**Multi-cohort fecal metagenomic analysis reveals the carcinogenic potential of Aspergillus rambellii and the altered fungal signatures in colorectal cancer**

Introduction

Colorectal cancer (CRC) is the third most common cancer and has the second-highest mortality rate after lung cancer globally 1,2. It is estimated that CRC occurrence will increase by approximately 80% to over two million cases in the next two decades3. Interestingly, sporadic CRCs, which arise without known contribution from germline causes or significant family history of CRC, accounted for about 75% of CRCs, implying the importance of environmental factors in CRC pathogenesis4. The gut microbiome provides numerous essential metabolic and physiological functions, including digestion, manufacturing vitamins and training of our immune system2. Recent studies have demonstrated the link between the gut microbiota alteration and CRC5. For instance, a perturbed enteric microbiome was shown to be a significant risk factor for CRC development, in which F. nucleatum promoted glycolysis and oncogenesis of CRC through targeting IncRNA ENO1-IT15. Meta-analysis with approximately 1,000 individuals from five cohorts have revealed the microbial signatures or genes specific for CRC6 and the association between the gut microbiome and choline degradation7.

Despite the fact that >90% of gut microbiome are composed of bacteria, a al composition had also been descrbed to i89. Fungi could influence the 8–12For instance, the mammalian intestinal fungal community interacts with the immune system through the innate immune receptor Dectin-111. The commensal fungi was shown to protect from colitis-associated colon cancer by prompting inflammasome activation and IL-18 maturation in murine model13. Therefore, it is apparent that fungus is playing a more significant role on CRC development than previously anticipatedHowever, e14aiming to discover potential for detection, the exact role of fungi in CRC pathogenesis remains largely unexplored due to their relative low abundance and lack of well-characterized reference fungal genomes.

In this study, we performed a meta-analysis of eight available datasets. After rigorous and consistent data processing, a total of 1,329 samples from four continents, including 525 healthy individuals, 350 adenoma patients, and 454 CRC patients were included in this analysis. Specific fungal diversity and features associated with different stages of CRC were identified. We also explored the intra-fungi and fungi-bacteria co-occurrence patterns in CRC and compared the correlations in CRC, adenoma and healthy control.

Methodology

Study inclusion and data attainment

Faecal shotgun metagenomic data on CRC-related research with a minimal of 2 subject categories (CRC patients and healthy controls) were retrieved from NCBI database and from year 2014 to 2020. We downloaded seven public faecal shotgun CRC datasets from European Nucleotide Archive (ENA) using the following ENA identifiers: ERP005534 for Zeller et al.15, ERP008729 for Feng et al.16, PRJEB10878 for Yu et al.17, PRJEB12449 for Vogtmann et al.18, PRJNA389927 for Hanningan et al.19, PRJEB27928 for Wirbel et al.6, and SRP136711 for Thomas et al.7. And the eighth cohort was downloaded from the DNA Data Bank of Japan (DDBJ) with the Accession numbers: DRA006684, and DRA008156 for Yachida et al.20. All published datasets contained at least two groups, CRC patients and healthy individuals; five published encompassed the adenoma patients7,15,16,19,20 (table 1 and supplementary table 1). Our cohort was generated with the new fecal metagenomic data from samples collected in Hong Kong from 2009 to 2012. A subset of samples from this patient collective was published previously14. These nine studies were organized from eight countries and various sampling procedures, sample storage, and DNA extraction protocols.

Hong Kong study recruitment and sequencing

Recruitment criteria included presentations of digestive symptoms to the outpatient gastroenterology clinics and asymptomatic individuals 50 years or older receiving colonoscopy screening from the Chinese University of Hong Kong Jockey Club Bowel Cancer Education Centre. Stool samples were collected by participants and stored at –20°C within 4 hours. For long-term storage, all samples were stored at –80°C within 24 hours of stool collection. Total DNA was extracted from stool samples by using the QIAamp DNA Stool Mini Kit according to the manufacturer's instructions (Qiagen, Germany). All subjects had intact colonic lesions at the time of stool collection. An independent Chinese cohort of 112 control subjects, 111 patients with CRC, and 197 patients with colorectal adenomas were recruited. Part of the samples had been published in the previous research21.

Sample filter criteria

To ensure consistent and high quality data, samples were subjected to filtering before analysis. Abnormal conditions, history sugery patients, and ambiguous stage patients were discarded. We only included the PCR-free cohort, because the PCR-free kits could reduce bias and cell spike-in controls for more accurate quantification22. Last, we excluded the low-aliganment reads samples (less than 1,000,000), which might cause by the low sequencing depth, host reads contamination, and so on. In the second part, we tended to remove the outlier or suspected contaminated cases, such as the high-fungi (the fungi contained more than 1% in all intestinal microbiome), low-Fungi (the fungi covered less than 0.01% in the gut microbiota), and bacteria or fungi contamination (one species more than 50% of intestinal microbiome) samples. For the last division, the low-fungal sequence depth sample (fungi aligned read counts less than 10,000) would be dropped, which was consistent with previous study exploring that at least 30% of individuals couldn’t be detected fungi in all gastrointestinal segments23.

Sequence pre-processing and taxonomic and functional profiling

We applied the KneadData's default parameters to quality control all the metagenomic samples, which aims to perform principled in silico separation of bacterial reads from these "contaminant" reads, be they from the host or other user-defined sources. In the second step, taxonomic profiles were generated with the Kraken2 v2.0.9-beta across the custom database. Our custom library contained 9,543 bacterial and 909 fungi references from NCBI (https://www.ncbi.nlm.nih.gov/), FungiDB (https://fungidb.org/fungidb/), Ensemble (http://fungi.ensembl.org/index.html), and Broad Institute (https://www.broadinstitute.org/); and was established with the Jellyfish program by counting distinct 31-mer. We discarded all reads of quality less than 20 and shorter than 50 nucleotides, and the other parameters were default. Each query was classified to a taxon with the highest total hits of k-mer matched by pruning the general taxonomic trees affiliated with mapped genomes. The final metagenomic read counts were normalized by multiple methods, rarefied abundance, relative abundance (see supplementary table 2 and supplementary table 8), and median normailzed(see supplementary table 3 and supplementary table 9) with the script (<https://github.com/ifanlyn95/multi-CRC-fungi>). In order to prevent the denominator from being zero, all zero values will be replaced by the normal distribution with a mean value of one-tenth of the non-zero minimum value and one-hundredth of the non-zero minimum value of the variance. The median normalized means dividing the median of the control group of each feature in various cohorts as :

: means the relative abundance of fungi or bacteria in sample , which belongs to cohort . While cohort has exactly sample to sample .

Feature selections criteria

We had three criteria to select the potential candidates, whether it is bacteria or fungi. In the most beginning, we excluded the rarefied candidates with an average abundance of less than 0.1% in all fungi. We selected the same trend features (SSTF), required more than 3/4 cohorts (not less than six cohorts) to perform the same trends. And the log2 of Multiple Median Fold Change (log2MultMedFC) was the evaluation index of SSTF. We define the log2MultMedFC as:

means the counts of CRC/CTRL samples in an individual cohort.

means the fungal names.

means the relative abundance of species in sample .

The second measure was based on the Wilcoxon test. The significance of differential abundance was tested on a per species basis using a Wilcoxon test and adjust the p-value with the conservative Bonferroni correction. And the last strict criterion was Fold Change. We only focused on the absolute value of log2 of features' Fold Change larger than 0.5. In addition, we ignored the unclassified strain of bacteria because we could not explain it. The scripts were also shared on the github (https://github.com/ifanlyn95/multi-CRC-fungi).

The rankdom forest based machine learning approach

Our machine learning analyses exploited the taxonomic species-level median normalized relative abundance by Kraken2 and its plugin, Bracken was an import. To get estimates of test error with lower bias, the LOSO (leave one set out) was used to do nested cross-validation. The features selection and model training was calculated with randomForest package in R. In order to choose the best model, we utilized the largest average AUC and best AUC in multi-features and single-feature as the selected criteria, respectively. The code generating the analyses and the figures is available at https://github.com/ifanlyn95/multi-CRC-fungi.

Association calculation and comparison

Co-occurrence and co-exclusion relationships within fungi and between fungi and bacteria were estimated using the DGCA algorithm24, which is the methodology for systematical assessing the difference in feature-feature regulatory relationships under different conditions. P-values less than 0.05 were considered significant. The inclusion criterion for network plot features had a correlation index less than -0.2 or larger than 0.5.

In the comparison of different stages, DGCA leverages the permutation samples to calculate empirical p-values. Another important index is the z-score, which represents the relative strength of differential correlation. We considered the empirical p-values less than 0.05, and the absolute values of the z-score larger than 5 were a significantly different correlation between different stages. The positive z-score refers to the correlation in CRC is weaker relative strength than the healthy control, while the negative z-score means the correlation is more positive in CRC.

Based upon a threshold for correlation significance (p-value less than 0.05) and the sign of correlation in each condition (i.e., healthy control or CRCs), species-species correlations in each condition could be categorized into 3 classes, i.e. significant positive correlation, no significant correlation, and significant negative correlation. Therefore, there were 9 classes for differential correlation between two stages, namely ‘+/+’, ‘+/0’, ‘+/-’, ‘0/+’, ‘0/0’, ‘0/-’, ‘-/+’, ‘-/0’, and ‘-/-’. It revealed the altered trends between the two stages.

Additional validation experiments on cancer cell line

TBA

Results

Filtering and pre-processing of a large population from various regions for the fungal meta-analysis

We collected shotgun metagenomic sequencing data from eight cohorts: xxx. All the raw sequencing data were reprocessed using the KneadData, Kraken225, and Bracken26 for taxonomic profiling. Each sample has about 107.19 (median) high-quality paired reads that match the bacterial database, and 104.31 (median) paired sequences were aligned to the fungal genome (figure 1a). The median ratio of fungi to bacteria was 10-2.80 (figure 1a), which is in consistent with a previous study27 which reported that fungi make up about 0.1% of the total enteric microbes. Rarefaction curve (figure 1b) showed that all cohort samples reached a plateau at 10,000 sequencing reads. Hence, the minimum rarefied fungal counts of each individual were defined as 10,000 in the downstream analysis. To ensure rigorous outcomes and reduce outlier effect, we applied strict sample filtering criteria (figure 1c). Collectively, we discarded 216 samples with unsatified sequencing quality, 211 outlier or contaminated samples. Additionally, 296 samples (~??% of the filtered samples) with low-fungi sequencing depth were filterd. It is consistent with a previous study23 that fungi could be detected in approximately 70% of individuals in the gut.

Alterations of enteric fungal and bacterial composition in CRC

Considering the overall fungal composition, *Ascomycota* was observed to be the most abundant fungal phylum among all the cohorts, while other dominating fungal phyla showed significant inter-cohort variations. For instance, unlike all other cohorts, the second-most abundant phylum in Yachida's Japanese cohort was *Mucoromycota* but not *Basidiomycota*. Other examples include *Microsporidia* taking up a smaller proportion in Asians compared with non-Asians (figure 2b).

For the altered microbial composition in CRC, that as the groupin the CRC group When we investigated the individual cohort, 6 of the total cohorts showed significant enrichment of *Fusobacteria* (p-value < 0.05) in the CRC group (supplementary figure 2). Although no fungi was identified to show a stronger relationship with CRC than *Fusobacteria*, there were also differentially abundant fungi identified, which will be discussed in later sessions (figure 2c).

In agreement with a previous research showing distorted microbiome diversity in the diseased group28, we observed that alpha diversity were reduced in CRC patients compared to healthy individuals when considering all the cohorts together (figure 2d). When considering individual cohort, majority showed reduced alpha diversity (chao1 index). Despite the less apparent alteration of fungal composition as compared to the bacterial composition in CRC as well as the heterogeneity in different cohorts, we could still observe remarkable differences in both fungi phyla composition and alpha diversity in the CRC group which are not negligible.

Identification of fungal species associated with CRC by univariate meta-analysis

We next searched for the potential enteric fungal shifts in CRC patients as compared to heathy individuals. After filtering low abundant (< 0.1%) fungi from the 592 aligned species, 296 species were obtained for further analysis (figure 3a and supplementary table 2, 3). Using the Wilcoxon rank-sum test to compare data from all the cohorts together, 74 differentially abundant fungi were identified which was named as the main set (FDR < 0.1). Among the 74 identified species, we further shortlisted 33 species which demonstrated significant alterations (FDR < 0.01) as the core set (figure 3a and supplementary table 4).

We then evaluated if these 74 fungi (main set) were consistently altered across all the 8 cohorts using SSTF and non-parametric tests (see Methods). We observed that the enrichment and depletion status of the 74 species were consistent in most of the cohorts except the 2019\_Thomas and 2019\_Yachida cohorts. Interestly, most of the 74 species in the 2019\_Thomas cohorts either showed significant enrichment in CRC patients or no significant difference between CRC versus healthy individuals but very few showed deleption in CRC patients. Whereas in the 2019\_Yachida group, most of the identified 74 fungi showed weak variance in CRC patients versus heathy individuals which was unlike other cohorts. We also discovered that 3 of the 74 species showed consistent changes across all the cohorts with *Aspergillus* *rambellii* and *Erysiphe* *pulchra* being enriched while *Trichophyton* *mentagrophytes* being depleted in CRC (figure 3b and supplementary table 6). We further identified 15 species which were consistently altered in 7 out of the 8 cohorts. 10 of them were enriched in CRC patietns while the remaining 5 were depleted (see supplementary table 6). Notably, only *Aspergillus* *rambellii* showed a significant difference (p-value < 0.05) in all the cohorts, except the 2019\_ThomasA cohort (figure 3d and supplementary table 7).

For the 33 species in the core set, 10of them patients and the remaining23 were depleted Notably, we identified that *Aspergillus* *rambellii* showed the most remarkable difference between the CRC patients and the healthy control groups (-log10FDR = 17.29). We observed that alteration of these 33 species in CRC patients versus healthy individuals were relatively consistent in most of the cohorts except 2019\_Thomas and 2019\_Yachida. Moreover, most of the species showed consistent alterations in at least 3 cohorts.

In the meantime, we also compared the fungal community between CRC and adenoma patients (supplementary table 5). 6 of the identified spcies were significantly differed (FDR < 0.01) in both CRC patients versus adenoma patients and CRC patients versus healthy individuals. These species include *Aspergillus* *rambellii*, *Erysiphe* *pulchra*, *Thielaviopsis* *punctulata*, *Moniliophthora* *perniciosa*, *Sphaerulina* *musiva*, *Aspergillus* *ochraceoroseus*. All of these species belong to the *Ascomycota* phylum except *Moniliophthora* *perniciosa* which belongs to the *Basidiomycota* phylum (supplementary table 12).

In summary, we identified 74 differentially abundant fungi between CRC patients and heathy individuals, of which 33 significant species were further selected as core-set for the downstream analysis. Despite the presence of cohort heterogeneity, we could still identified fungal features which were consistently altered in most of the cohorts.

Identification of fungi with most significant association with CRC

To identify the most crucial candidates associated with CRC, we utilized the stricter criteria (see methods) and found that *A.* *rambellii* and *A.* *kawachii* were the only two significant fungi (figure 3c). *A. rambellii* was the only candidate with a significant difference among seven cohorts, excluding the 2019\_Thomas cohort (figure 3d). Whereas *A. kawachii* was significantly different among 2014\_ZellerG, 2016\_VogtmannE, 2017\_JunY, and our unpublished dataset (figure 3d). Although they belong to the same genus, they were shown to have different association with CRC with *A. rambellii* being enriched in CRC, while *A. kawachii* being opposite. Importantly, they were also reported to have opposing actions in previous studies. *A. rambellii* has been demonstrated to have to ability of accumulating aflatoxins (AF) and the aflatoxin precursor sterigmatocystin (ST)29. AF and ST are well known as the most carcinogenic natural products 30. In contrast, *A. kawachii* was reported with the ability of enhancing anticancer effects of cancer-curing herbs, such as Korean mistletoe31 and fermented silkworm larvae32. All these previous literatures supported our findings. Collectively, our meta-analysis revealed the key fungi, *A.* *rambellii*, and *A. kawachii*, that were significantly associated with CRC across multiple metagenomic studies.

The fungal-bacterial combination performed best in CRC diagnosis

To identify the potential value of gut fungi in clinical, we trained the model with single or multiple features to diagnose CRC. Among the single-feature models, the average AUC value of six features was greater than 60%. They were composed of four bacteria (*Fusobacterium nucleatum*, *Parvimonas micra*, *Gemella morbillorum*, and *Porphyromonas asaccharolytica*) and two fungi (*Aspergillus rambellii*, and *Aspergillus kawachii*) (table 2).

Correlation analysis of CRC associated fungi

Due to the complex and multifactorial nature of CRC, we asked whether the interactions among fungi were associated with CRC. We performed the correlation analysis with DGCA24 on the 33 differentially abundant fungi in the core set and observed that the correlations within the core-set network were stronger in CRC than healthy control (figure 4). There were only four strong positive (correlation index ≥ 0.5) and three negative interactions (correlation index ≤ -0.15) in healthy individuals (figure 4a); meanwhile, there were nine strong positive and four negative associations exhibited in CRC patients (figure 4b). In addition, nine strong positives and one negative interrelationship were shown in adenoma patients. *Aspergillus* *rambellii*, *Rhizophagus* *irregularis*, *Rhizophagus* *clarus*, *Phytopythium* *vexans*, and *Edhazardia* *aedis* showed significant correlations with one another in all three stages (figure 4a and supplementary figure 5). However, their correlation was not consistent throughout the stages. A more exciting discovery was that 4 of the CRC-associated fungi *Aspergillus* *rambellii*, *Eysiphe* *pulchra*, *Thielaviopsis* *punctulata*, and *Sphaerulina* *musiva* were found to be positively correlated with one another in CRC patients. However, these correlations disappeared in the adenoma or healthy control group (figure 4a and supplementary figure 5). This suggested that these fungi may cooperate to promote CRC carcinogenesis (figure 4a). In adenoma or healthy control group, *T. punctulata*, *S. musiva*, and *E. pulchra* were not strongly correlated with any other candidates, but *A. rambellii* were still positively correlated with *Moniliophthora* *Pemiciosa* in these two stages. Thus, our analysis revealed that co-occurrence fungal relationships might be crucial for enteric homeostasis in a healthy gut. In contrast, fungal dysbiosis might break the balance and provide a suitable environment for the harmful fungi clustering developing, which might cause colorectal carcinogenesis.

Correlation between CRC-associated bacteria and CRC-associated fungi in CRC

To identify the significant differentially abundant bacteria between CRC and healthy individuals, we used Wilcoxon rank-sum test with a stricter selection criteria (q-value < 0.01, , and removed the unclassified species) (supplementary table 10). We identified 31 differentially abundant bacteria in CRC which was more significant as compared to those in fungi (supplementary table 10). At least half of the bacteria identified have been reported to be CRC-related33–44 or probiotics45–50, including well-known CRC-associated pathogens, such as *Fusobacterium* *nucleatum*, *Parvimonas* micra, and *Gemella* *morbillorum*; as well as famous probiotics, such as, *Roseburia* *intestinalis*, *Bifidobacterium* *bifidum*, and *Streptococcus* *thermophilus*. The consistent result in differentially abundant bacteria analysis implied that our methods used in previous differentially abundant fungi analysis were credible. Next, we asked whether the correlations between the fungi and bacteria are associated with CRC. We performed DGCA24, to calculate the correlation between fungi and bacteria. We discovered that the fungal-bacterial corrlations were weaker than inter-fungi correlations across all three groups (CRC, Adenoma and Healthy control) (supplementary table 11). However, the fungal-bacterial correlation was still stronger in CRC when compared to healthy control, which is the same as inter-fungi correlations (figure 4 and supplementary table 11).

Alternative enteric microbiome in CRC, and fungi were the important potential biomarkers or inducement for CRC

Our results revealed a marked difference in correlations among fungi and bacteria in CRC and healthy (figure 5a). Correlations among fungi were higher in healthy controls compared to CRC (figure 5a). In contrast, intra-bacterial relationships were increased in CRC (figure 5a). When assessing fungi-bacteria correlations, two peaks at -5 and 5 were observed, indicating the fungi-bacteria correlations didn't exist in the gentle relationship. Collectively, our results implicated those correlations among fungi were weakened in CRC, while bacteria-bacteria correlations were utterly opposite. And associations of fungi-bacteria existed above two situations.

We also defined the nine cases in the pair correlation comparison (figure 5b and see methods). Our results showed that the most significant correlations were '+/+', '+/0', and '0/+', indicating that the most meaningful comparisons (pmVal < 0.05) were based on the positive correlations; in other words, negative correlations were rare (figure 5c). Notably, only the internal fungi had six '-/+' cases, which means the feature pair correlation in CRC was negative, while its association in healthy control was positive (figure 5c). It might reveal some potential markers or changes in the stage alteration.

Sixty-four candidates (31 bacteria and 33 fungi) were separated into six clusterings with affinity propagation cluster (figure 5d). Among these, two clusterings contained most of the candidates were identified. We named the biggest one as the Fun\_cluster because 18 of 22 participants were fungi. In this cluster, the correlations between the enriched fungi were enhanced in CRC compared. We identified that *E. pulchra* and *A. rambellii* were the core fungi in the eEuk\_cluster. Three CRC enriched bacteria, including *F. nucleatum*, *F. periodonticum*, and *P. micra* had strong correlations with these fungi (figure 5d). The most outstanding fungi, *A. rambellii*, and the cancer-related pathogens, *F. nucleatum*, were in the same cluster. Its z-score was -5.95, and it belonged to the '+/+' case (see supplementary table 11), indicating that fungi *A. rambellii* and bacteria *F. nucleatum* were both positive relative CRC and healthy control. Still, their pair association was more potent in CRC compared with healthy control. Also, *A. rambellii* showed a strong correlation with another CRC-enriched bacteria *P. micra*, with a z-score -5.07, belonged' +/0' (see supplementary table 11). In contrast, no direct correlation between another key fungi *E. pulchra,* and these three CRC-enriched bacteria was identified. It revealed that *A. rambellii* was an important potential carcinogen and related to other reported CRC-pathogenes.

The second clustering included twenty-one candidates, and 17 of 21 were bacteria, namely Bac\_cluster. It had the most reported probiotics or potential probiotics, including *S. thermophilus*, *S. salivarius*, *A. hadrus*, and *E. eligens* (figure 5d). While some cancer enriched such as *D. pneumosintes*, *S. anginosus*, and *P. intermedia* were also included in this clustering (figure 5d). And these results were consistent with many previous studies reporting the correlations among bacteria. Interestingly, *P. kudriavzevii* was the only one correlated with three of four probiotics. In previous research51, colon cancer cell proliferation, induction of endogenous and exogenous apoptosis inhibition would be inhibited by metabolites from *P. kudriavzevii*. Collectively, these findings revealed that the correlations among intra-fungi, fungi-bacteria, and bacteria-bacteria were quite different in CRC. The specific relationships between *A. rambellii* with CRC-carcinogen, and *P. kudriavzevii* with CRC-probiotics revealed that fungi were the important potential biomarkers or inducement for CRC.

*Aspergillus rambellii* and its conditioned medium promote the viability of colon cancer cells.

TBA

Discussion

Fungi are one of the crucial inhabitants of the human gastrointestinal tract52. In the previous, most of the research would like to force on the relationship between gut bacteria and the host. Fungi were ignored in the microbiota studies for an extended period because of their low proportion in the enteric microbiome53. During further research, some studies9,11,12,54 revealed that fungi also played an essential role in the associations with the host. Our study was the first to report the enteric fungi altered in CRC across multiple cohorts according to the acknowledge we know.

According to the alpha diversity rarefaction curve, our results indicated that the fungal plateau was 10,000. The previous study revealed that beyond 60 million and 80 million reads per sample were required to recover all bacterial classification55 and the full richness of different Antimicrobial resistance (AMR) gene families56 in the fecal sample. Our research could suggest that future studies on enteric fungi should be conducted at a sequencing depth of at least 10 thousand fungi in each case. We have supplemented the gaps in sequencing depth for fungi and provided reliable support for follow-up fungal research based on the meta-analysis of more than 2,000 samples.

Previous studies have revealed that genetic background, age, dietary habits, lifestyle, and local environments play crucial roles in the heterogeneity of enteric microbiota among humans57. We observed a significant p-value for the fungal composition difference in different cohorts from the principal component analysis results. This observation is consistent with previous bacterial research58,59 exposing the effects of ethnicity and technical various on gut microbiota and highlights the compulsion for combined analysis. We showed that the enteric fungal profile in CRC is distinct from healthy controls, and it owned significant regional differences. Ascomycota and Basidiomycota dominated both CRC and control subjects in almost all cohorts. The alpha diversity in CRC was altered and decreased compared with healthy controls. It was similar to other intestinal diseases IBD in the previous research8. In the IBD research, *Saccharomyces cerevisiae* and *Candida albicans* were enriched in IBD, which may be the primary reason for low alpha diversity in IBD. However, this study showed they have no apparent difference between CRC and healthy control. This revealed that even though the alpha diversity of the various disease would be disturbed, their mechanisms may be different.

The meta-analysis approach has been used to evaluate and combine results of comparable studies60 with significant advantages of reducing the influence by most highly abundant features and weakening the batch effect. We applied the rank-sum method to identify 33 fungi and 31 bacteria that significantly differed in CRC across eight cohorts. Since few studies described the fungi associated with CRC, the bacterial populations we screened could prove that the fungal organisms we found were correct. Our results were supported by the fact that more than half of the cancer-related bacteria were carcinoma potential probiotics or pathogens, and four out of the top five were reported colorectal cancer pathogen, namely, F. nucleatum61, P. micra17, G. morbillorum62,63, and A. hadrus64 have been reported previously to promote CRC development and occurrence. It also proved that there might be existed several potential probiotics or pathogenic fungi we found, and the possibility of the top rankings p-value difference or the biggest value of would be more significant. Among the most effective two fungi, *A. rambellii* and *R. irregularis* were reported cancer related. *A. rambellii* was the most significantly different fungi in CRC-control, as well as in CRC-adenoma comparison. It would create carcinogenic products, aflatoxin and aflatoxin precursor sterigmatocystin29,30. Salvianolic acids, as a result of the *A. officinalis*-*R. irregularis* symbiosis, showed a significant impact on cancer treatment65,66. Furthermore, other major affected compounds from *A. officinalis*-*R. irregularis* symbiosis, such as rosmarinic acid, ferulic acid, caffeic acids, and derivatives, present several health-related properties, such as antioxidant, anti-inflammatory, and antimicrobial activities67,68. *A. rambellii* and *A. kawachii* were the most enriched or depleted fungus in CRC. Even though both were from the same genus *Aspergillus*, the latter plays the opposite function in cancer. The crude enzyme extract derived from *A. rambellii* could enhance the antioxidative activities of Viscum album var. coloratum31 (Korean mistletoe; KM), a promising agent for treating colon cancer69, hepatoma, and immunomodulation in the previous study. In the other research, solid-state fermentation with *A. kawachii* would create the fermented silkworm was investigated anticancer activity in human hepatocellular carcinoma cells32. Collectively, our results evidenced the correctness of fungi selection, and the fungi were associated with CRC or adenoma, especially *A. rambellii*, *R. irregularis*, and *A. kawachii*.

As we know, the human gut microbiome was a large, complicated, and mutually microbial community. Most previous studies have focused on one key microorganism or one metabolite effect on CRC development70–72. However, diseased intestinal microbiota dysbiosis may be associated with a list of the microbiota, the microbial community perturbed, instead of only one or several microorganisms altered28,73. Except for the transformed fungal abundance, alteration in fungal internal relationship and fungi-bacteria association could partially explain colorectal tumorigenesis. The fungal internal relationship in CRC was quite different from the healthy control and adenoma. Multiple strong positive or negative correlations in CRC were disappeared or weaken in the compared stages. Its establishment may potentially contribute to colorectal carcinogenesis. The strongest relationship clustering in these three communities was also quite different, but *Aspergillus* *rambellii*, *Rhizophagus* *irregularis*, *Rhizophagus* *clarus*, *Phytopythium* *vexans*, and *Edhazardia* *aedis* appeared in all stages. It revealed that they might play a vital role in the stability of the entire intestinal ecology.

Although we recognized that their relationships between stages were quite different, we demanded to perceive which were statistically different. We defined the z-score and correlation classes to represent the relative strength and the trend of differential correlation in CRC and the healthy control, respectively. Our results showed that the fungal internal correlations were interrupted in CRC, while bacterial were enhance. The disruption from fungal interrelationship may break the healthy intestinal environment and induce colorectal carcinogenesis. On the other hand, the new increased bacterial correlations in CRC may potentially contribute to colorectal carcinogenesis. We also observed some interesting comparisons of the relationship between fungi and bacteria. Our results showed that the less changed differential fungal-bacterial correlations contained low proportions. The CRC strengthen and weaken correlations performed the primary and secondary ratios in fungal-bacterial correlation comparisons, respectively. It revealed that the internal-kingdom associations and external-kingdom correlations were quite different. This suggested that bacterial kingdom dysbiosis may cause the fungi to tremble rapidly, which was not similar to the warm alteration of internal-kingdom relationships as previously described14,74.

In the present study, the candidates were automatically divided into six clusters. The two main clusters, named Fun\_Cluster and Bac\_Cluster, had the highest proportions of bacteria and fungi. All reported potential probiotics, excluded *R. intestinalis*, were separated in the Bac\_Cluster. We disclosed that *P. kudriavzevii* owned multiple correlations with these probiotics. And itssecretion metabolites exert anticancer effects by inhibiting cell proliferation and inducing intrinsic and extrinsic apoptosis in colon cancer cells51. Collectively, this discovery exposed that may exist other potential probiotics in this cluster. Another interesting finding, there were strong correlations among *A. rambellii*, *F. nucleatum*33,61,75, and *P. micra*17, and the latter two were the famous CRC-related pathogens. It may indicate colorectal carcinogenesis under their synergistic effect. Conclusively, we showed the probiotic group (*P. kudriavzevii*, *S. thermophilus*, *A. hadrus*, and *S. salivarius*) and the pathogenic bunch (*A. rambellii*, *F. nucleatum*, and *P. micra*), and the potentially pathogenic and probiotic candidates among the different clusters.

TBA (in vitro)

In conclusion, our study elucidated the following findings through the eight cohorts with more than 1,300 cases. We indicated the fungal internal network and fungal-bacterial relationship alterations in CRC, indicating that synergistic intra-fungi and micro-fungi-bacteria interaction might contribute to colorectal carcinogenesis. Several fungi, *R. rambellii*, *R. kawachii*, and *P. kudriavzevii,* play critical roles in promoting or inhibited CRC. [TBA].

Contribution

**Yufeng Lin**, study design, dry lab analysis, wet lab validation, and written.

**Thomas Kwong**, written and web lab technology guidance.

**Yali Liu**, wet lab validation.

**Nick Ding**, written.

**Jason Kang**, web lab technology guidance

**Yiwei Wang**, dry lab guidance (SBS, cuhk).

**Sunny Wong**, clinical guidance.

**Han Jing**, fungal database establishment (Med-X, XJTU).

**Harry Chan**, written guidance.

**Changan Liu**, dry lab guidance.

**Weixin Liu**, dry lab guidance.

**SK**, dry lab guidance.

**Yanqiang Ding**, dry lab guidance.

**Bisi**, dry lab guidance.

Reference

1. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* **136**, E359–E386 (2015).

2. Lin, Y., Wang, G., Yu, J. & Sung, J. J. Y. Artificial intelligence and metagenomics in intestinal diseases. *Journal of Gastroenterology and Hepatology* **36**, 841–847 (2021).

3. Siegel, R., DeSantis, C. & Jemal, A. Colorectal cancer statistics, 2014. *CA: A Cancer Journal for Clinicians* **64**, 104–117 (2014).

4. Yamagishi, H., Kuroda, H., Imai, Y. & Hiraishi, H. Molecular pathogenesis of sporadic colorectal cancers. *Chin J Cancer* **35**, 4 (2016).

5. Hong, J. *et al.* *F. nucleatum* targets lncRNA ENO1-IT1 to promote glycolysis and oncogenesis in colorectal cancer. *Gut* gutjnl-2020-322780 (2020) doi:10.1136/gutjnl-2020-322780.

6. Wirbel, J. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nature Medicine* **25**, 27 (2019).

7. Thomas, A. M. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nature Medicine* **25**, 27 (2019).

8. Botschuijver, S. Intestinal Fungal Dysbiosis Is Associated With Visceral Hypersensitivity in Patients With Irritable Bowel Syndrome and Rats. **153**, 14 (2017).

9. Bajaj, J. S. *et al.* Fungal dysbiosis in cirrhosis. *Gut* **67**, 1146–1154 (2018).

10. Iliev, I. D. & Leonardi, I. Fungal dysbiosis: immunity and interactions at mucosal barriers. *Nat Rev Immunol* **17**, 635–646 (2017).

11. Iliev, I. D. *et al.* Interactions Between Commensal Fungi and the C-Type Lectin Receptor Dectin-1 Influence Colitis. *Science* (2012).

12. Wheeler, M. L. *et al.* Immunological Consequences of Intestinal Fungal Dysbiosis. *Cell Host & Microbe* **19**, 865–873 (2016).

13. Malik, A. *et al.* SYK-CARD9 Signaling Axis Promotes Gut Fungi-Mediated Inflammasome Activation to Restrict Colitis and Colon Cancer. *Immunity* **49**, 515-530.e5 (2018).

14. Coker, O. O. *et al.* Enteric fungal microbiota dysbiosis and ecological alterations in colorectal cancer. *Gut* **68**, 654–662 (2019).

15. Zeller, G. *et al.* Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol* **10**, 766 (2014).

16. Feng, Q. *et al.* Gut microbiome development along the colorectal adenoma–carcinoma sequence. *Nature Communications* **6**, 6528 (2015).

17. Yu, J. *et al.* Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* **66**, 70–78 (2017).

18. Vogtmann, E. *et al.* Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing. *PLoS ONE* **11**, e0155362 (2016).

19. Hannigan, G. D., Duhaime, M. B., Ruffin, M. T., Koumpouras, C. C. & Schloss, P. D. Diagnostic Potential and Interactive Dynamics of the Colorectal Cancer Virome. *mBio* **9**, (2018).

20. Yachida, S. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nature Medicine* **25**, 27 (2019).

21. Nakatsu, G. *et al.* Alterations in Enteric Virome Are Associated With Colorectal Cancer and Survival Outcomes. *Gastroenterology* **155**, 529-541.e5 (2018).

22. Jones, M. B. *et al.* Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proc Natl Acad Sci USA* **112**, 14024–14029 (2015).

23. Schulze, J. & Sonnenborn, U. Yeasts in the Gut: From Commensals to Infectious Agents. *Dtsch Arztebl Int* **106**, 837–842 (2009).

24. McKenzie, A. T., Katsyv, I., Song, W.-M., Wang, M. & Zhang, B. DGCA: A comprehensive R package for Differential Gene Correlation Analysis. *BMC Systems Biology* **10**, 106 (2016).

25. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biology* **20**, 257 (2019).

26. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **3**, e104 (2017).

27. Chin, V. K. *et al.* Mycobiome in the Gut: A Multiperspective Review. *Mediators of Inflammation* **2020**, e9560684 (2020).

28. Nakatsu, G. *et al.* Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nature Communications* **6**, 8727 (2015).

29. Cary, J. W., Ehrlich, K. C., Beltz, S. B., Harris-Coward, P. & Klich, M. A. Characterization of the Aspergillus ochraceoroseus aflatoxin/sterigmatocystin biosynthetic gene cluster. *Mycologia* **101**, 352–362 (2009).

30. Frisvad, J. C., Skouboe, P. & Samson, R. A. Taxonomic comparison of three different groups of aflatoxin producers and a new efficient producer of aflatoxin B1, sterigmatocystin and 3-O-methylsterigmatocystin, Aspergillus rambellii sp. nov. *Systematic and Applied Microbiology* **28**, 442–453 (2005).

31. Kim, S.-Y., Yang, E.-J., Son, Y. K., Yeo, J.-H. & Song, K.-S. Enhanced anti-oxidative effect of fermented Korean mistletoe is originated from an increase in the contents of caffeic acid and lyoniresinol. *Food Funct.* **7**, 2270–2277 (2016).

32. Cho, H.-D. *et al.* Solid state fermentation process with Aspergillus kawachii enhances the cancer-suppressive potential of silkworm larva in hepatocellular carcinoma cells. *BMC Complement Altern Med* **19**, 241 (2019).

33. Parhi, L. *et al.* Breast cancer colonization by Fusobacterium nucleatum accelerates tumor growth and metastatic progression. *Nat Commun* **11**, 3259 (2020).

34. Stott, K. J., Phillips, B., Parry, L. & May, S. Recent advancements in the exploitation of the gut microbiome in the diagnosis and treatment of colorectal cancer. *Biosci Rep* **41**, BSR20204113 (2021).

35. Reyes, R., Abay, A. & Siegel, M. Gemella morbillorum bacteremia associated with adenocarcinoma of the cecum. *The American Journal of Medicine* **111**, 164–165 (2001).

36. Ai, D. *et al.* Identifying Gut Microbiota Associated With Colorectal Cancer Using a Zero-Inflated Lognormal Model. *Front. Microbiol.* **10**, 826 (2019).

37. Loftus, M., Hassouneh, S. A.-D. & Yooseph, S. Bacterial community structure alterations within the colorectal cancer gut microbiome. *BMC Microbiol* **21**, 98 (2021).

38. Mu, W. *et al.* Intracellular Porphyromonas gingivalis Promotes the Proliferation of Colorectal Cancer Cells via the MAPK/ERK Signaling Pathway. *Front. Cell. Infect. Microbiol.* **10**, 584798 (2020).

39. Zhou, Y. & Luo, G.-H. Porphyromonas gingivalis and digestive system cancers. *WJCC* **7**, 819–829 (2019).

40. Yang, C.-Y. *et al.* Oral Microbiota Community Dynamics Associated With Oral Squamous Cell Carcinoma Staging. *Front. Microbiol.* **9**, 862 (2018).

41. Masood, U., Sharma, A., Lowe, D., Khan, R. & Manocha, D. Colorectal Cancer Associated with Streptococcus anginosus Bacteremia and Liver Abscesses. *Case Rep Gastroenterol* **10**, 769–774 (2016).

42. Suzuki, H., Hase, R., Otsuka, Y. & Hosokawa, N. Bloodstream infections caused by Streptococcus anginosus group bacteria: A retrospective analysis of 78 cases at a Japanese tertiary hospital. *Journal of Infection and Chemotherapy* **22**, 456–460 (2016).

43. Zupancic, K., Kriksic, V., Kovacevic, I. & Kovacevic, D. Influence of Oral Probiotic Streptococcus salivarius K12 on Ear and Oral Cavity Health in Humans: Systematic Review. *Probiotics & Antimicro. Prot.* **9**, 102–110 (2017).

44. Ternes, D. *et al.* Microbiome in Colorectal Cancer: How to Get from Meta-omics to Mechanism? *Trends in Microbiology* **28**, 401–423 (2020).

45. Liang, Q. *et al.* Fecal Bacteria Act as Novel Biomarkers for Noninvasive Diagnosis of Colorectal Cancer. *Clin Cancer Res* **23**, 2061–2070 (2017).

46. Jia, W., Rajani, C., Xu, H. & Zheng, X. Gut microbiota alterations are distinct for primary colorectal cancer and hepatocellular carcinoma. *Protein Cell* **12**, 374–393 (2021).

47. Yu, X. *et al.* A Comparative Characterization of Different Host-sourced Lactobacillus ruminis Strains and Their Adhesive, Inhibitory, and Immunomodulating Functions. *Front. Microbiol.* **8**, (2017).

48. Department of Microbiology, Islamic Azad University School of Science, Fars, Iran *et al.* Anti-colon cancer activity of Bifidobacterium metabolites on colon cancer cell line SW742. *Turk J Gastroenterol* **30**, 835–842 (2019).

49. Tarrah, A. *et al.* In vitro Probiotic Potential and Anti-cancer Activity of Newly Isolated Folate-Producing Streptococcus thermophilus Strains. *Front. Microbiol.* **9**, 2214 (2018).

50. Singh, J. Bifidobacterium longum, a lactic acid-producing intestinal bacterium inhibits colon cancer and modulates the intermediate biomarkers of colon carcinogenesis. *Carcinogenesis* **18**, 833–841 (1997).

51. Saber, A., Alipour, B., Faghfoori, Z., Mousavi jam, A. & Yari Khosroushahi, A. Secretion metabolites of probiotic yeast, Pichia kudriavzevii AS-12, induces apoptosis pathways in human colorectal cancer cell lines. *Nutrition Research* **41**, 36–46 (2017).

52. Pérez, J. C. Fungi of the human gut microbiota: Roles and significance. *International Journal of Medical Microbiology* **311**, 151490 (2021).

53. Pérez, J. C. & Johnson, A. D. Regulatory Circuits That Enable Proliferation of the Fungus Candida albicans in a Mammalian Host. *PLOS Pathogens* **9**, e1003780 (2013).

54. Aykut, B. *et al.* The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL. *Nature* **574**, 264–267 (2019).

55. Rajan, S. K., Lindqvist, M., Brummer, R. J., Schoultz, I. & Repsilber, D. Phylogenetic microbiota profiling in fecal samples depends on combination of sequencing depth and choice of NGS analysis method. *PLoS ONE* **14**, e0222171 (2019).

56. Gweon, H. S. *et al.* The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples. *Environmental Microbiome* **14**, 7 (2019).

57. Fontana *et al.* Gut Microbiota Profiles Differ among Individuals Depending on Their Region of Origin: An Italian Pilot Study. *IJERPH* **16**, 4065 (2019).

58. Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A. & Alm, E. J. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun* **8**, 1784 (2017).

59. Chong, C. W. *et al.* Effect of ethnicity and socioeconomic variation to the gut microbiota composition among pre-adolescent in Malaysia. *Sci Rep* **5**, 13338 (2015).

60. Pereira, M. B., Wallroth, M., Jonsson, V. & Kristiansson, E. Comparison of normalization methods for the analysis of metagenomic gene abundance data. *BMC Genomics* **19**, 274 (2018).

61. Yu, T. *et al.* Fusobacterium nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell* **170**, 548-563.e16 (2017).

62. Kwong, T. N. Y. *et al.* Association Between Bacteremia From Specific Microbes and Subsequent Diagnosis of Colorectal Cancer. *Gastroenterology* **155**, 383-390.e8 (2018).

63. Lopez-Dupla, M., Creus, C., Navarro, O. & Raga, X. Association of Gemella morbillorum Endocarditis with Adenomatous Polyps and Carcinoma of the Colon: Case Report and Review. *Clinical Infectious Diseases* **22**, 379–379 (1996).

64. Zhang, Q. *et al.* Accelerated dysbiosis of gut microbiota during aggravation of DSS-induced colitis by a butyrate-producing bacterium. *Sci Rep* **6**, 27572 (2016).

65. Ma, L., Tang, L. & Yi, Q. Salvianolic Acids: Potential Source of Natural Drugs for the Treatment of Fibrosis Disease and Cancer. *Frontiers in Pharmacology* **10**, 97 (2019).

66. Das, G. *et al.* Cordyceps spp.: A Review on Its Immune-Stimulatory and Other Biological Potentials. *Frontiers in Pharmacology* **11**, 2250 (2021).

67. Boskovic, I., Đukić, D. A., Maskovic, P., Mandić, L. & Perovic, S. Phytochemical composition and antimicrobial, antioxidant and cytotoxic activities of Anchusa officinalis L. extracts. *Biologia* **73**, 1035–1041 (2018).

68. Luo, C. *et al.* A Review of the Anti-Inflammatory Effects of Rosmarinic Acid on Inflammatory Diseases. *Frontiers in Pharmacology* **11**, 153 (2020).

69. Khil, L.-Y. *et al.* Mechanisms involved in Korean mistletoe lectin-induced apoptosis of cancer cells. *WJG* **13**, 2811 (2007).

70. Li, Q. *et al.* Streptococcus thermophilus Inhibits Colorectal Tumorigenesis Through Secreting β-Galactosidase. *Gastroenterology* **160**, 1179-1193.e14 (2021).

71. Zhao, L. *et al.* *Parvimonas Micra Promotes Intestinal Tumorigenesis in Conventional Apcmin/+ Mice and in Germ-Free Mice*. https://www.researchsquare.com/article/rs-25974/v1 (2020) doi:10.21203/rs.3.rs-25974/v1.

72. Long, X. *et al.* Peptostreptococcus anaerobius promotes colorectal carcinogenesis and modulates tumour immunity. *Nat Microbiol* **4**, 2319–2330 (2019).

73. Dai, Z. *et al.* Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. *Microbiome* **6**, 70 (2018).

74. Sovran, B. *et al.* Enterobacteriaceae are essential for the modulation of colitis severity by fungi. *Microbiome* **6**, 152 (2018).

75. Guo, S. *et al.* Exosomes derived from *Fusobacterium nucleatum* -infected colorectal cancer cells facilitate tumour metastasis by selectively carrying miR-1246/92b-3p/27a-3p and CXCL16. *Gut* **70**, 1507–1519 (2021).