JOURNAL OF
EMERGING INVESTIGATORS

Article

The correlation between bacteria and colorectal cancer

Qi Zhang[1] and Qin Zhao[1]
[1] Adlai E. Stevenson High School, Lincolnshire, Illinois

SUMMARY

Colorectal cancer (CRC) is the third most commonly diagnosed cancer with a high mortality rate worldwide. The global burden of CRC imposes the need to understand the mechanism of CRC progression. Accumulating evidence reveals gut bacteria may contribute to the risk of CRC. The purpose of this project was to determine whether the bacterial composition is altered in stool samples of CRC patients. We hypothesized that patients with CRC would have a set of differentially abundant bacteria species in their stool samples compared to the control samples. Using 11 CRC datasets across 9 countries, consisting of 1395 samples from the curatedMetagenomicData database, we examined the bacterial composition and compared the prevalence and relative abundance of bacteria species across 701 CRC patients and 694 control samples. We found that CRC-associated bacteria species and their prevalence varied across different demographic regions. Combining 11 datasets, we performed differential abundance analysis to analyze the prevalence and compare the relative abundance of bacteria species between CRC samples and control samples, and we found a set of CRC-associated bacteria species that were significantly more abundant than those of control samples. Many of these species have previously been found to influence CRC development. Moreover, we identified a set of control-associated species that were important in maintaining a beneficial microbiome. This project provided insights that support further investigation into the role of CRC-associated bacteria in CRC tumorigenesis.

INTRODUCTION

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and the second leading cause of cancer deaths worldwide, with about 2 million new incidences and 1 million deaths in 2020 (1, 2). The burden of CRC is expected to grow steadily, resulting in 3.2 million new cases and 1.6 million deaths by 2040 (2). The global crisis of CRC is imposing a great need to understand the mechanism of CRC progression.

Extensive research has revealed that the prevalence of CRC can be influenced by genetics, environment, lifestyle factors such as alcohol and smoking, and a Western diet (2, 3). More specifically, a diet high in animal fat and red meat with low fiber, fruits, and vegetables increases the risk of developing CRC (2). In addition to these well-known risk factors, accumulating evidence has demonstrated that the gut microenvironment can influence human health, and microbiota may be linked to CRC (4).

The gut microbiota has evolved over thousands of years and developed a beneficial relationship with humans (5, 6). More than 10^{14} microorganisms inhabit the human GI tract (4). The gut microbiota plays an important role in shaping the intestinal epithelium, harvesting energy, and regulating host immunity (7, 8). Under healthy conditions, the host and its microbiome exist in symbiosis, providing a nutrient-rich microenvironment for bacteria in return for aid in metabolism and digestion (9). By regulating human immunity, metabolic homeostasis, and protecting against pathogens, gut bacteria are often beneficial to human health. However, an altered microbiome may promote chronic inflammation, tissue impairment, and induce the progression of cancer such as CRC (10).

Since the introduction of next generation sequencing (NGS), it has been broadly applied to microbiota and host studies. Notably, 16S rRNA sequencing and metagenomic sequencing technology have enabled the discovery of more microorganisms colonizing the intestines and understanding of their roles (11). Over the last decade, scientists have studied the human gut bacteria and identified the important role of gut bacterial composition in pathological processes, such as cancer (9, 10). Clarifying which gut bacteria species are linked to CRC with high prevalence might identify CRC biomarkers, which could help predict and develop strategies to slow the progression to CRC in the future (12–14).

The scientific community is just beginning to understand the role of the bacteria in cancer, and a lot is still unknown about the effect of the gut microbiome on cancers. Thus, the relationship between gut bacteria and CRC remains an important question to address. If the gut's bacteria are associated with CRC, alterations in the bacterial composition may signify carcinogenesis. In this project, we examined the bacteria taxonomy present in CRC patients' stool samples to understand the correlation of bacteria species with CRC using the publicly available curatedMetagenomicData database (15). From stool samples from patients with CRC, we identified a set of bacteria species that were significantly more abundant than those of control samples from healthy patients, and many of these species are known to promote CRC tumorigenesis. We also found a set of control-associated bacteria species, which have been found to be important in maintaining a healthy gut microbiome.

https://doi.org/10.59720/24-084

RESULTS

We processed and analyzed metagenomic data from the curatedMetagenomicData database, which provided curated metagenomic data across multiple cohort studies, with MetaPhlAn3 to calculate bacterial taxonomic abundances. We downloaded CRC cohort datasets and used the relative abundance data to perform the downstream statistics analysis. From the curatedMetagenomicData database, we collected 11 public CRC metagenomic datasets, including a total of 1395 samples, with 694 control samples and 701 CRC patient samples. In each CRC dataset, the number of CRC patients ranged between 27 and 258, and the number of controls ranged between 24 and 251. The participants' ages ranged between 21 and 90 years old with 40% of them in their 60s. Among the 11 studies, 9 of them had deeply sequenced datasets while GuptaA_2019 and HanniganGD_2017 had lower numbers of sequenced reads with fewer controls and CRC patients than most of the other studies (**Figure 1**). This depth of sequencing data was an important quality control metric of sequencing quality, and the combined datasets provide high quality data for the following analysis.

To test our hypothesis, we compared the bacterial composition and associated abundance of each identified bacteria species in the CRC samples with those in the control samples. We then further analyzed the differentially abundant gut bacteria species and evaluated their association with CRC pathogenesis. In all cohort study datasets, we observed more bacteria species in CRC samples than controls (**Figure 2**). We calculated the prevalence of each identified species as the fraction of samples with non-zero relative abundance for both the CRC samples and control samples. Next, we performed the differential abundance analysis to identify the bacteria species that were differentially abundant between CRC samples and control samples.

CRC-associated species varied across geographic regions

We calculated and compared the prevalence of various bacteria species between CRC and control samples in each cohort dataset. Since the human gut microbiota is influenced by multiple factors such as environment, diet and eating style, and ethnic diversity, we expected that the bacterial composition would vary among different geographic regions and thus that the observation based on any particular dataset could be biased (16). Therefore, we compared the bacterial composition in the CRC and control samples at the species level using datasets from three different geographic regions: FengQ_2015 for Austria, VogtmannE_2016 for USA, and ThomasAM_2018a and ThomasAM_2018b for Italy. Bacteria species were considered unique to CRC when they were prevalent in CRC samples indicated by the presence of sequencing reads but were not found in any control sample. A total of 32 bacteria species unique to CRC patients were identified across the datasets for the three geographic regions (**Table 1**). Among those species unique to CRC, we identified 17 bacteria species common in all three regions. The remaining 15 species either only partially overlapped in the different regions or were specific to a single region.

Overall, species under three phyla—Bacteriodetes, Firmicutes, and Fusobacteria—were significantly represented in the CRC samples compared to control across the three geographic regions ($p < 0.05$, **Table 1**). At the genus level, *Dialister*, *Fusobacterium*, *Parvimonas,* and *Streptococcus* were significantly over-enriched in CRC samples than control samples ($p < 0.05$, **Table 1**). Among these 32 species, 10 were known oral pathogens (17–21). As these species were observed across all three geographic regions, they were not geographic-specific species. The top abundant species associated with CRC samples across all three geographic regions were *Fusobacteria nucleatum* and *Dialister pneumosintes*, which were both known oral pathogens (18–22).

In contrast to these species identified across all three different geographic datasets, some species were only found in one or two of the investigated geographic regions (**Table 1**). For example, eight species were found in CRC patients in Austria and USA but not in CRC patients in Italy (**Table 1**). These species could be region-specific and may be affected by environment, diet, eating style, and ethnic diversity (23).

Eight bacteria species were more abundant in CRC samples

In addition to the different bacterial composition, the prevalence of the same bacteria species differed across the different cohorts. As a result of this observed variability between geographic regions, we combined 11 cohort

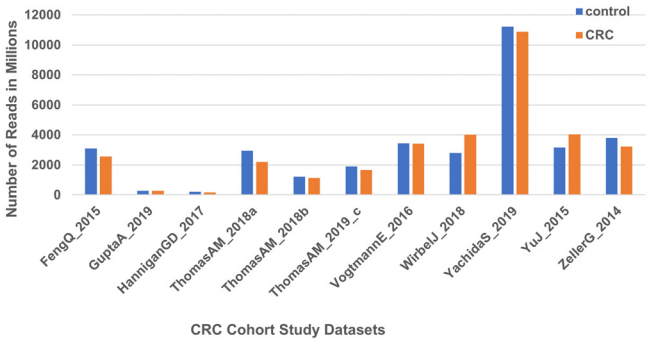

Figure 1: Number of sequencing reads in millions for each cohort study dataset. Number of sequencing reads of DNA in millions for the control (blue) and CRC (orange) group in each cohort study from the public metagenomic database curatedMetagenomicData (15).

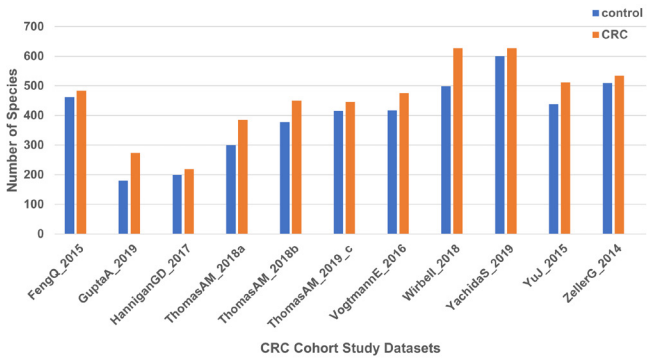

Figure 2: Number of bacteria species in each CRC cohort study dataset. Number of bacteria species identified in stool samples from control (blue) and CRC (orange) samples in each cohort study.

https://doi.org/10.59720/24-084

Phyla	Species	Prevalence		
		Austria	USA	Italy
Actinobacteria	*Slackia exigua*	0.087	0.038	0.082
Bacteriodetes	*Prevotella oris*	0.065	0.019	0.049
	Prevotella nigrescens	0.065	0.038	0.049
	Prevotella intermedia	0.13	0.077	0.066
	Prevotella denticola	0.087	0.038	0.016
	Porphyromonas uenonis	0.24	0.12	0.049
	Porphyromonas gingivalis	0.065	0.019	0
	Phocaeicola coprophilus	0.13	0.038	0.082
	Alloprevotella tannerae	0.087	0.058	0.082
Enterobacteria	*Cronobacter sakazakii*	0.065	0.019	0
Firmicutes	***Streptococcus gallolyticus***	0.087	0.12	0.049
	Selenomonas sputigena	0.065	0.019	0.033
	Peptostreptococcus sp. MV1	0.11	0.038	0
	Parvimonas sp. oral taxon 393	0.087	0.019	0
	Parvimonas sp. KA00067	0.13	0.096	0.049
	Murimonas intestini	0.043	0.038	0
	Leuconostoc carnosum	0.087	0.019	0.016
	Eggerthia catenaformis	0.043	0.019	0
	Dialister pneumosintes	0.22	0.19	0.11
	Dialister micraerophilus	0.043	0.038	0
	Anaerococcus vaginalis	0.2	0.15	0.049
Fusobacteria	***Fusobacterium varium***	0.065	0.019	0.033
Fusobacteria	*Fusobacterium sp. oral taxon 370*	0.087	0.058	0
	Fusobacterium nucleatum	0.33	0.23	0.13
	Fusobacterium necrophorum	0.13	0.019	0.016
	Fusobacterium naviforme	0.13	0.019	0.016
	Fusobacterium equinum	0.065	0.038	0.016
Proteobacteria	***Eikenella corrodens***	0.11	0.038	0.13
	Campylobacter ureolyticus	0.11	0.077	0.033
	Campylobacter gracilis	0.065	0.058	0.033
Synergistota	***Pyramidobacter piscolens***	0.065	0.058	0.098
Terrabacteria	***Schaalia cardiffensis***	0.087	0.038	0.049

Prevalence
>0.1
[0.05, 0.1]
<0.05

Table 1: CRC-associated bacteria species and prevalence across different geographic regions: Austria, USA and Italy. The bacteria composition was compared across datasets from Austria, USA, and Italy to test if it varied across different geographic regions. Among 32 CRC-associated bacteria species, 17 species were common in all 3regions and the remaining species were specific to 1 or 2 regions. The known pathogenic bacteria are in bold (17–19, 21, 32, 35–36, 38–41, 44, 57). The species prevalences are highlighted in different colors based on a prevalence rate greater than 0.1, between 0.05 and 0.1, or lower than 0.05.

datasets across 9 countries to overcome the limitations of using geographic-specific datasets. Using this combined dataset, we performed differential abundance analysis to compare the relative abundance of bacteria species between CRC samples and control samples, and we found a set of bacteria species that were more abundant in CRC samples. *Parvimonas micra*, *Gemella morbillorum*, *Solobacterium moorei*, *Peptostreptococcus stomatis*, *F. nucleatum*, *D. pneumosintes*, *Porphyromonas asaccharolytica,* and *Clostridium symbiosum* were found to be significantly more abundant in CRC samples but still existed in low levels in control samples ($p<0.001$, **Table 2**). The abundance of *C. symbiosum*, *D. pneumosintes*, *P. stomatis*, *G. morbillorum*, *P. micra*, and *F. nucleatum* was about 14– 25 times higher in CRC samples than in control samples (**Table 2**). The abundance of *S. moorei*, and *P. asaccharolytica* was 31 times and 90 times higher in CRC than in control samples, respectively (**Table 2**). Clearly, the abundance of eight species was higher in CRC samples than in control samples. Of these eight species, *F. nucleatum* and *D. pneumosintes* were identified in CRC patients across all geographic regions in the previous study (**Table 1**).

Five bacteria species were more abundant in control samples

In addition to the highly abundant species in the CRC patient samples, we found five species to be significantly more abundant in control samples but at low levels in CRC samples ($p<0.001$, **Table 3**). For example, the abundance of *Bifidobacterium cantenulatum* and *Clostridium sp. CAG:167* was about two times higher in control than in CRC samples (**Table 3**). The abundance of *Roseburia sp. CAG:303* and *Lactobacillus ruminus* were about 4 times and 16 times higher, respectively, and the abundance of *Terrisporobacter othiniensis* was about 612 times higher in control than in CRC samples (**Table 3**). Many of these species have been found to have beneficial effects. *B. catenulatum* helps digest fiber and promote intestinal health (24). *R. sp. CAG:303* is in the genus *Roseburia*, which is anti-inflammatory and known for its protective role in the nervous system (25). Finally, *L. ruminus* is a known beneficial species with important functional roles and metabolic capabilities (19). Taken together, these bacteria usually maintain a beneficial relationship with the host in the gut, and significant pathology can result when they are lost (26).

Species	Relative Abundance*		Fold change*	p-value
	CRC	Control		
Clostridium symbiosum	0.096	0.044	2.185	5.41×10^{-12}
Dialister pneumosintes	0.038	0.003	14.144	1.91×10^{-27}
Peptostreptococcus stomatis	0.059	0.003	18.781	2.51×10^{-35}
Gemella morbillorum	0.037	0.002	19.583	1.04×10^{-37}
Parvimonas micra	0.154	0.006	24.927	1.08×10^{-42}
Fusobacterium nucleatum	0.022	0.001	25.503	7.53×10^{-29}
Solobacterium moorei	0.039	0.001	30.861	6.30×10^{-28}
Porphyromonas asaccharolytica	0.157	0.002	90.126	5.25×10^{-16}

*Number was rounded to the third decimal place.

Table 2: Top eight species with increased abundance in CRC stool samples. The relative abundance of bacteria species in CRC and control stool samples across 11 datasets was determined, and the eight species with increased relative abundance in CRC samples were identified. The fold difference was calculated by dividing the relative abundance of species in CRC by that in control samples. p-value was calculated by a t-test between CRC and control samples using animalcule package (54).

Species	Relative Abundance*		Fold change	p-value
	CRC	Control		
Clostridium sp. CAG:167	0.024	0.042	1.742	4×10^{-3}
Bifidobacterium catenulatum	0.046	0.113	2.476	1×10^{-3}
Roseburia sp. CAG:303	0.013	0.050	3.974	6.29×10^{-6}
Lactobacillus ruminis	0.004	0.060	15.847	3.67×10^{-5}
Terrisporobacter othiniensis	0.000	0.001	612.218	4×10^{-3}

*Number was rounded to the third decimal place.

Table 3: Top five species that had increased abundance in control stool samples. The relative abundance of bacteria species in CRC and control stool samples were determined, and the five species differentially abundant in control samples were identified. The fold difference was calculated by dividing the relative abundance of species in control by that in CRC samples. p-value was calculated by a t-test between CRC and control samples using animalcule package (54).

Potential bacteria biomarkers were identified for CRC samples

Given the high performance of random forest models in biomarker identification, we applied it to create a machine learning model, using the relative abundance profile at taxonomic species level across 1395 samples in all 11 cohort datasets. We then evaluated the top eight bacteria species with higher relative abundance in the CRC samples for their association with CRC to determine their potential to serve as bacterial biomarkers to predict CRC. With the large datasets

https://doi.org/10.59720/24-084

covering different populations and biogeography, the prediction model performed with a high area under the curve (AUC) value of 0.8318, indicating high model performance at predicting CRC based on the levels of these eight bacteria species (**Figure 3**).

We also calculated the feature importance score of each bacteria species in the random forest model, as the importance indicated the strength of the association of each bacteria species with CRC. *P. micra* had an importance of 100%, so it was highly important for the model to predict whether a sample was from a CRC patient or a control (**Figure 4**). *P. micra* has been shown to be highly abundant in CRC, be involved in immunoregulation of CRC, and promote colorectal tumorigenesis (27–31). *G. morbillorum* had the second highest importance, and its pathogenicity has been reported to be involved in immunoregulation of CRC (28–31). *D. pneumosintes, P. stomatis*, and *F. nucleatum* had the third, fourth, and fifth highest importance, respectively, between 50% and 75%. The remaining species all had an importance below 50%. Overall, they were important biomarkers for the model to predict CRC risk from patient samples.

DISCUSSION

Recently, the human gut microbiome has emerged as a relevant factor for human diseases, including CRC (12). We were interested in exploring how gut bacteria may contribute to promoting CRC. We therefore sought to determine whether patients with CRC have differentially abundant gut bacteria species than those without CRC using stool samples.

We investigated the bacteria species across geographic regions to examine which were differentially abundant in CRC samples compared to control samples. Among those species only found in CRC, we identified 15 species found only in CRC samples from a subset of the regions we investigated, specifically Austria, Italy, and the USA. This observation agreed with previous reports that CRC-related bacteria species could be affected by diverse geographical regions (27). The most abundant species in CRC samples across the three regions were *F. nucleatum* and *D. pneumosintes* (18–22). It has previously been demonstrated that *F. nucleatum* is correlated with CRC and that the bacteria modulate the tumor-immune microenvironment and promote colorectal carcinogenesis (22, 32–34). *D. pneumosintes* has also been found in CRC patients with oral infections; the oral species thrived in the CRC-associated bacterial community and drove tumor progression through their metabolic actions, such as the production of acetate and lactate (19, 35, 36).

The difference in bacterial composition and prevalence across the different regional cohorts can likely be explained by the demographic composition, biogeography, and diet of the cohort participants (37). To investigate broadly CRC-associated species, we then combined multiple cohort datasets across various geographic regions to compare the bacterial composition and prevalence in CRC samples and control samples.

Our hypothesis was strongly supported by our analysis results, which showed CRC samples had a set of differentially abundant bacteria species compared to control samples. Notably, the top eight bacteria species that were more abundant in CRC samples have been reported to be correlated with CRC (17, 21-22, 25, 27, 29, 33, 36, 38–40). Among these eight bacteria, *F. nucleatum* has been extensively reported to promote CRC (22, 33, 34, 40). Along with *F. nucleatum*, *P. micra* and *P. stomatis* were reported as oral pathogens, and *G. morbillorum* was suggested as a relevant biomarker for CRC by previous cohorts (31, 41–43). Previous reports have shown that oral pathogens can migrate from the oral cavity to the colon via the circulatory system or gastrointestinal tract, where they can promote colonic inflammation and tumorigenesis (17–19, 27, 44). There has been evidence showing that the pathogenic bacteria likely promote carcinogenesis via various mechanisms such as inflammation, modulating immune response, and producing deleterious metabolic byproducts (10, 27, 45–48). For example, *F. nucleatum* is known to lead to a pro-inflammatory intestinal microenvironment, increase the levels of intestinal short chain fatty acids, and modulate the immune response (49).

In contrast to these CRC-associated bacteria, a set of well-known beneficial bacteria were found to be more abundant in control samples than CRC patient samples. *B.catenulatum*, *T. othiniensis, R. sp. CAG:303*, *L. ruminus*, and *C. sp. CAG:167* were the top five bacteria that were present in increased levels in control samples compared to CRC samples. Bacteria in the *Bifidobacterium* genus are marketed as probiotics due

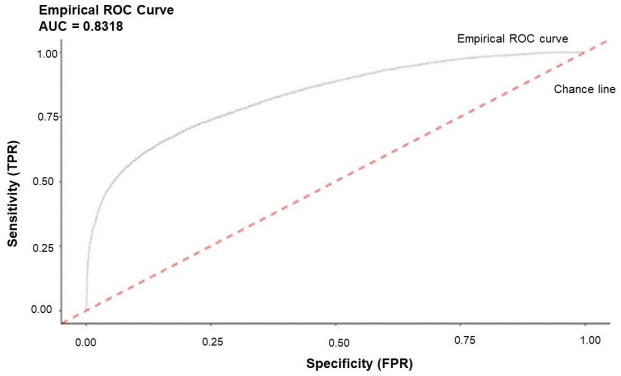

Figure 3: Empirical receiver operating characteristic (ROC) curve from the random forest model predicting CRC based on bacteria species abundance. The ROC curve plotted the sensitivity (TPR for true positive rate) against specificity (FPR for false positive rate). The area under the curve (AUC) value, 0.8318, was generated to show the prediction performance of the identified biomarker.

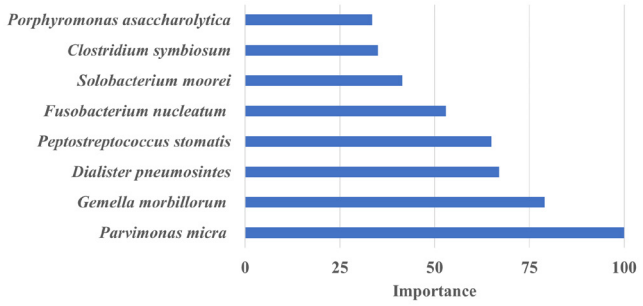

Figure 4: Importance plot of bacterial biomarkers at the species level. The importance of each bacteria species in the random forest model was calculated to predict CRC-associated bacteria species. Three cross validation folds, three cross-validation repeats, and 0.015% top biomarker proportion were used in the random forest model. Higher importance indicates higher association of the bacteria species with CRC.

https://doi.org/10.59720/24-084

to their functions to promote health (50). The *Roseburia* and *Bifidobacterium* genera are known to mediate microbial metabolites, ferment saccharolytic, and produce short-chain fatty acids such as acetate, propionate, and butyrate (51). *L. ruminis* has been identified as an important homofermentative bacteria in the intestinal tract of humans, with functional roles in fermenting cellobiose, galactose, maltose, mannose, raffinose, salicin, and sucrose (52).

Overall, the comparison of bacteria between CRC and control samples strongly supported the correlation between CRC and specific species of gut bacteria. However, our study design and analyses were unable to determine whether any of the bacteria identified in our studies have a causative relationship with CRC. It would be important to perform more experiments in the future to investigate if any of the bacteria species over-represented in the CRC patient stool samples promote the progression of CRC, or if the increased growth of these bacteria species in CRC patients is a consequence of their environment.

Due to their abundance in CRC samples, the gut bacteria we identified may potentially be applied to clinical practice as biomarkers for CRC screening or risk prediction, or as modifiable factors for CRC treatment (53). As a large proportion of CRC incidence and mortality is preventable through CRC screening, CRC bacteria biomarkers may be used to develop an efficient, quick, simple and cost-effective tool for early detection of CRC in addition to current screening approaches. Next, investigating CRC-associated bacteria species may also be beneficial to understand what bacterial composition may promote CRC, further develop more effective medicine to improve personalized treatment, and manage CRC risks. Therefore, in the near future, integrating metagenomics and cancer patient data may enable efficient and powerful prevention and intervention strategies to reduce CRC risk.

MATERIALS AND METHODS

Data collection

The curatedMetagenomicData database provides curated metagenomic data across multiple cohort studies. At the time of this study, the curatedMetagenomicData database included approximately 90 projects of human microbiome sequencing samples from multiple body sites. Since the curatedMetagenomicData database already processed and analyzed the raw metagenomic data with MetaPhlAn3 for bacterial taxonomic abundances, we queried the database, collected preprocessed human bacteria species data from the CRC cohort studies, and performed the downstream statistics analysis.

CRC cohort data were downloaded from the public metagenomic database curatedMetagenomicData. The datasets were queried using selection criteria including datatype “relative_abundance” and study_condition “CRC” from all public datasets in the database, following the database grammatical rules (15). The CRC cohort studies used in this study were as follows: FengQ_2015, GuptaA_2019, HanniganGD_2017, ThomasAM_2018a, ThomasAM_2018b, ThomasAM_2019_c, VogtmannE_2016, Wirbell_2018, YachidaS_2019, YuJ_2015, and ZellerG_2014. Each cohort study included both CRC and control stool sample data. The control samples were collected from stools of individuals without CRC during the CRC cohort studies. In total, 1395 samples with 694 control samples and 701 CRC patient samples were collected from all 11 cohort datasets.

Data analysis

The prevalence of all identified bacteria species was calculated as the fraction of samples with non-zero relative abundance from the total number of samples for both the CRC group and control samples. For example, 371 of 701 CRC samples contained non-zero relative abundance of *C. symbiosum*. Thus, the prevalence of *C. symbiosum* in the CRC group was calculated as 371/701=0.53.

The differential abundance analysis on CRC cohort datasets was performed using animalcules. The analysis was conducted to compare the relative abundance of bacteria species between CRC samples and control samples using the limma R package. Species with differential abundance in CRC with $p<0.01$ were considered significantly correlated with CRC and further characterized. The statistical analysis was first performed for geography-specific datasets to compare CRC-associated bacteria species and prevalence across different geographic regions, with FengQ_2015 for Austria, VogtmannE_2016 for USA, and ThomasAM_2018a and ThomasAM_2018b for Italy. Next, the same statistical approach was applied for the combined datasets across all 11 cohort studies to avoid the bias limited to specific geographic regions.

A random forest model was created using the relative abundance profile at taxonomic species level across the combined 11 cohort datasets to identify top bacteria biomarker species based on the combined cohort datasets, with three-fold cross-validation, three cross-validation repeats, and 0.015% top biomarker proportion. The model was created using animalcules and the AUC value was generated to show how well an identified species predicted CRC. To calculate importance, the sum of feature values in the tree was divided by the total number of trees in the random forest model. The higher the importance, the higher association the bacteria had with CRC.

External validation

For the bacteria species that were differentially abundant in CRC samples compared to control samples, a literature search was performed using PubMed with the bacteria species or genus as the keyword (56). Each species identified in the data analysis was evaluated in a careful literature review to understand its function and potential pathogenicity and validate its correlation with CRC.

ACKNOWLEDGMENTS

We would like to sincerely thank Mrs. Christina Palffy who made this work possible. Her support, encouragement and guidance carried out this research project. Thank Mrs. Palffy for taking the time and energy to mentor this project.

Received: March 17, 2024
Accepted: May 30, 2024
Published: October 21, 2024

https://doi.org/10.59720/24-084

REFERENCES

1. Keum, Na Na and Edward Giovannucci. "Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies" *Nature Reviews Gastroenterology and Hepatology*, vol. 16, 2019, https://doi.org/10.1038/s41575-019-0189-8.
2. Morgan, Eileen, et al., "Global burden of colorectal cancer in 2020 and 2040: Incidence and mortality estimates from GLOBOCAN" Gut, vol. 72, no. 2, 2023, https://doi.org/10.1136/gutjnl-2022-327736.
3. Drost, Jarno, et al., "Sequential cancer mutations in cultured human intestinal stem cells" Nature, vol. 521, no. 7550, 2015, https://doi.org/10.1038/nature14415.
4. Thursby, Elizabeth and Nathalie Juge, "Introduction to the human gut microbiota" *Biochemical Journa*l, vol. 474, no. 11, 2017, https://doi.org/10.1042/BCJ20160510.
5. Bäckhed, Fredrik, et al., "Host-bacterial mutualism in the human intestine" *Science,* vol. 307, no. 5717, 2005, https://doi.org/10.1126/science.1104816.
6. Neish, Andrew S., "Microbes in Gastrointestinal Health and Disease" *Gastroenterology,* vol. 136, no. 1, 2009, https://doi.org/10.1053/j.gastro.2008.10.080.
7. Natividad, Jane M.M., and Elena F. Verdu, "Modulation of intestinal barrier by intestinal microbiota: Pathological and therapeutic implications" *Pharmacological Research,* vol. 69, no.1, 2013, pp. 42-51, https://doi.org/10.1016/j.phrs.2012.10.007.
8. Besten, Gijs Den, et al., "The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism" *Journal of Lipid Research,* vol. 54, no. 9, 2013, pp. 2325-2340, https://doi.org/10.1194/jlr.R036012.
9. Xue, Xiaoyu, et al., "The role of the symbiotic microecosystem in cancer: gut microbiota, metabolome, and host immunome" *Frontiers in Immunology*, vol. 14, 2023, https://doi.org/10.3389/fimmu.2023.1235827.
10. Schwabe, Robert F., and Christian Jobin, "The microbiome and cancer" *Cancer Cell*, vol. 39, no. 10, 2021, pp. 1317-1341, https://doi.org/10.1016/j.ccell.2021.08.006.
11. Schloss, Patrick D., "identifying and overcoming threats to reproducibility, replicability, robustness, and generalizability in microbiome research" *mBio*, vol. 9, no. 3, 2018, https://doi.org/10.1128/mBio.00525-18.
12. Flemer, Burkhardt, et al., "Tumour-associated and non-tumour-associated microbiota in colorectal cancer" *Gut*, vol. 66, no. 4, 2017, https://doi.org/10.1136/gutjnl-2015-309595.
13. Weiss, Robin A., and Peter K. Vogt, "100 years of rous sarcoma virus" *Journal of Experimental Medicine,* vol. 208, no. 12, 2011, pp. 2351-2355, https://doi.org/10.1084/jem.20112160.
14. Wong, Sunny H., and Jun Yu., "Gut microbiota in colorectal cancer: mechanisms of action and clinical applications" *Nature Reviews Gastroenterology and Hepatology,* no. 16, 2019, pp.690-701, https://doi.org/10.1038/s41575-019-0209-8.
15. Pasolli, Edoardo, et al., "Accessible, curated metagenomic data through ExperimentHub" *Nature Methods*, no. 14, 2017, pp. 1023-1024, https://doi.org/10.1038/nmeth.4468.
16. Senghor, Bruno, et al., "Gut microbiota diversity according to dietary habits and geographical provenance" *Human Microbiome Journal,* vol. 7-8, 2018, pp.1-9, https://doi.org/10.1016/j.humic.2018.01.001.
17. Xu, Jun, et al., "Alteration of the abundance of parvimonas micra in the gut along the adenoma-carcinoma sequence" *Oncol Lett*, vol. 20, no. 4, 2020, https://doi.org/10.3892/ol.2020.11967.
18. Abed, Jawad, et al., "Colon Cancer-Associated Fusobacterium nucleatum May Originate From the Oral Cavity and Reach Colon Tumors via the Circulatory System" *Front Cell Infect Microbiol*, vol. 10, 2020, https://doi.org/10.3389/fcimb.2020.00400.
19. Wang, Ni, and Jing Yuan Fang, "Fusobacterium nucleatum, a key pathogenic factor and microbial biomarker for colorectal cancer" *Trends in Microbiology,* vol. 31, no. 2, 2023, pp.159-172, https://doi.org/10.1016/j.tim.2022.08.010.
20. Komiya, Yasuhiko, et al., "Patients with colorectal cancer have identical strains of Fusobacterium nucleatum in their colorectal cancer and oral cavity" *Gut,* 2019, https://doi.org/10.1136/gutjnl-2018-316661.
21. Xuan, Kun, et al., "Is periodontal disease associated with increased risk of colorectal cancer? A meta-analysis" *International Journal of Dental Hygiene*, 2021, https://doi.org/10.1111/idh.12483.
22. Kostic, Aleksandar D., et al., "Fusobacterium nucleatum Potentiates Intestinal Tumorigenesis and Modulates the Tumor-Immune Microenvironment" *Cell Host Microbe*, vol. 14, no. 2, 2013, https://doi.org/10.1016/j.chom.2013.07.007.
23. Borrello, Kirra, et al., "Dietary Intake Mediates Ethnic Differences in Gut Microbial Composition" *Nutrients*, vol. 14, no. 3, 2022, https://doi.org/10.3390/nu14030660.
24. Jensen, N. S., and E. Canale-Parola, "Bacteroides pectinophilus sp. nov. and Bacteroides galacturonicus sp. nov.: Two pectinolytic bacteria from the human intestinal tract" *Appl Environ Microbiol*, vol. 52, no. 4, 1986, https://doi.org/10.1128/aem.52.4.880-887.1986.
25. Keshavarzian, Ali, et al., "Colonic bacterial composition in Parkinson's disease" *Movement Disorders*, vol. 30, no. 10, 2015, https://doi.org/10.1002/mds.26307.
26. Wang, Hui, et al., "Interaction between dietary fiber and bifidobacteria in promoting intestinal health" *Food Chem*, vol. 393, 2022, https://doi.org/10.1016/j.foodchem.2022.133407.
27. Avuthu, Nagavardhini, and Chittibabu Guda, "Meta-Analysis of Altered Gut Microbiota Reveals Microbial and Metabolic Biomarkers for Colorectal Cancer" *Microbiol Spectr*, vol. 10, no. 4, 2022, https://doi.org/10.1128/spectrum.00013-22.
28. Cozza, E, et al., "P22 an unexpected diagnosis; gemella morbillorum, a poorly known killer" *European Heart Journal Supplements*, vol. 25, no. Supplement_D, 2023, https://doi.org/10.1093/eurheartjsupp/suad111.109.
29. Reyes, Ramon, et al., "Gemella morbillorum bacteremia associated with adenocarcinoma of the cecum" *American Journal of Medicine*, 2001, https://doi.org/10.1016/S0002-9343(01)00783-5.
30. Abdulla, Maha Hamadien, et al., "Association of the microbiome with colorectal cancer development (Review)" *Int J Oncol*, vol. 58, no. 5, 2021, https://doi.org/10.3892/IJO.2021.5197.
31. Thomas, Andrew Maltez, et al., "Metagenomic analysis of colorectal cancer datasets identifies cross-cohort

https://doi.org/10.59720/24-084

microbial diagnostic signatures and a link with choline degradation" *Nat Med*, vol. 25, no. 4, 2019, https://doi.org/10.1038/s41591-019-0405-7.
32. Zeller, Georg, et al., "Potential of fecal microbiota for early-stage detection of colorectal cancer" *Mol Syst Biol*, vol. 10, no. 11, 2014, https://doi.org/10.15252/msb.20145645.
33. Castellarin, Mauro, et al., "Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma" *Genome Res*, vol. 22, no. 2, 2012, https://doi.org/10.1101/gr.126516.111.
34. Rubinstein, Mara Roxana, et al., "Fusobacterium nucleatum Promotes Colorectal Carcinogenesis by Modulating E-Cadherin/β-Catenin Signaling via its FadA Adhesin" *Cell Host Microbe*, vol. 14, no. 2, 2013, https://doi.org/10.1016/j.chom.2013.07.012.
35. Jumas-Bilak, Estelle, et al., "Dialister micraerophilus sp. nov. and Dialister propionicifaciens sp. nov., isolated from human clinical samples" *Int J Syst Evol Microbiol*, vol. 55, no. 6, 2005, https://doi.org/10.1099/ijs.0.63715-0.
36. Demirci, Mehmet, "Dialister in microbiome of cancer patients: A systematic review and meta-analysis" *Eurasian J Med Oncol*, vol. 5, no. 3, 2021, https://doi.org/10.14744/ejmo.2021.65073.
37. Kemp, Keri M. et al, "Associations between dietary habits, socio-demographics and gut microbial composition in adolescents" *British Journal of Nutrition*, vol. 131, no. 5, 2024, https://doi.org/10.1017/S0007114523002271.
38. Downes, J., et al., "Dialister invisus sp. nov., isolated from the human oral cavity" *Int J Syst Evol Microbiol*, vol. 53, no. 6, 2003, https://doi.org/10.1099/ijs.0.02640-0.
39. Feng, Qiang, et al., "Gut microbiome development along the colorectal adenoma-carcinoma sequence" *Nat Commun*, vol. 6, 2015, https://doi.org/10.1038/ncomms7528.
40. Kostic, Aleksandar D., et al., "Genomic analysis identifies association of Fusobacterium with colorectal carcinoma" *Genome Res*, vol. 22, no. 2, 2012, https://doi.org/10.1101/gr.126573.111.
41. Yu, Jun, et al., "Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer" *Gut*, vol. 66, no. 1, 2017, https://doi.org/10.1136/gutjnl-2015-309800.
42. Hu, Yi, et al., "Altered Gut Microbiota and Short-Chain Fatty Acids After Vonoprazan-Amoxicillin Dual Therapy for Helicobacter pylori Eradication" *Front Cell Infect Microbiol*, vol. 12, 2022, https://doi.org/10.3389/fcimb.2022.881968.
43. Zhao, Liuyang, et al., "Parvimonas micra promotes colorectal tumorigenesis and is associated with prognosis of colorectal cancer patients" *Oncogene*, vol. 41, no. 36, 2022, https://doi.org/10.1038/s41388-022-02395-7.
44. Mohammadi, Mehrdad, et al., "The role of anaerobic bacteria in the development and prevention of colorectal cancer: A review study" *Anaerobe,* 2022, https://doi.org/10.1016/j.anaerobe.2021.102501.
45. Elliott, Kerryn, and Erik Larsson, "Non-coding driver mutations in human cancer" *Nature Reviews Cancer,* 2021, https://doi.org/10.1038/s41568-021-00371-z.
46. Ternes, Dominik, et al., "Microbiome in Colorectal Cancer: How to Get from Meta-omics to Mechanism?" *Trends in Microbiology*, 2020, https://doi.org/10.1016/j.tim.2020.01.001.
47. Gopalakrishnan, Vancheswaran, et al., "The Influence of the Gut Microbiome on Cancer, Immunity, and Cancer Immunotherapy" *Cancer Cell,* 2018, https://doi.org/10.1016/j.ccell.2018.03.015.
48. McQuade, Jennifer L., et al., "Modulating the microbiome to improve therapeutic response in cancer" *The Lancet Oncology*, 2019, https://doi.org/10.1016/S1470-2045(18)30952-5.
49. Hou, Xinxin, et al., "Effects of gut microbiota on immune responses and immunotherapy in colorectal cancer" *Frontiers in Immunology,* 2022, https://doi.org/10.3389/fimmu.2022.1030745.
50. Fang, Daiqiong, et al., "Bifidobacterium pseudocatenulatum LI09 and Bifidobacterium catenulatum LI10 attenuate D-galactosamine-induced liver injury by modifying the gut microbiota" *Sci Rep*, vol. 7, no. 1, 2017, https://doi.org/10.1038/s41598-017-09395-8.
51. Cheng, Hsin Lin, et al., "The next generation beneficial actions of novel probiotics as potential therapeutic targets and prediction tool for metabolic diseases" *Journal of Food and Drug Analysis*, 2022, https://doi.org/10.38212/2224-6614.3396.
52. Wang, Shuo, et al., "Comparative genomics analysis of Lactobacillus ruminis from different niches" *Genes*, vol. 11, no. 1, 2020, https://doi.org/10.3390/genes11010070.
53. Montalban-Arques, Ana, and Michael Scharl, "Intestinal microbiota and colorectal carcinoma: Implications for pathogenesis, diagnosis, and therapy" *EBioMedicine*, 2019, https://doi.org/10.1016/j.ebiom.2019.09.050.
54. Zhao, Yue, et al., "animalcules: interactive microbiome analytics and visualization in R" *Microbiome*, vol. 9, no. 1, 2021, https://doi.org/10.1186/s40168-021-01013-0.
55. Smyth, Gordon K, "Linear models and empirical bayes methods for assessing differential expression in microarray experiments" *Stat Appl Genet Mol Biol*, vol. 3, no. 1, 2004, https://doi.org/10.2202/1544-6115.1027.
56. Sayers, Eric W., et al., "Database resources of the national center for biotechnology information" *Nucleic Acids Res*, vol. 50, no. D1, 2022, https://doi.org/10.1093/nar/gkab1112.
57. Morio, F.H., et al., "Antimicrobial susceptibilities and clinical sources of Dialister species" *Antimicrob Agents Chemother*, vol. 51, no. 12, 2007, https://doi.org/10.1128/AAC.00538-07.