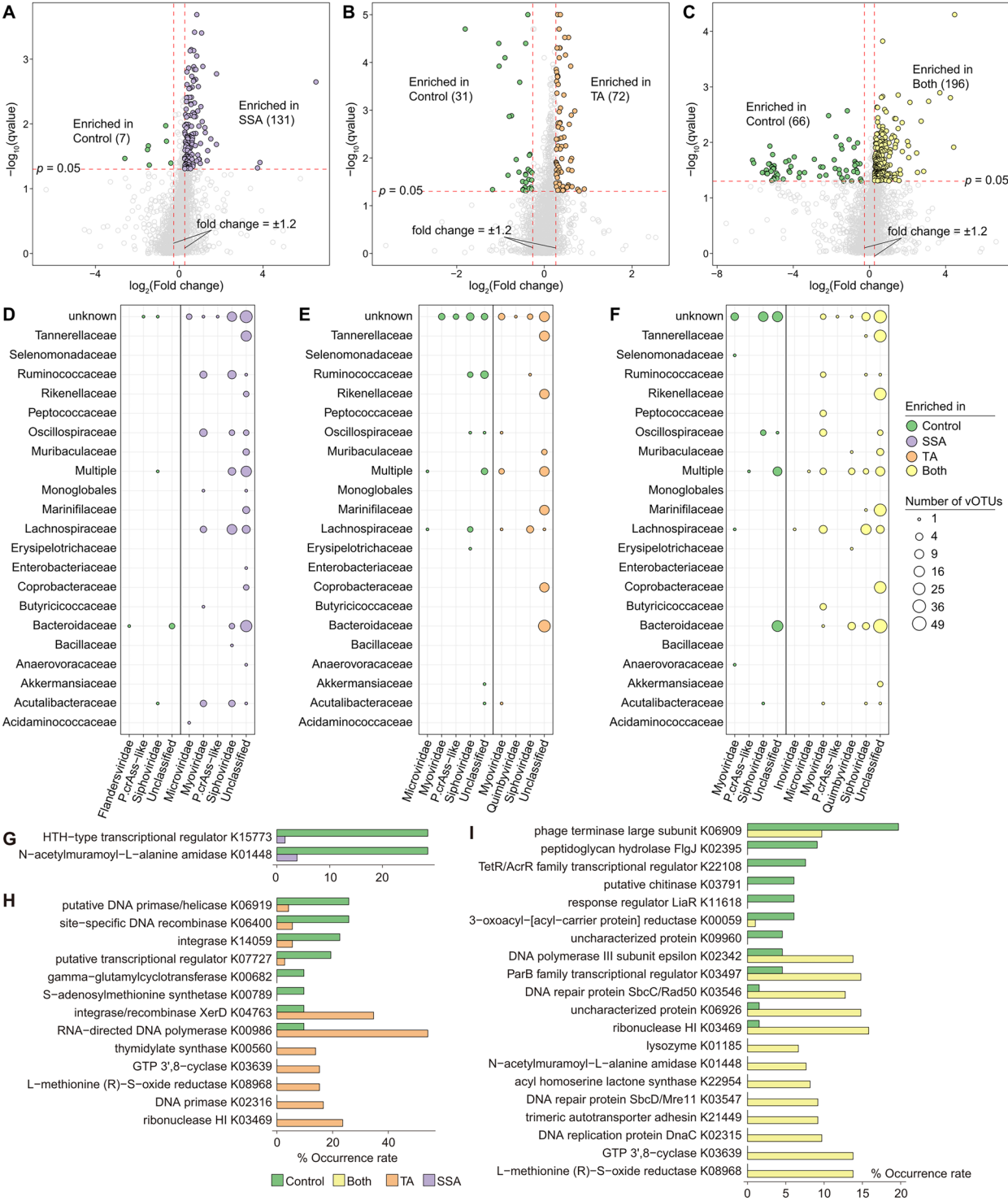


Fig. 3 Adenoma-associated gut viral signatures and functions. **A–C** Volcano map showing the fold change and q -values of all viral operational taxonomic units (vOTUs) for the comparisons between healthy controls and patients with sessile serrated adenomas (SSA) (**A**), tubular adenomas (TA) (**B**), and both (**C**). vOTUs with absolute value of fold change greater than 1.2 and q -value less than 0.05 were regarded as significantly different between patients and controls. **D–F** Taxonomic and host information of SSA-associated (**D**), TA-associated (**E**), and both-associated (**F**) vOTUs. **G–I** Bar plot showing the occurrence rates of differentially abundant KEGG orthologs (KOs) between the adenoma-enriched and adenoma-depleted vOTUs

and 140 family-unclassified viruses (71.4%), whereas the vOTUs depleted in these patients were 14 members of *Siphoviridae* (corresponding to 21.2% of the 66 both-enriched vOTUs), 6 *Myoviridae* (9.1%), and 1 *Podoviridae_crAss-like* and 45 unclassified viruses (68.2%) (Fig. 3F). We further examined the host assignment of the vOTU signatures and found that the viruses enriched in three adenoma patient groups exhibited a remarkable preference for infecting *Bacteroidaceae* (corresponding to 19.1, 40.3, and 45.4% of SSA-, TA-, and both-enriched vOTUs, respectively) and *Lachnospiraceae* (13.7, 8.3, and 13.8% of SSA-, TA-, and both-enriched vOTUs, respectively) (Table S3).

Next, we annotated the functions of the protein-coding genes of adenoma-associated viruses based on the KEGG database. Out of the total 33,738 genes identified from the adenoma-associated vOTUs, 12.9% could be assigned to KEGG orthologs (KOs). To further investigate whether these functions are related to the disease, we compared the occurrence rates of each KO in disease-enriched and disease-depleted vOTUs within three patient groups (i.e., SSA, TA, and both). This analysis revealed differences in the occurrence rates of 2, 13, and 20 KOs in SSA-enriched/depleted, TA-enriched/depleted, and both-enriched/depleted vOTUs, respectively (Fig. 3G–I; Table S4). Notably, K01448, an N-acetylmuramoyl-L-alanine amidase potentially involved in bacterial peptidoglycan hydrolysis [41], was significantly more frequently observed in SSA-depleted vOTUs compared to SSA-enriched vOTUs. Conversely, K08968, an L-methionine (R)-S-oxide reductase involved in methionine biosynthesis, was more common in TA-enriched and both-enriched vOTUs compared with control-enriched vOTUs.

Correlations between adenoma-associated gut viral and bacterial signatures

To investigate whether gut viral signatures can interact with bacterial signatures, we first identified a set of differentially abundant bacteria for three groups of patients, including 30 bacteria associated with SSA, 28 bacteria associated with TA, and 18 bacteria associated with both conditions (Wilcoxon rank-sum test $p < 0.05$; Table S5). Using Spearman correlation analysis, we constructed correlation networks between viral and bacterial signatures for the three groups, revealing 167 correlations between SSA-associated vOTUs and bacteria, 141 correlations between TA-associated vOTUs and bacteria, and 248 correlations between both-associated vOTUs and bacteria (Spearman correlation coefficient $|\rho| > 0.2$ and $q < 0.05$; Fig. 4). We found that in the SSA network, SSA-depleted bacteria such as *Faecalicatena fissicatena* and *Clostridium symbiosum*, as well as SSA-enriched bacteria

like *Ruminococcus torques* and *Clostridiaceae bacterium OM08 6BH*, were correlated with numerous viruses. Similarly, in the TA network, TA-enriched bacteria such as *Phocaeicola massiliensis* were associated with many viruses. Strikingly, in the “both” network, we observed that a both-enriched bacterium, *Bacteroides uniformis*, contained the majority (55.2%) of correlations, followed by several both-enriched bacteria such as *Alistipes shahii* and *Eubacterium rectale*. These findings suggest that these bacteria and viruses may frequently interact with each other, potentially influencing the disease.

Distinguishing colorectal adenoma states using gut viral signatures

Finally, to assess whether the identified gut viral signatures could aid in colorectal cancer prediction, we constructed random forest models for each of the three patient groups based on the relative abundances of their respective adenoma-associated vOTUs. To achieve optimal classification performance, we initially trained the random forest models using all the viral signatures obtained from each group as features and calculated the importance of each signature (Mean Decrease Accuracy). Subsequently, we ranked the signatures by their importance and iteratively added them for model training (for example, the top 1, top 2, top 3, etc.), calculating the AUC for each model to evaluate its classification performance. All model training and predictions were performed using five times of five-fold cross-validation. This analysis revealed that the ability of viral signatures to distinguish between SSA, TA, and both patients and healthy individuals reached AUCs of 0.80, 0.66, and 0.76, respectively (Fig. 5A–D). Using the same approach, we also built models using adenoma-associated bacterial signatures, which showed significantly lower discriminatory ability between patients and healthy individuals compared to vOTUs. Furthermore, we combined the viral and bacterial signatures to construct models, and the performance of the combined signatures was superior to that of either viral or bacterial signatures alone in all three patient groups, with AUCs of 0.82 (95% CI 0.76–0.87), 0.68 (95% CI 0.65–0.73), and 0.76 (95% CI 0.70–0.83) in the SSA, TA, and both groups, respectively. Additionally, we investigated the importance of each signature in the combined models and found that in all three groups, the majority of (76.6%, 71.9%, and 98.4% in the SSA, TA, and both groups, respectively) the most important signatures were viruses (Fig. 5E–G). To assess generalizability, we tested the combined bacterial-viral model in three independent external cohorts for distinguishing CRC patients from healthy controls. The model demonstrated consistent performance, with AUCs ranging from 0.61 to 0.65

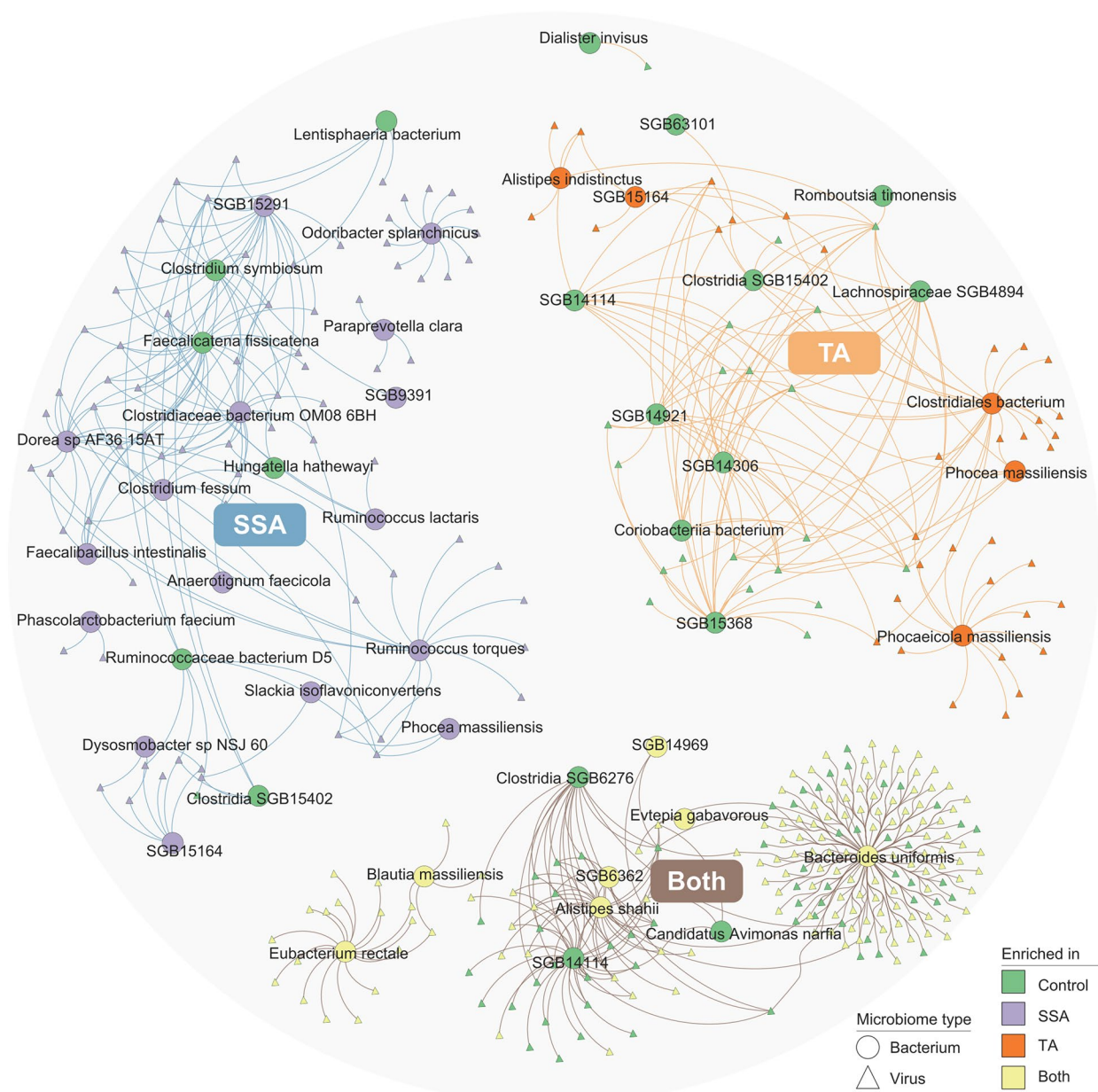


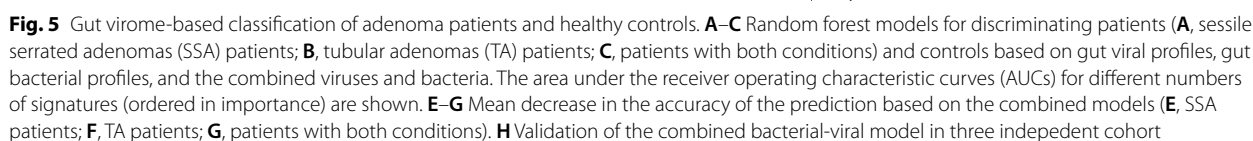
Fig. 4 Correlation networks between adenoma-associated viral operational taxonomic units (vOTUs) and bacteria. Network showing the correlations of the adenoma-associated vOTUs and bacterial species. Spearman's correlation coefficient was used to evaluate the correlation, and correlations with a correlation coefficient $|\rho| > 0.2$ and $q < 0.05$ are shown in the network

(Fig. 5H), reinforcing the potential of multi-kingdom microbial signatures in CRC risk prediction.

Discussion

Almost all colorectal tumors arise from precancerous polyps known as colorectal adenomas [42]. A few studies have explored the potential roles of gut viromes in the progression from adenoma to carcinoma; however, all these current investigations lack systematicity and

comprehensiveness. In this study, the fecal metagenome datasets sourced from 419 patients with colorectal adenomas and 552 healthy controls were investigated to characterize the diversity, taxonomic composition, and functional profile of the gut virome community associated with premalignant colorectal adenomas. Furthermore, the interaction between the gut bacteriome and virome was also investigated, providing valuable insights into the cross-talk between these two distinct kingdoms.



20]. Additionally, studies have also suggested that bacteriophages can directly infiltrate colonic epithelial cells, thereby facilitating the progression and invasiveness of CRC [47, 48]. However, the impact of bacteriophages on CRC carcinogenesis is still in its early stages and requires further investigation.

Several studies have indicated that the CRC cohorts exhibited a greater virome diversity in the gut compared to the control group [12, 49]. Similarly, we found an elevated diversity in the enteric virome among individuals with colorectal adenoma. The observed increase in colon virome diversity, particularly the diversity of

bacteriophage, may be associated with a simultaneous decrease in bacterial diversity among patients with adenoma and CRC. In addition to the modification of virome diversity, discernible alterations in viral profiles at the vOTU level were observed between patients with adenoma and healthy controls. These findings suggest that the utilization of vOTU-based relative abundance profiles can serve as a valuable approach for discerning fecal samples obtained from individuals exhibiting a healthy status, adenomatous polyps, or CRC.

The modulation of gut bacteriophages can shape the dynamics of gut bacterial communities. In this study, we have observed a clear correlation between phages and the composition of bacterial communities in the colon, however, the interaction between phages and bacteria lacks specificity. These findings suggest that the oncogenesis of CRC is not attributed to the direct alterations of influential bacteria by specific phages, but arises from the modulation of overall bacterial community. In our study, we observed a significant reduction in the abundance of *Faecalicatena fissicatena* within the adenoma cohort. *Faecalicatena fissicatena* has been reported to play a crucial role in the production of short-chain fatty acids and has shown an inverse correlation with CRC [50]. However, it was noted that certain other probiotics demonstrated an increase in abundance among adenoma patients. These also corroborate the above notion that phages exert an impact on cancer by modifying the overall bacterial population rather than directly manipulating the influential bacteria.

However, the following limitations should be considered in this study. All recruited adenoma patients were classified based on their clinical pathology, regardless of their clinical symptoms. This may allow us to disregard the potential association between the virus and specific clinical manifestations. Besides, the utilization of the sequencing approach in this study imposes certain limitations on obtaining more comprehensive information. To maximize access to virus data, it is recommended to employ the viral-like particle (VLP) enriched metagenome sequencing method.

In conclusion, we found that temperate phages constitute the predominant constituents of the gut virome. Alterations in phage composition induce changes in the overall gut bacterial composition, thus promoting the development of adenomas and potentially leading to carcinogenesis. Our findings will enhance the understanding of enterovirus profiles among individuals with colorectal adenomas, thus facilitating further investigation into virome-level mechanisms involved in CRC carcinogenesis. Additionally, based on the findings of this study, exploring enterovirus biomarkers in patients

with adenoma would be beneficial for early detection and improved clinical efficacy in managing CRC patients. However, this study does have several limitations. On one hand, our model did not perform optimally in the validation cohort, which may be attributed to population heterogeneity; therefore, a larger sample size is needed to further refine the model's performance and enhance its predictive accuracy. On the other hand, this study focused solely on the association between viruses and adenoma development, without establishing causal relationships. Future work should incorporate intervention experiments using animal models to validate these findings.

Abbreviations

vOTU	Viral operational taxonomic unit
CRC	Colorectal cancer
TA	Tubular adenomas
SSA	Sessile serrated adenomas
IBD	Inflammatory bowel disease
SRA	Sequence read archive
dbRDA	Distance-based redundancy analysis
AUC	Area under the receiver operating characteristic curve
KOs	KEGG orthologs
VLP	Viral-like particle

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-025-06404-7>.

Supplementary Material 1

Acknowledgements

We sincerely thank all the staff of the Department of Gastroenterology at the Second Affiliated Hospital of Xi'an Jiaotong University for their support and contributions to this study.

Author contributions

PZ, HS, LL, SM, JJ, YZ, JZ, XT, CD were involved in the design of the study. PZ, GZ, YC, LS, YY, RG, HZ¹ (Huan Zhang), and HZ² (Hongli Zhao) carried out the statistical analyses and visualizations. Interpretation of data was performed by PZ, JJ, YZ, YC, YY, RG, HZ¹, JZ, and XT. The first draft of the manuscript was written by PZ, HS, JJ, YZ, JZ, XT, and CD. The manuscript was revised by PZ, HS, LL, SM, GZ, YC, LS, YY, RG, and HZ².

Funding

This work was supported by Natural Science Basic Research Program of Shaanxi, China (Program No.S2024-JC-QN-1554).

Availability of data and materials

All data generated are available in the article and supplementary data files, as well as from the corresponding authors upon request.

Declarations

Ethical approval and consent to participate

Not applicable. This study involves secondary analysis of publicly available data, hence no ethical approval was required.

Consent for publication

All authors have agreed to this manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Gastroenterology, The Second Affiliated Hospital of Xi'an Jiaotong University, No. 157 Xiwu Road, Xi'an 710004, Shaanxi, China. ²Shaanxi Key Laboratory of Gastrointestinal Motility Disorders, Xi'an 710004, Shaanxi, China. ³Shaanxi Provincial Clinical Research Center for Gastrointestinal Diseases, Xi'an 710004, Shaanxi, China. ⁴Digestive Disease Quality Control Center of Shaanxi Province, Xi'an 710004, Shaanxi, China. ⁵Puensum Genetech Institute, Wuhan 430076, China. ⁶Department of Medical Oncology, The Second Affiliated Hospital Xi'an Jiaotong University, Xi'an 710004, Shaanxi, China. ⁷Department of Critical Care Nephrology and Blood Purification, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, Shaanxi, China.

Received: 10 October 2024 Accepted: 20 March 2025

Published online: 28 May 2025

References

- Center MM, Jemal A, Smith RA, Ward E. Worldwide variations in colorectal cancer. *Ca Cancer J Clin*. 2009;59:366–78.
- Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*. 2017;66:683–91.
- Marmol I, Sanchez-de-Diego C, Pradilla Dieste A, Cerrada E, Rodriguez YM. Colorectal carcinoma: a general overview and future perspectives in colorectal cancer. *Int J Mol Sci*. 2017;18:197.
- Sharma R. An examination of colorectal cancer burden by socioeconomic status: evidence from GLOBOCAN. 2018. *EPMA J*. 2020;11:95–117.
- Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. *Transl Oncol*. 2021;14: 101174.
- Fujita K, Yamamoto H, Matsumoto T, Hirahashi M, Gushima M, Kishimoto J, et al. Sessile serrated adenoma with early neoplastic progression: a clinicopathologic and molecular study. *Am J Surg Pathol*. 2011;35:295–304.
- Lochhead P, Chan AT, Giovannucci E, Fuchs CS, Wu K, Nishihara R, et al. Progress and opportunities in molecular pathological epidemiology of colorectal premalignant lesions. *Am J Gastroenterol*. 2014;109:1205–14.
- Terry MB, Neugut AI, Bostick RM, Sandler RS, Haile RW, Jacobson JS, et al. Risk factors for advanced colorectal adenomas: a pooled analysis. *Cancer Epidemiol Biomark Prev*. 2002;11:622–9.
- Edwards TL, Shrubsole MJ, Cai Q, Li G, Dai Q, Rex DK, et al. Genome-wide association study identifies possible genetic risk factors for colorectal adenomas. *Cancer Epidemiol Biomark Prev*. 2013;22:1219–26.
- Gupta S, Jacobs ET, Baron JA, Lieberman DA, Murphy G, Ladabaum U, et al. Risk stratification of individuals with low-risk colorectal adenomas using clinical characteristics: a pooled analysis. *Gut*. 2017;66:446–53.
- Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, et al. Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat Commun*. 2015;6:6528.
- Nakatsu G, Zhou H, Wu WKK, Wong SH, Coker OO, Dai Z, et al. Alterations in enteric virome are associated with colorectal cancer and survival outcomes. *Gastroenterology*. 2018;155:529–41.e5.
- Dahmus JD, Kotler DL, Kastenberger DM, Kistler CA. The gut microbiome and colorectal cancer: a review of bacterial pathogenesis. *J Gastrointest Oncol*. 2018;9:769–77.
- Lee JWJ, Plichta DR, Asher S, Delsignore M, Jeong T, McGoldrick J, et al. Association of distinct microbial signatures with premalignant colorectal adenomas. *Cell Host Microb*. 2023;31:827–38.e3.
- Yang Y, Du L, Shi D, Kong C, Liu J, Liu G, et al. Dysbiosis of human gut microbiome in young-onset colorectal cancer. *Nat Commun*. 2021;12:6757.
- Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med*. 2019;25:679–89.
- Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microb*. 2013;14:207–15.
- Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/beta-catenin signaling via its FadA adhesin. *Cell Host Microb*. 2013;14:195–206.
- Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M, et al. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat Med*. 2019;25:667–78.
- Hannigan GD, Duhaime MB, Ruffin MT, Koupouras CC, Schloss PD. Diagnostic potential and interactive dynamics of the colorectal cancer virome. *mBio*. 2018;9:e02248-18.
- Chen F, Li S, Guo R, Song F, Zhang Y, Wang X, et al. Meta-analysis of fecal viromes demonstrates high diagnostic potential of the gut viral signatures for colorectal cancer and adenoma risk assessment. *J Adv Res*. 2023;49:103–14.
- Clooney AG, Sutton TDS, Shkoporov AN, Holohan RK, Daly KM, O'Regan O, et al. Whole-virome analysis sheds light on viral dark matter in inflammatory bowel disease. *Cell Host Microb*. 2019;26(6):764–78.e5.
- Guo R, Li S, Zhang Y, Zhang Y, Wang G, Ullah H, et al. Dysbiotic oral and gut viromes in untreated and treated rheumatoid arthritis patients. *Microbiol Spectrum*. 2022;10: e0034822.
- Chen C, Yan Q, Yao X, Li S, Lv Q, Wang G, et al. Alterations of the gut virome in patients with systemic lupus erythematosus. *Front Immunol*. 2022;13:1050895.
- Li C, Zhang Y, Yan Q, Guo R, Chen C, Li S, et al. Alterations in the gut virome in patients with ankylosing spondylitis. *Front Immunol*. 2023;14:1154380.
- Hsu CL, Duan Y, Fouts DE, Schnabl B. Intestinal virome and therapeutic potential of bacteriophages in liver disease. *J Hepatol*. 2021;75:1465–75.
- Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:1884–90.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Method*. 2012;9:357–9.
- Nayfach S, Paez-Espino D, Call L, Low SJ, Sberro H, Ivanova NN, et al. Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. *Nat Microbiol*. 2021;6:960–70.
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf*. 2010;11:119.
- Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, et al. Linking virus genomes with host taxonomy. *Viruses*. 2016;8:66.
- Guerin E, Shkoporov A, Stockdale SR, Clooney AG, Ryan FJ, Sutton TDS, et al. Biology and taxonomy of crAss-like bacteriophages, the most abundant virus in the human gut. *Cell Host Microbe*. 2018;24:653–64.e6.
- Benler S, Yutin N, Antipov D, Rayko M, Shmakov S, Gussow AB, et al. Thousands of previously unknown phages discovered in whole-community human gut metagenomes. *Microbiome*. 2021;9:78.
- Li S, Guo R, Zhang Y, Li P, Chen F, Wang X, et al. A catalogue of 48,425 nonredundant viruses from oral metagenomes expands the horizon of the human oral virome. *iScience*. 2022;25:104418.
- Wang G, Li S, Yan Q, Guo R, Zhang Y, Chen F, et al. Optimization and evaluation of viral metagenomic amplification and sequencing procedures toward a genome-level resolution of the human fecal DNA virome. *J Adv Res*. 2023;48:75–86.
- Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucl Acid Res*. 2017;45:D353–61.
- Blanco-Míguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, Zolfo M, et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat Biotechnol*. 2023;41:1633–44.
- Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci*. 2003;14:927–30.
- Su G, Morris JH, Demchak B, Bader GD. Biological network exploration with Cytoscape 3. *Curr Protoc Bioinf*. 2014. <https://doi.org/10.1002/0471250953.bi0813s47>.
- Yan Q, Wang Y, Chen X, Jin H, Wang G, Guan K, et al. Characterization of the gut DNA and RNA viromes in a cohort of Chinese residents and visiting Pakistanis. *Virus Evol*. 2021;7:veab022.
- Lazarevic V, Margot P, Soldo B, Karamata D. Sequencing and analysis of the *Bacillus subtilis* lytRABC divergon: a regulatory unit encompassing the structural genes of the N-acetylmuramoyl-L-alanine amidase and its modifier. *J Gen Microbiol*. 1992;138:1949–61.

42. Markowitz SD, Bertagnolli MM. Molecular origins of cancer: Molecular basis of colorectal cancer. *N Engl J Med*. 2009;361:2449–60.
43. Monaco CL, Gootenberg DB, Zhao G, Handley SA, Ghebremichael MS, Lim ES, et al. Altered virome and bacterial microbiome in human immunodeficiency virus-associated acquired immunodeficiency syndrome. *Cell Host Microb*. 2016;19:311–22.
44. Norman JM, Handley SA, Baldrige MT, Droit L, Liu CY, Keller BC, et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell*. 2015;160:447–60.
45. Avellaneda-Franco L, Dahlman S, Barr JJ. The gut virome and the relevance of temperate phages in human health. *Front Cell Infect Microbiol*. 2023;13:1241058.
46. Hsu BB, Gibson TE, Yeliseyev V, Liu Q, Lyon L, Bry L, et al. Dynamic modulation of the gut microbiota and metabolome by bacteriophages in a mouse model. *Cell Host Microb*. 2019;25:803–14.e5.
47. Nguyen S, Baker K, Padman BS, Patwa R, Dunstan RA, Weston TA, et al. Bacteriophage transcytosis provides a mechanism to cross epithelial cell layers. *mBio*. 2017;8:e01874–17.
48. Lehti TA, Pajunen MI, Skog MS, Finne J. Internalization of a polysialic acid-binding *Escherichia coli* bacteriophage into eukaryotic neuroblastoma cells. *Nat Commun*. 2017;8:1915.
49. Li G, Jin Y, Chen B, Lin A, Wang E, Xu F, et al. Exploring the relationship between the gut mucosal virome and colorectal cancer: characteristics and correlations. *Cancers (Basel)*. 2023;15:3555.
50. Sakamoto M, Iino T, Ohkuma M. *Faecalimonas umbilicata* gen. nov., sp. nov., isolated from human faeces, and reclassification of *Eubacterium contortum*, *Eubacterium fissicatena* and *Clostridium oroticum* as *Faecalicatena contorta* gen. nov., comb. nov., *Faecalicatena fissicatena* comb. nov. and *Faecalicatena orotica* comb. nov. *Int J Syst Evol Microbiol*. 2017;67:1219–27.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.