Viral and Bacterial Communities of Colorectal Cancer

Geoffrey D Hannigan1, Melissa B Duhaime2, Mack T Ruffin IV3, Charlie C Koumpouras1, and Patrick D Schloss1,\*

1Department of Microbiology & Immunology, University of Michigan, Ann Arbor, Michigan, 48109 2Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, 48109 3Department of Family and Community Medicine, Pennsylvania State University Hershey Medical Center, Hershey, Pennsylvania, 17033 \*To whom correspondence may be addressed.

***Corresponding Author Information***  
Patrick D Schloss, PhD  
1150 W Medical Center Dr. 1526 MSRB I  
Ann Arbor, Michigan 48109  
Phone: (734) 647-5801  
Email: pschloss@umich.edu

***Journal***: PNAS ([*Preparation Details*](http://www.pnas.org/site/authors/procedures.xhtml))  
***Major Classification***: Biological Sciences  
***Minor Classification***: Microbiology

***Keywords***: Colorectal Cancer, Virome, Machine Learning

\* *Figures included for internal editing purposes*

# Abstract

Colorectal cancer is the second leading cause of cancer-related death in the United States and is a primary cause of morbidity and mortality throughout the world. Its risk has been linked to changes in colonic bacterial community composition. Viruses are another important component of the colonic microbial community, however they have yet to be studied in colorectal cancer despite their oncogenic potential. We evaluated the colorectal cancer virome (virus community) in stool using a cohort of 90 human subjects with either healthy, adenomatous (precancerous), or cancerous colons. We utilized 16S rRNA gene, whole shotgun metagenomic, and purified virus metagenomic sequencing methods to compare the colorectal cancer virome to the bacterial community. We identified no detectable difference in diversity (alpha or beta) between healthy, adenomatous, or cancerous colonic samples, but more sophisticated random forest models identified striking changes in the virus community. The majority of the cancer-associated virome consisted of temperate bacteriophages, suggesting that the community was indirectly linked to colorectal cancer by modulating bacterial community structure and function. Our data suggested that the influential phages did not exclusively infect influential bacteria, but rather acted through the community as a whole. These results provide foundational evidence that bacteriophage communities are associated with colorectal cancer and likely impact cancer progression by altering the bacterial host communities.

**Word Count**: 213 / 250

# Significance Statement

Colorectal cancer is a leading cause of cancer-related death in the United States and worldwide. Its risk and severity have been linked to colonic bacterial community composition. Although viruses have been linked to other cancers and diseases, little is known about colorectal cancer virus communities. We addressed this knowledge gap by identifying changes in colonic virus communities in the stool of colorectal cancer patients and how they compared to bacterial community changes. The results suggested an indirect role for the virome in impacting colorectal cancer by modulating their associated bacterial community. These findings both support a biological role for viruses in colorectal cancer and provide a new understanding of basic colorectal cancer etiology.

**Word Count**: 111 / 120

\body

# Introduction

Due to their mutagenic abilities and propensity for functional manipulation, human viruses are strongly associated with, and in many cases cause, cancer (1–4). Because bacteriophages are crucial for bacterial community stability and composition (5–7) and have been implicated as oncogenic agents (8–11), bacteriophages have the potential to indirectly impact cancer. The gut virome (the virus community of the gut, including eukaryotic and prokaryotic viruses) therefore has the potential to impact health and disease (e.g. cancer). Altered human virome composition and diversity have been identified in diseases including periodontal disease (12), HIV (13), cystic fibrosis (14), antibiotic exposure (15, 16), urinary tract infections (17), and inflammatory bowel disease (18). The strong association of bacterial communities with colorectal cancer and the precedent for the virome to impact other human diseases suggest that colorectal cancer may be associated with altered virus communities.

Colorectal cancer is the second leading cause of cancer-related deaths in the United States (19). The US National Cancer Institute estimates over 1.5 million Americans have been diagnosed with colorectal cancer in 2016 and over 500,000 Americans have died from the disease (19). An important component of colorectal cancer etiology is variation in colorectal bacterial community composition (8, 10, 11, 20, 21). Work in this area has led to a proposed disease model in which bacteria colonize the colon, develop biofilms, promote inflammation, and enter an oncogenic synergy with the cancerous human cells (22). This association also has allowed researchers to leverage bacterial community signatures as biomarkers to provide accurate, noninvasive colorectal cancer detection from stool (8, 23, 24). While an understanding of colorectal cancer bacterial communities has proven fruitful both for disease classification and for identifying the underlying disease etiology, bacteria are only a subset of the colon microbiome. Viruses are another important component of the colon microbial community that have yet to be studied in the context of colorectal cancer. We evaluated disruptions in virus and bacterial community composition in a human cohort whose stool was sampled at the three relevant stages of cancer development: healthy, adenomatous, and cancerous.

Colorectal cancer progresses in a stepwise process that begins when healthy tissue develops into a precancerous polyp (i.e., adenoma) in the large intestine (25). If left untreated, the adenoma may develop into a cancerous lesion that can invade and metastasize, leading to severe illness and death. Progression to cancer can be prevented when precancerous adenomas are detected and removed during routine screening (26, 27). Survival for colorectal cancer patients may exceed 90% when the lesions are detected early and removed (26). Thus, work that aims to facilitate early detection and prevention of progression beyond early cancer stages has great potential to inform therapeutic development.

Here we address the knowledge gap of whether virus community composition is altered in colorectal cancer and, if it is, how those changes might impact cancer progression and severity. We also aimed to evaluate the virome's potential for use as a diagnostic biomarker. The implications of this study are threefold. *First*, this work supports a biological role for the virome in colorectal cancer development and suggests that more than the bacterial members of the associated microbial communities are involved in the process. *Second*, we present a supplementary, or even alternative, virus-based approach for classification modeling of colorectal cancer using stool samples. *Third*, we provide initial support for the importance of studying the virome as a component of the microbiome ecological network, especially in cancer.

# Results

## Cohort Design, Sample Collection, and Processing

Our study cohort consisted of 90 human subjects, 30 of whom had healthy colons, 30 of whom had adenomas, and 30 of whom had carcinomas **(Figure )**. Half of each stool sample was used to sequence the bacterial communities using both 16S rRNA gene and shotgun sequencing techniques. The other half of each stool sample was purified for virus like particles (VLPs) before genomic DNA extraction and shotgun metagenomic sequencing. In the VLP purification, cells were disrupted and extracellular DNA degraded **(Figure )** to allow the exclusive analysis of viral DNA within virus capsids. In this manner, the *extracellular virome* of encapsulated viruses was targeted.

Each extraction was performed with a blank buffer control to detect contaminants from reagents or other unintentional sources. Only one of the nine controls contained detectable DNA at a minimal concentration of 0.011 ng/µl, thus providing evidence of the enrichment and purification of VLP genomic DNA over potential contaminants **(Figure A)**. As was expected, these controls yielded few sequences and were almost entirely removed while rarefying the datasets to a common number of sequences **(Figure B)**. The high quality phage and bacterial sequences were assembled into highly covered contigs longer than 1 kb **(Figure )**. Because contigs represent genome fragments, we further clustered related bacterial contigs into operational genomic units (OGUs) and viral contigs into operational viral units (OVUs) **(Figure - )** to approximate organismal units.

## Unaltered Virome Diversity in Colorectal Cancer

Microbiome and disease associations are often described as being of an altered diversity (i.e., "dysbiotic"). We therefore first evaluated the influence of colorectal cancer on virome OVU diversity. We evaluated differences in communities between disease states using the Shannon entropy, richness, and Bray-Curtis metrics. We observed no significant alterations in either Shannon entropy or richness in the diseased states as compared to the healthy state **(Figure C-D)**. There was no statistically significant clustering of the disease groups (ANOSIM p-value = 0.4, **Figure** ). Notably, there was a significant difference between the few blank controls that remained after rarefying the data and the other study groups (ANOSIM p-value < 0.001, **Figure )**, further supporting the quality of the sample set. In summary, standard alpha and beta diversity metrics were insufficient for capturing virus community differences between disease states **(Figure )**. This is consistent with what has been observed when the same metrics were applied to metagenomic samples (24) and points to the need for alternate approaches to detect the impact of colorectal cancer disease state on these communities.

## Altered Virome Composition in Colorectal Cancer

As opposed to the diversity metrics discussed above, 16S rRNA gene relative abundance profiles are effective feature sets for classifying stool samples as originating from individuals with healthy, adenomatous, or cancerous colons (8, 23). The exceptional performance of bacteria in these classification models supports a role for bacteria in colorectal cancer. We built off of these findings by evaluating the ability of virus community signatures to classify stool samples and compared their performance to models built using bacterial community signatures.

To identify the altered virus communities associated with colorectal cancer, we built and tested random forest models for classifying stool samples as belonging to individuals with either cancerous or healthy colons. We confirmed that our bacterial 16S rRNA gene model replicated the performance of the original report which used logit models instead of random forest models **(Figure A)** (8). We then compared the bacterial 16S rRNA gene model to a model built using virome relative abundance. The viral model performed as well as the bacterial model (corrected p-value = 0.4), with the viral and bacterial models achieving mean AUC (area under the curve) values of 0.793 and 0.796, respectively **(Figure A - B)**.

To evaluate the ability of both bacterial and viral biomarkers to classify samples, we built a combined model that used both bacterial and viral community data. The combined model yielded a modest but statistically significant performance improvement beyond the viral (corrected p-value = 0.002) and bacterial (corrected p-value = 0.002) models, yielding an AUC of 0.816 **(Figure A - B)**. The combined features from the virus and bacterial communities improved our ability to classify stool as belonging to individuals with cancerous colons.

To determine the advantage of viral metagenomic methods over bacterial metagenomic methods, we compared the viral model to a model built using relative abundance profiles from bacterial metagenomic shotgun sequencing data. This model performed worse than the other models (mean AUC = 0.505) **(Figure A - B)**. Further investigation revealed that the bacterial 16S rRNA gene model was strongly driven by sparse and low abundance OTUs **(Figure )**. Removal of OTUs with a median abundance of zero resulted in the removal of six OTUs, and a loss of model performance down to what was observed in the metagenome-based model **(Figure A)**. The majority of these OTUs had a relative abundance lower than 1% across the samples **(Figure B)**. Although the features in the viral model also were of low abundance **(Figure F)**, the coverage was sufficient for high model performance, likely because viral genomes are orders of magnitude smaller than bacterial genomes.

The association between the bacterial and viral communities and colorectal cancer was driven by a few important microbes, measured using the mean decrease in model accuracy when each was removed. *Fusobacterium* was the primary driver of the bacterial association with colorectal cancer, which was consistent with its previously described oncogenic potential **(Figure C)**(22). The virome signature also was driven by a few OVUs, suggesting a role for these viruses in cancer development **(Figure D)**. The identified viruses were bacteriophages, belonging to *Siphoviridae*, *Myoviridae*, and "unclassified" phage taxa. Many of the important viruses were unidentifiable (denoted "unknown"). This is common in viromes across habitats; studies have reported as much as 95% of virus sequences belonging to unknown genomic units (14, 28–30). When the bacterial and viral community signatures were combined, both bacterial and viral organisms drove the community association with cancer **(Figure E)**.

## Shifted Phage Influence Between Cancer Progression Stages

Because previous work has identified shifts in which bacteria were most important at different stages of colorectal cancer (8, 20, 22), we explored whether shifts in the relative influence of specific phages could be detected between healthy, adenomatous, and cancerous colons. We evaluated community shifts between the two disease stage transitions (healthy to adenomatous and adenomatous to cancerous) by building random forest models to compare only the diagnosis groups around the transitions. While bacterial 16S rRNA gene models performed equally well for all disease class comparisons, the virome model performances differed **(Figure A-B)**. Like bacteria **(Figure F-H)**, different virome members were important between the healthy to adenomatous and adenomatous to cancerous stages **(Figure C-E)**.

After evaluating our ability to classify samples between two disease states, we performed a three-class random forest model including all disease states. The 16S rRNA gene model yielded a mean AUC of 0.771 and outperformed the viral community model, which yielded a mean AUC of 0.699 (p-value < 0.001, **Figure A-C**). The microbes important for the cancer versus healthy and healthy versus adenoma models were also important for the three-class model **(Figure D-E)**. The most important bacterium in the two and three class models was the same *Fusobacterium* (OTU 4) **(Figure C, Figure D)**. The viruses most important to the three-class model were identified as bacteriophages **(Figure D, Figure E)**, but not all important OVUs were of increased abundance in the diseased state **(Figure F)**.

## Bacteriophage Dominance in Colorectal Cancer Virome

Changes in the colorectal cancer virome could have been driven directly by eukaryotic viruses or indirectly by bacteriophages acting through their bacterial hosts. To better understand the types of viruses that were important for colorectal cancer, we identified the virome OVUs as being similar to either eukaryotic viruses or bacteriophages. The most important viruses to the classification model were identified as bacteriophages (**Figure )**. Overall, we were able to identify 78.8% of the OVUs as known viruses, and 93.8% of those viral OVUs aligned to bacteriophage reference genomes. It is important to note that this could have been influenced by our methodological biases against enveloped viruses (more common of eukaryotic viruses than bacteriophage), due to chloroform and DNase treatment for purification.

We evaluated whether the phages in the community were primarily lytic (obligately lyse their hosts after replication) or temperate (able to integrate into their host's genome to form a lysogen, and subsequently transition to a lytic mode). We accomplished this by identifying three markers for temperate phages in the OVU representative sequences: 1) presence of phage integrase genes, 2) presence of known prophage genes, according the the ACLAME (A CLAssification of Mobile genetic Elements) database, and 3) nucleotide similarity to regions of bacterial genomes, as previously described (29, 31, 32). We found that the majority of the colon phages were temperate and that the overall fraction of temperate phages remained consistent throughout the healthy, adenomatous, and cancerous stages **(Figure E)**. Thus, the majority of the OVUs were temperate bacteriophages and not eukaryotic viruses, indicating that the association between the virome and colorectal cancer was reliant on bacteriophage communities that have the ability to integrate into host bacterial genomes. These findings were consistent with previous reports suggesting the gut virome is primarily composed of temperate phages (13, 18, 31, 33).

## Community Context of Influential Phages

Because the link between colorectal cancer and the virome was driven by bacteriophages, we hypothesized that the influential phages were primarily predators of the influential bacteria, and thus influenced their relative abundance through predation. If this hypothesis were true, we would expect a correlation between the relative abundances of influential bacteria and phages. Instead, we observed a strikingly low correlation between bacterial and phage relative abundances **(Figure A,C)**. Overall, there was an absence of correlation between the most influential OVUs and bacterial OTUs **(Figure B)**. This evidence supported our null hypothesis that the influential phages were not primarily predators of influential bacteria.

Given these findings, we hypothesized that the most influential phages were acting by infecting a wide range of bacteria in the overall community, instead of just the influential bacteria. In other words, we hypothesized that the influential bacteriophages were community hubs (central members) within the bacteria and phage interactive network. We investigated the potential host ranges of all phage OVUs using a previously developed random forest model that relies on sequence features to predict which phages infected which bacteria in the community **(Figure A)** (34). The predicted interactions were then used to identify phage community hubs. We calculated the alpha centrality (measure of importance in the ecological network) of each phage OVU's connection to the rest of the network. The phages with high centrality values were defined as community hubs. Next, the centrality of each OVU was compared to its importance in the colorectal cancer classification model. Phage OVU centrality was significantly and positively correlated with importance to the disease model (p-value = 0.02, R = 0.14), suggesting that phages important in driving colorectal cancer also were more likely to be community hubs **(Figure B)**. Together these findings supported our hypothesis that influential phages were hubs within their microbial communities.

# Discussion

Because of their propensity for mutagenesis and capacity for modulating their host functionality, many viruses are oncogenic (1–4). Some bacteria also have oncogenic properties, suggesting that bacteriophages may play an indirect role in promoting carcinogenesis by influencing bacterial community composition and dynamics (8–10). Despite their carcinogenic potential and the strong association between bacteria and colorectal cancer, the link between virus colorectal communities and colorectal cancer has yet to be evaluated. Here we show that, like colonic bacterial communities, the colon virome was altered in colorectal cancer. Our findings support a working hypothesis for oncogenesis by phage-modulated bacterial community composition.

Here, we have begun to delineate the role the colonic virome plays in colorectal cancer **(Figure A)**. We found that basic diversity metrics of alpha diversity (richness and Shannon entropy) and beta diversity (Bray-Curtis dissimilarity) were insufficient for identifying virome community changes between healthy and cancerous states. By implementing a more sophisticated machine learning approach (random forest classification), we detected strong associations between the colon virus community composition and colorectal cancer. The colorectal cancer virome was composed primarily of bacteriophages. These phage communities were not exclusively predators of the most influential bacteria, as demonstrated by the lack of correlation between the abundances of the bacterial and phage populations. Instead, we identified influential phages as being community hubs, suggesting phages influence cancer by altering the greater bacterial community instead of directly modulating the influential bacteria. Our previous work has shown that modifying colon bacterial communities alters colorectal cancer progression and tumor burden in mice (10, 20). This provides a precedent for phage indirectly influencing colorectal cancer progression by altering the bacterial community composition. Overall, our data support a model in which the bacteriophage community modulates the bacterial community, and through those interactions indirectly influences the bacteria driving colorectal cancer progression **(Figure A)**. Although our evidence suggested phages indirectly influenced colorectal cancer development, we were not able to rule out the role of phages directly interacting with the human host (35, 36).

In addition to modeling the potential connections between virus communities, bacteria communities, and colorectal cancer, we also used our data and existing knowledge of phage biology to develop a working hypothesis for the mechanisms by which this may occur. This was done by incorporating our findings into the current model for colorectal cancer development **(Figure B)** (22). We hypothesize that the process began with broadly infectious phages in the colon lysing and thereby disrupting the existing bacterial communities. This shift led to novel niche space that enabled opportunistic bacteria (such as *Fusobacterium nucleatum*) to colonize. Once the initial influential founder bacteria established themselves in the epithelium, secondary opportunistic bacteria were able to adhere to the founders, colonize, and begin establishing a biofilm. Phages may have played a role in biofilm dispersal and growth by lysing bacteria within the biofilm, a process important for effective biofilm growth (37). The oncogenic bacteria may then have been able to transform the epithelial cells and disrupt tight junctions to infiltrate the epithelium, thereby initiating an inflammatory immune response. As the adenomatous polyps developed and progressed towards carcinogenesis, we observed a shift in the phages and bacteria whose relative abundances were most influential. As the bacteria entered their oncogenic synergy with the epithelium, we conjecture that the phages continued mediating biofilm dispersal. This process would thereby support the colonized oncogenic bacteria by lysing competing cells and releasing nutrients to other bacteria in the form of cellular lysates. In addition to highlighting the likely mechanisms by which the colorectal cancer virome is interacting with the bacterial communities, this outline will guide future research investigations of the role the virome plays colorectal cancer.

A notable finding was the poor performance observed using bacterial metagenomic methods compared to the performance of models using viral metagenomes or 16S rRNA gene sequences. We believe this observation speaks to the importance of sequencing coverage in microbial community studies and the advantage of the high coverage of 16S rRNA gene sequencing relative to the lower per OTU coverage possible using whole metagenomic shotgun sequencing. To demonstrate this concept, consider that six bacterial OTUs drove the performance of the 16S rRNA gene classification model and these OTUs were all sparsely present and lowly abundant. Filtration of OTUs with a median relative abundance of zero resulted in the removal of the six important OTUs and reduced model performance to being nearly random like the bacterial metagenomic model. The bacterial metagenomic OGUs represented only the most abundant taxa, which was not informative for this application. There has been some success in using shotgun metagenomic approaches for stool colorectal cancer classification (24), but this previous approach relied on lowly abundant signatures and did not utilize OGU clustering, as done here. In that former case, the models only performed as well as the 16S rRNA gene model (24). Thus, the targeted 16S rRNA gene sequencing approach, which yielded only a fraction of the bacterial metagenomic sequences, was more effective for detecting colorectal cancer in stool samples. Despite a loss of enthusiasm for 16S rRNA gene sequencing in favor of shotgun metagenomic techniques, 16S rRNA gene sequencing is still a superior methodological approach for some important applications.

In addition to the therapeutic ramifications for understanding the colorectal cancer microbiome, our findings provide a proof-of-principle that viruses, while understudied and currently under-appreciated in the human microbiome, are an important contributor to human disease. Viral community dynamices have the potential to provide an abundance of information to supplement those of bacterial communities. Evidence has suggested that the virome is a crucial component to the microbiome and that bacteriophages are important players. Bacteriophage and bacterial communities cannot maintain stability and co-evolution without one another (6, 38). Not only is the human virome an important element to consider in human health and disease (12–18), but our findings support that it is likely to have a significant impact on cancer etiology and progression.

# Methods

## Analysis Source Code & Data Availability

All study sequences are available on the NCBI Sequence Read Archive under the BioProject ID PRJNA389927.

All associated source code is available at the following GitHub repository:

https://github.com/SchlossLab/Hannigan\_CRCVirome\_PNAS\_2017

## Study Design and Patient Sampling

This study was approved by the University of Michigan Institutional Review Board and all subjects provided informed consent. Design and sampling of this sample set have been reported previously (8). Briefly, whole evacuated stool was collected from patients who were 18 years of age or older, able to provide informed consent, have had colonoscopy and histologically confirmed colonic disease status, had not had surgery, had not had chemotherapy or radiation, and were free of known co-morbidities including HIV, chronic viral hepatitis, HNPCC, FAP, and inflammatory bowel disease. Samples were collected from four geographic locations: Toronto (Ontario, Canada), Boston (Massachusetts, USA), Houston (Texas, USA), and Ann Arbor (Michigan, USA). Ninety patients were recruited to the study, thirty of which were designated healthy, thirty with detected adenomas, and thirty with detected carcinomas.

# Conflicts of Interest

The authors declare no conflicts of interest.

# Acknowledgments

The authors thank the Schloss lab members for their underlying contributions, and the Great Lakes-New England Early Detection Research Network for providing the fecal samples that were used in this study. GD Hannigan was supported in part by the Molecular Mechanisms in Microbial Pathogenesis Training Program (T32 AI007528). PD Schloss was supported by funding from the National Institutes of Health (P30DK034933). MT Ruffin was supported by funding from the National Institutes of Health (5U01CA86400).

# References

1. Feng H, Shuda M, Chang Y, Moore PS (2008) Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* 319(5866):1096–1100.

2. Shuda M, Kwun HJ, Feng H, Chang Y, Moore PS (2011) Human Merkel cell polyomavirus small T antigen is an oncoprotein targeting the 4E-BP1 translation regulator. *Journal of Clinical Investigation* 121(9):3623–3634.

3. Schiller JT, Castellsagué X, Garland SM (2012) A review of clinical trials of human papillomavirus prophylactic vaccines. *Vaccine* 30 Suppl 5:F123–38.

4. Chang Y, et al. (1994) Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi’s sarcoma. *Science* 266(5192):1865–1869.

5. Harcombe WR, Bull JJ (2005) Impact of phages on two-species bacterial communities. *Applied and Environmental Microbiology* 71(9):5254–5259.

6. Rodriguez-Valera F, et al. (2009) Explaining microbial population genomics through phage predation. *Nature Reviews Microbiology* 7(11):828–836.

7. Cortez MH, Weitz JS (2014) Coevolution can reverse predator-prey cycles. *Proceedings of the National Academy of Sciences of the United States of America* 111(20):7486–7491.

8. Zackular JP, Rogers MAM, Ruffin MT, Schloss PD (2014) The human gut microbiome as a screening tool for colorectal cancer. *Cancer prevention research (Philadelphia, Pa)* 7(11):1112–1121.

9. Garrett WS (2015) Cancer and the microbiota. *Science* 348(6230):80–86.

10. Baxter NT, Zackular JP, Chen GY, Schloss PD (2014) Structure of the gut microbiome following colonization with human feces determines colonic tumor burden. *Microbiome* 2(1):20.

11. Arthur JC, et al. (2012) Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* 338(6103):120–123.

12. Ly M, et al. (2014) Altered Oral Viral Ecology in Association with Periodontal Disease. *mBio* 5(3):e01133–14–e01133–14.

13. Monaco CL, et al. (2016) Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. *Cell Host and Microbe* 19(3):311–322.

14. Willner D, et al. (2009) Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLOS ONE* 4(10):e7370.

15. Abeles SR, Ly M, Santiago-Rodriguez TM, Pride DT (2015) Effects of Long Term Antibiotic Therapy on Human Oral and Fecal Viromes. *PLOS ONE* 10(8):e0134941.

16. Modi SR, Lee HH, Spina CS, Collins JJ (2013) Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* 499(7457):219–222.

17. Santiago-Rodriguez TM, Ly M, Bonilla N, Pride DT (2015) The human urine virome in association with urinary tract infections. *Frontiers in Microbiology* 6:14.

18. Norman JM, et al. (2015) Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* 160(3):447–460.

19. Siegel R, Desantis C, Jemal A (2014) Colorectal cancer statistics, 2014. *CA: a cancer journal for clinicians* 64(2):104–117.

20. Zackular JP, Baxter NT, Chen GY, Schloss PD (2016) Manipulation of the Gut Microbiota Reveals Role in Colon Tumorigenesis. *mSphere* 1(1):e00001–15.

21. Dejea CM, et al. (2014) Microbiota organization is a distinct feature of proximal colorectal cancers. *Proceedings of the National Academy of Sciences of the United States of America* 111(51):18321–18326.

22. Flynn KJ, Baxter NT, Schloss PD (2016) Metabolic and Community Synergy of Oral Bacteria in Colorectal Cancer. *mSphere* 1(3):e00102–16.

23. Baxter NT, Ruffin MT, Rogers MAM, Schloss PD (2016) Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome medicine* 8(1):37.

24. Zeller G, et al. (2014) Potential of fecal microbiota for early-stage detection of colorectal cancer. *Molecular systems biology* 10(11):766–766.

25. Fearon ER (2011) Molecular genetics of colorectal cancer. *Annual review of pathology* 6(1):479–507.

26. Levin B, et al. (2008) Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA: A Cancer Journal for Clinicians* (The University of Texas MD Anderson Cancer Center, Houston, TX, USA. John Wiley & Sons, Ltd.), pp 130–160.

27. Zauber AG (2015) The impact of screening on colorectal cancer mortality and incidence: has it really made a difference? *Digestive diseases and sciences* 60(3):681–691.

28. Pedulla ML, et al. (2003) Origins of highly mosaic mycobacteriophage genomes. *Cell* 113(2):171–182.

29. Hannigan GD, et al. (2015) The Human Skin Double-Stranded DNA Virome: Topographical and Temporal Diversity, Genetic Enrichment, and Dynamic Associations with the Host Microbiome. *mBio* 6(5):e01578–15.

30. Brum JR, et al. (2015) Ocean plankton. Patterns and ecological drivers of ocean viral communities. *Science* 348(6237):1261498–1261498.

31. Minot S, et al. (2011) The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Research* 21(10):1616–1625.

32. Hannigan GD, et al. (2017) Evolutionary and functional implications of hypervariable loci within the skin virome. *PeerJ* 5(4):e2959.

33. Reyes A, et al. (2010) Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 466(7304):334–338.

34. Hannigan GD, Duhaime MB, Koutra D, Schloss PD (2017) Biogeography & Environmental Conditions Shape Phage & Bacteria Interaction Networks Across The Human Microbiome. *bioRxiv*:1–40.

35. Lengeling A, Mahajan A, Gally DL (2013) Bacteriophages as Pathogens and Immune Modulators? *mBio* 4(6):e00868–13–e00868–13.

36. G rski A, et al. (2012) Phage as a Modulator of Immune Responses. *Bacteriophages, Part B* (Bacteriophage Laboratory, Ludwik Hirszfeld Institute of Immunology; Experimental Therapy, Polish Academy of Sciences, Wrocław, Poland. agorski@ikp.pl; Elsevier), pp 41–71.

37. Rossmann FS, et al. (2015) Phage-mediated Dispersal of Biofilm and Distribution of Bacterial Virulence Genes Is Induced by Quorum Sensing. *PLoS Pathogens* 11(2):e1004653–17.

38. Brockhurst MA, Koskella B (2013) Experimental coevolution of species interactions. *Trends in ecology & evolution* 28(6):367–375.

39. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology* 79(17):5112–5120.

40. Schloss PD, et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* 75(23):7537–7541.

41. Pruesse E, et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* 35(21):7188–7196.

42. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27(16):2194–2200.

43. Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F (2009) Laboratory procedures to generate viral metagenomes. *Nature protocols* 4(4):470–483.

44. Kleiner M, Hooper LV, Duerkop BA (2015) Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. *BMC Genomics* 16(1):7.

45. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 17(1):10.

46. Hannon GJ FASTX-Toolkit. GNU Affero General Public License.

47. Schmieder R, Edwards R (2011) Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLOS ONE* 6(3):e17288.

48. Li D, et al. (2016) MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *METHODS* 102:3–11.

49. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9(4):357–359.

50. Alneberg J, et al. (2014) Binning metagenomic contigs by coverage and composition. *Nature Methods*:1–7.

51. Oksanen J, et al. vegan: Community Ecology Package.

52. Kuhn M caret: Classification and Regression Training.

53. Guidi L, et al. (2016) Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532(7600):465–470.

54. Camacho C, et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10(1):1.