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

Yufeng Lin1, Yali Liu1, Nick Lung-Ngai Ting1, Thomas Ngai-Yeung Kwong1, Han Jing2, Yiwei Wang3, Harry Cheuk-Hay Lau1, Xing Kang1, Changan Liu1, Junjun She2, Sunny Hei Wong1,4, Joseph JY Sung1,4, Jun Yu1

1 Institute of Digestive Disease and Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China

2 Center for Gut Microbiome Research, Med-X Institute, The First Affiliated Hospital of Xi’an Jiao tong University, Xi’an, Shaanxi, China

3 School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China

4 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

**Author contribution:**

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

**Wet Lab:** Yali Liu, Xing Kang, Thomas Kwong.

**Dry Lab:** Yiwei Wang, Han Jing, Weixin Liu, SK, Yanqiang Ding.

**Clinical guidance:** Sunny Wong

**Manuscript revision:** Nick Ting, Thomas Kwong, Harry Lau.

**Study design and supervise:** Jun Yu

**Achnowleagment**

Weixin Liu, Siu-Kin Ng, Yanqiang Ding, Andrew Maltez Thomas, Nicola Segata

**Background** Gut microbiota alterations are associated with colorectal cancer (CRC) pathogenesis. However, the role of enteric fungi, an essential component of gut microbiota, in CRC remains largely elusive. We aim to characterize the contribution of enteric fungi to the development of CRC.

**Methods** We performed shotgun metagenomic analyses of 1325 fecal samples from seven public datasets and one new cohort (454 CRC patients, 350 Adenoma patients and 216 healthy controls).xx

**Results xxx** We identified 33 differentially abundant fungal species in CRC versus healthy individuals (false discovery rate (FDR) < 0.01), of which Aspergillus rambellii showed the most significant difference (FDR = 5.13E-18). In seven of the eight cohorts, the combined bacterial and fungal biomarkers classified CRC from healthy individuals with an AUC 1.44% - 10.60% higher than the bacterial classifier. Among 14 biomarkers in the combined classifier, *A. rambellii* was the most important fungal species. Further abundance correlation analyses of the 64 differentially abundant species (33 fungi and 31 bacteria) showed that cross-kingdom interactions are associated with CRC. Particularly, strong differential correlations were shown between *A. rambellii* and two CRC-associated pathogens, *Fusobacterium* *nucleatum* and *Parvimonas* *micra*. The role of *A. rambellii* in colorectal carcinogenesis was further confirmed by our colorectal cancer stem cell proliferation experiment.

**Conclusions** This study revealed the mycobiome alterations in CRC particularly the enrichment of *A. rambellii* implying that the role of mycobiome in CRC is not negligeable.

Introduction

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer-related death globally1,2. Sporadic CRC, which arises without known contribution from germline mutations or significant family history, accounted for about 75% of CRC, implying the importance of environmental factors in CRC pathogenesis3. Recent studies have linked gut microbiota alteration to CRC occurence4–6. A meta-analysis with approximately 1,000 individuals from five cohorts has revealed the microbial signatures of genes specific for CRC7 and the association between the gut microbiome and choline degradation8. Even though >90% of the gut microbiome are composed of bacteria, a perturbed gut fungal composition has also been described to be associated with inflammatory bowel disease9, liver cirrhosis10 and CRC11. Fungi could influence the immunological responses of the host by dampening or promoting local inflammatory reactions9,10,12–14. The commensal fungi were shown to prevent hosts from colitis-associated colon cancer by prompting inflammasome activation and IL-18 maturation in murine model15. Therefore, it is apparent that fungi play a significant role in CRC development than our previous anticipation. However, apart from our previous study aiming to discover potential fungal biomarkers for CRC detection11, the compositional feature and the role of fungi in CRC pathogenesis remains unexplored.

In this study, we performed a multi-cohort fecal metagenomic analysis of eight available datasets from France16, Germany7,16, Australia17, America18,19, Canada19, Italy8, Chinese5, and Japan20. After rigorous and stringent data processing, a total of 1,329 samples from four continents, European, Occeanian, North American and Asian, including 525 healthy individuals, 350 adenoma patients, and 454 CRC patients, were included in this analysis. After consistent data reprocessing, we determined the compositional and ecologic alteraction of fungi in CRC. Our abundant differential analysis targeting the fungi genome revealed the significant enrichment of *Aspergillus rambellii* in CRC patients. Oncogenic function of *A. rambellii* in CRC was further validated in vitro and in vivo. We then evaluated and compared the differential correlations across the stages of colorectal carcinogenesis from healthy controls to adenoma, and CRC, respectively. *A. rambellii* and *Pichia kudriavzevii* exhibited the strong differential correlations with carcinogens (*F. nucleatum* and *P. micra*) and reported potential probiotics (*S. salivarius*, *A. hadrus* and *S. thermophilus*), respectively, in CRC compared with healthy controls. For clinical application, the trans-kingdoms CRC-screening models were 1.44%-10.60% higher than pure bacterial models in seven of eight studies. And *A. rambellii* ranked fourth in the importance of trans-kingdoms model’s features, second only to the three famous carcinogens *P. micra*, *F. nucleatum*, and *G. morbillorum*. All these suggested that enteric fungi, especially*A. rambellii*might play a potential role in CRC carcinogenesis.

Methodology

Sample collection and data retrieva

Hong Kong cohort with CRC, adenoma patients and healthy controls

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 at the time of stool collection. An independent Chinese cohort with 112 healthy individuals, 111 patients with CRC, and 197 patients with colorectal adenoma were recruited. A subset of these samples have been published in a previous research21.

Samples collected from NCBI dataset

Fecal shotgun metagenomic sequencing data of CRC-related studies from 2014 to 2020 with a minimum of 2 subject categories (CRC patients and healthy controls) were retrieved from the NCBI. Eight published cohorts and our recently completed but unpublished cohort were included in this meta-analysis; five of them also included adenoma patients8,16,17,19,20 (table 1 and supplementary table 1). We downloaded seven public fecal shotgun metagenomic CRC datasets from European Nucleotide Archive (ENA) using the following ENA identifiers: ERP005534 for Zeller et al.16, ERP008729 for Feng et al.17, PRJEB10878 for Yu et al.5, PRJEB12449 for Vogtmann et al.18, PRJNA389927 for Hanningan et al.19, PRJEB27928 for Wirbel et al.7, and SRP136711 for Thomas et al.8. The cohort from Yachida et al. was downloaded from the DNA Data Bank of Japan (DDBJ) with the Accession numbers: DRA006684 and DRA00815620. For our cohort, fecal metagenomic sequencing data were used from samples collected in Hong Kong from 2009 to 2012. A subset of samples in this cohort were published previously11. These nine studies were organized from eight countries and various sampling procedures, sample storage, and DNA extraction protocols.

Sample filtering

To ensure consistent and high-quality data, samples were subjected to filtering before analysis. Patients after surgery, or with ambiguous conditions (CRC, adenoma or healthy controls) were discarded. We only included the PCR-free cohort because the PCR-free kits could reduce bias and cell spike-in controls for a more accurate quantification22. Finally, we excluded the samples with low-alignment reads (less than 1,000,000), which might be due to low sequencing depth and host reads contamination. In the second part, we removed the outliers or suspected contaminated cases. These include samples with high-fungi composition (the fungi composition is more than 1% of total gut microbiota), low-Fungi composition (the fungi composition is less than 0.01% of the gut microbiota), and bacterial or fungal contaminated samples (a particular species constitutes more than 50% of the gut microbiota). Finally, the samples with low-fungal sequence depth (fungi aligned read counts less than 10,000 reads) would be discarded, which was consistent with a previous study revealing that fungi could not be detected in at least 30% of individuals23.

Sequence pre-processing and taxonomic profiling

We applied the KneadData default parameters for the quality control of all the metagenomic sequencing data. This separated microbial reads from the contaminated reads from the host or other user-defined sources using principled in silico methods. Next, taxonomic profiles were generated with the Kraken2 v2.0.9-beta across the custom database. Our custom library contained 9,543 bacterial and 909 fungal 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/>). The library was then established with the Jellyfish program by counting distinct 31-mer. We used the default parameters and discarded all reads with quality less than 20 and shorter than 50 nucleotides. 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, rarefaction, relative abundance (supplementary table 2 and supplementary table 8), and median normalization (supplementary table 3 and supplementary table 9) with the script (<https://github.com/ifanlyn95/multi-CRC-fungi>). 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 normalization means dividing the relative abundance of each feature by the median of the control group:

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

Overview of statistical analyses

To overcome the limitations of univariate statistics, relative abundance and median normalization were applied. The confounding factors, such as age, BMI, and tumor location, were not considered because these meta information were not complete in each cohort. The non-parametric test was used throughout for univariate association testing between the abundances of fungi and bacteria; all were two-sided Wilcoxon tests except were otherwise mentioned. Co-occurrence and co-exclusion relationships within fungi and between fungi and bacterial kingdoms were estimated using Differential Gene Correlation Analysis (DGCA)24. *P* value less than 0.05 after multiple comparisons correction using the false discovery rated method were considered significant except were otherwise mentioned. Areas under the receiver operating characteristic curves (AUCs) were determined by random forest model and the relative change values between trans-kingdom AUCs and pure bacteria AUCs present the models difference. Network parameters were estimated by using the network analyzer algorithm of Cytoscape, version 3.0.425, with default parameters and clustered by affinity propagation clusters methodology26 .

Differential abundance analysis

Three criteria were used to select the potential differentially abundant candidates between CRC and healthy individuals. First, we excluded the candidates with an average rarefied abundance less than 0.1% fungal composition. 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 rank-sum test. We identified differentially abundant features between two groups on a per species basis using Wilcoxon rank-sum test and with p-values being adjusted using the conservative Bonferroni correction. For the last criteria, we discarded features with an absolute value of log2 of features’ Fold Change less than 0.5. In addition, we ignored the unclassified strain of bacteria. The scripts were available on Github (https://github.com/ifanlyn95/multi-CRC-fungi).

Microbial association and network analysis

Inter-fungal and fungal-bacterial co-occurrence and co-exclusion relationships were estimated using the DGCA algorithm24. DGCA is an algorithm for systematically assessing the difference in feature-feature regulatory relationships under different conditions. In our case, DGCA was used to assess the difference of inter-fungal and fungal-bacterial correlations between different groups (CRC vs Adenoma vs Healthy controls). *P* values less than 0.05 were considered significant. When comparing the inter-fungal and fungal-bacterial correlation in different groups, DGCA leverages the permutation samples to calculate empirical p-values. The inclusion criterion for network plot features is correlation index less than -0.2 or more than 0.5. To stabilized the variance of sample correlation coefficients in different stages, the Fisher z-transformation is untilized:

where presents the sample correlation coefficient. Another important index used is the z-score, which represents the relative strength of differential correlation. The z-score is defined as:

wheres refers to the variance of the z-score in condition x. Using the difference in z-scores, a two-sided *p*-value can be calculated with the standard normal distribution. We considered the empirical p-values less than 0.05, and the absolute values of the z-score larger than 5 as a significantly different correlation between different groups. When the inter-fungal or fungal-bacterial correlation in CRC is weaker than that of the healthy control, the z-score would be positive. Whereas, if the correlation is stronger in CRC, the z-score would be negative. Based on a threshold of correlation significance (p-value less than 0.05) and the direction of correlation changes in different conditions (i.e. the correlation is stronger or weaker in CRC compared to healthy controls), species-species correlations in each condition could be categorized into three classes: significant positive correlation, no significant correlation, and significant negative correlation. As we have two conditions (CRC vs Healthy controls), there were nine classes for differential correlation analysis, namely '+/+', '+/0', '+/-', '0/+', '0/0', '0/-', '-/+', '-/0', and '-/-'. The interactions between these selected features were clustered with affinity propagation clusters methodology26.

Additional validation experiments on cancer cell line

TBA

Results

Data pre-processing of metagenomic datasets for studying the associations between mycobiome and CRC

We collected shotgun metagenomic sequencing data from eight cohorts (table 1 and supplementary table 1). All the raw sequencing data were reprocessed using the KneadData, Kraken227, and Bracken28 for taxonomic profiling. Each sample has about 107.19 (median) high-quality paired reads that match the bacterial database. 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), consistent with a previous study29 reporting that fungi make up about 0.1% of the total enteric microbes. The rarefaction curve (figure 1b) showed that all cohort samples reached a plateau at 10,000 sequencing reads. We applied strict sample filtering criteria to ensure rigorous outcomes and reduce the outlier effect (figure 1c). We finally included 1,329 samples with 454 CRC patients, 350 adenoma and 525 healthy controls.

Alterations of enteric fungal composition in CRC

We assessed the overall fungal composition, *Ascomycota* was identified to be the most abundant fungal phylum among all the cohorts, while other dominating fungal phyla showed significant inter-cohort variations (figure 2a and 2b). 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 than non-Asians (figure 2b).

In agreement with previous researches showing distorted microbial diversity in the diseased group30, the alpha diversity of enteric fungi was reduced in CRC patients compared to healthy individuals when considering all the cohorts together (figure 2d) and in three individual cohorts (chao1 index) (figure 2d). Collectively, we observed significant differences in both fungi phyla composition and alpha diversity in the CRC group compared to healthy controls.

Identification of fungal species associated with CRC

We searched for the potential enteric fungal shifts in CRC patients as compared to healthy 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 that demonstrated significant alterations (FDR < 0.01) as the core set (figure 3a and supplementary table 4). We evaluated if these 74 fungi (main set) were consistently altered across all the eight cohorts using SSTF and Wilcoxon rank-sum test. The enrichment and depletion status of the 74 species were consistent in six cohorts except the 2019\_ThomasAM and 2019\_Yachida cohorts. Interestingly, most of the 74 species in the 2019\_ThomasAM cohorts either showed significant enrichment in CRC patients or no significant difference between CRC versus healthy individuals. Very few showed depletion in CRC patients. Whereas in the 2019\_Yachida group, most of the identified 74 fungi showed weak variance in CRC patients versus healthy individuals, 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 that were consistently altered in 7 out of the eight cohorts. Ten of them were enriched in CRC patients, while the remaining five were depleted (supplementary table 6). For the 33 species in the core set, 10 were enriched, and the remaining 23 were depleted in CRC patients (figure 3c). The alterations of these 33 species in CRC patients versus healthy individuals were relatively consistent in most cohorts except 2019\_ThomasAM and 2019\_Yachida cohots. Among them, *Aspergillus* *rambellii* showed the most remarkable difference between the CRC patients and the healthy control groups (-log10FDR = 17.29).

We also compared the fungal community between CRC and adenoma patients (supplementary table 5 and supplementary figure 2). Seven fungi species differed significantly (FDR < 0.01) in both CRC patients versus adenoma patients and CRC patients versus healthy individuals. These species include *Aspergillus rambellii*, *Moniliophthora perniciosa*, *Erysiphe pulchra*, *Sphaerulina musiva*, *Phytophthora capsici*, *Aspergillus kawachii*, and *Cordyceps sp. RAO-2017* . These species belong to the *Ascomycota* phylum except *Moniliophthora* *perniciosa* and *Phytophthora capsici* (supplementary table 12). Collectively, we identified universal differentially abundant fungi in CRC patients compared to healthy individuals.

*A. rambellii* is the most significant enriched fungus in CRC

We sought to idnetify the most significant fungi candidates associated with CRC using stringent criteria. As shown in figure 3c, enriched *A.* *rambellii* and depleted *A.* *kawachii* were the two significant altered fungi in CRC. *A. rambellii* was significantly enriched in seven cohorts (figure 3d); whilst, *A. kawachii* was significantly depleted in cohorts of 2014\_ZellerG, 2016\_VogtmannE, 2017\_JunY, and our unpublished dataset (figure 3d). Collectively, our multi-cohort analysis revealed that the enriched *A.* *rambellii* and depleted *A. kawachii*, were significantly associated with CRC in multiple cohorts.

*Aspergillus rambellii* promotes colon cancer cell growth

[TBA] figure 4

Ecological networks of CRC-enriched and depleted fungi increased with CRC progresion

We evaluated the interactions among 33 core CRC-enriched and depleted fungi acorss steps of CRC progression using the correlation analysis with DGCA24. As shown in figure 5, we observed that both co-occurrence and co-excluding interactions among CRC-enriched and -depleted fungi were significantly different across the stages of healthy control, adenoma and CRC) – progressively stronger towards carcinogenesis. Four CRC-associated fungi including *Aspergillus* *rambellii*, *Erysiphe* *pulchra*, *Thielaviopsis* *punctulata*, and *Sphaerulina* *musiva,* showed significant co-occurrence centralities. These correlations weaken in adenoma and disappear in healthy individuals (figure 5 and supplementary figure 4) , indicating they are the most significant fungi in the CRC interaction network. In both CRC and adenoma conditions, only *A. rambellii* showed correlations with other fungi.

Ecological interactions among differentially abundant fungi and bacteria with CRC progresion

We performed additional ecological network analyses on the potential interplay among differentially abundant fungi and bacteria in CRC progression using DGCA24. We discovered that the fungal-bacterial correlation was progressively stronger from healthy control, ademona to CRC (figure 5, supplementary figure 5 and supplementary table 11). The number of significant trans-kingdoms interactions was increased during the CRC progression, from 143 (healthy control), to 156 (adenoma), and 184 (CRC) (supplementary table 11). And 99 fungal-bacterial correlations only appeared in CRC, which presented the important altered trans-kingdom community in CRC, such as *A. rambellii-F. nucleatum*. the 17 pairs of relationships represented by *A.rambellii-P.micra* were only significant in disease states (adenoma and CRC). m. This suggested that the fungal-bacterial interactions might be associated with CRC tumorigenesis.

Differential inter-fungal and fungal-bacterial correlation analysis in CRC versus healthy controls

After determining the inter-fungal and fungal-bacterial correlations in CRC progression, we explored if these correlations were significantly different between CRC and healthy controls. When assessing fungal-bacterial correlations, two peaks at -2 and +4 were observed in the density graph with z-score (supplementary figure 6a) . And our differential correlation analysis demonstrated distinct differences in the correlation changes among inter-fungal, inter-bacterial, and fungal-bacterial interactions(supplementary figure 6a).

We also defined the nine cases in the pair correlation comparison (supplementary figure 6b left panel). In cases pair correlation analysis, negative correlations were rare (supplementary table 11). Notably, only the intra-fungi had six '-/+' cases (supplementary table 11), which means the feature pair correlation in CRC was negative, while its association in healthy control was positive (supplementary figure 6b right panel). It might reveal some potential markers or changes in the stage alteration.

Sixty-four microbes (31 bacteria and 33 fungi) were separated into six clusters by affinity propagation clusters method with z-score (figure 6). Among these, two clusters contained most of the candidates identified. We named the biggest one the Fun\_cluster because 18 of 22 microbes were fungi. We named the second biggest one the Bac-cluster because 17 of the 21 microbes were bacteria. As the clustering results were based on the z-score, we can observe that the alteration of inter-bacteria and inter-fungal correlation have distinct differences. Notably, some bacteria were present in the Fun\_cluster while some fungi were present in the Bac\_cluster. This implies that these might be the special species that have more trans-kingdom interactions and might be important in CRC pathogenesis.

Fecal fungal-bacterial biomarkers to distingruish CRC patients from healthy subjects

To identify the significant differentially abundant bacteria between CRC and healthy individuals, we performed Wilcoxon rank-sum test with stringent selection criteria (q-value < 0.01, , unclassified species removed) (supplementary table 10). Thirty-one differentially abundant bacteria were identified in CRC, which was more significant than fungi (supplementary table 10), including CRC-related enriched bateria *Fusobacterium nucleatum*, *Parvimonas micra*, and *Gemella morbillorum*, and depleted beneficial bacteria *Roseburia* *intestinalis*, *Bifidobacterium* *bifidum*35–46, and *Streptococcus* *thermophilus*47–52.

In previous researches7,8, the classifier distinguishing CRC patients from healthy individuals was trained using bacterial markers only. We examined if the fungal markers identified in this study could improve the accuracy of the classifier and further increase the potential of using fecal metagenomic markers to early identify CRC patients from the population. We trained the model with single feature or multiple features to distinguish CRC from healthy individuals. Single feature refers to using only one fungus and bacteria in the core set as the predictor of the model. Whereas, multiple features refer to using a consortium of pure bacteria, pure fungi or a combination of fungi and bacteria in the core set as the predictor. Among the single-feature models, the average AUC value of the six features was greater than 60%. These include four bacteria (*Fusobacterium nucleatum*, *Parvimonas micra*, *Gemella morbillorum*, and *Porphyromonas asaccharolytica*) and two fungi (*Aspergillus rambellii* and *Aspergillus kawachii*) (table 2). Among these, *P. micra* showed the best performance, with an average AUC value of 67.79%, but it played a bad performance in 2016\_VogtmannE (AUC: 56.15%), in which *A. rambellii* performed the best (AUC: 67.57%). It revealed that the predictive values of fungi might supplement the bacteria in some situations. Next, we wanted to know whether the classifier model would be improved when using a combination of fungi and bacteria as the predictors. We trained and compared the multi-features model with pure fungi, bacteria, and the fungal-bacterial combination, containing 17, 12, and 14 species respectively (figure 7a, supplementary figure 3). Unexpectedly, the fungal classifier played more accurate role than the bacterial one in 2016\_VogtmannE (fungi: 77.27% vs bacteria: 70.63%) and 2019\_WirbelJ (fungi: 93.23% vs bacterial 89.39%). The AUC of classifier with combined fungal and bacterial markers was 1.44% - 10.60% greater than the bacterial classifier in seven of eight cohorts (figure 7b). Altogether, the classifier with combined fungal and bacterial markers was more accurate than the conventional pure fungal or bacterial classifiers.

Discussion

Researchers mainly focused on the relationship between gut bacteria and host pathology. Mycobiota is usually overlooked because of their low proportion in the enteric microbiome53. We performed the first comprehensive multi-cohort analysis of enteric fungi shotgun metagenomics in CRC of eight publicly available cohorts and our unpublished dataset from ? ethinal population. We were able to demonstrate the universal mycobiota alteration in CRC patients versus healthy individuals. Using robust statistical methods, we identified differentially abundant fungi and their ecological networks in stages of CRC progression.

From our alpha diversity rarefaction curve, a sequencing depth of at least ten thousand reads is required to study enteric fungi. To improve credibility and accuracy, we adopted strict criteria to obtain 1,329 from 2,052, by filtering low quality samples23,54.

Cohort heterogeneity of the enteric mycobiota was observed across different studies. From our principal component analysis, there were significant p-values for the fungal compositional differences across cohorts. This is consistent with previous studies showing the crucial roles of genetic background, age, dietary habits, lifestyle, and local environments in microbiota composition across different populations55.

We first observed the altered mycobiota composition in CRC versus healthy controls. We found that the fungal chao1 index in CRC was lower than that of healthy controls, as in the case of gut bacteria56,57. This phenomenon was also observed in other intestinal diseases such as inflammatory bowel diseases (IBD)9. The multi-cohort-analysis approach has been used to evaluate and combine results from comparable studies58 with significant advantages of reducing the influence by cohort-specific bias and increasing statistical power. Using the rank-sum test and SSTF in our analysis, we identified 33 fungi that were associated with CRC across eight cohorts. Our results suggested that *A. rambellii* was the most significant CRC-enriched fungus, which showed universal associations with CRC in seven of eight cohorts. This finding was supported by previous studies showing the ability of *A. rambellii* in synthesizing carcinogenic products, aflatoxin and aflatoxin precursor sterigmatocystin31,32. XXX.

Moreover, the most significant CRC-depleted fungus was *A. kawachii*, which is also from the genus *Aspergillus*. Even though both fungi were from the same genus, they play opposite roles in CRC (ref?). The crude enzyme extract derived from *A. kawachii* could enhance the antioxidative activities of Viscum album var. coloratum33, a promising agent for immunomodulation, treating colon cancer59 and hepatoma. The fermented silkworm produced by *A. kawachii* solid-state fermentation could inhibit the human hepatocellular carcinoma cells34. *R.* *irregularis* was the second most CRC-depleted fungus. *A. Officinalis*-*R. irregularis* symbiosis was reported to induce the production of salvianolic acid, which has anticancer effects60,61. *A. Officinalis*-*R. irregularis* symbiosis could produce rosmarinic acid, ferulic acid and caffeic acid, which related to beneficial properties of antioxidant, anti-inflammatory, and antimicrobial effects62,63. These findings support their roles of the enriched- or depleted fungi in the promotion or inhibition of colorectal carcinogenesis.

The fungal-fungal interaction in CRC was significantly different between the healthy controls and CRC patients. Most previous studies have focused on the role of a single key microorganism or metabolite in CRC development64–66. However, microbes in our gut are interacting with one another. From our results, multiple strong positive or negative correlations in CRC disappeared or weakened in adenoma or healthy control groups. It maybe the interaction of multiple species that causes the carcinogenesis. Therefore, we cannot only focus on the abundance changes of fungi in two groups but also the alteration of inter-fungal correlation. The most inter-fungal interactions in these three groups were significantly different, but *Aspergillus* *rambellii*, *Rhizophagus* *irregularis*, *Rhizophagus* *clarus*, *Phytopythium* *vexans*, and *Edhazardia* *aedis* appeared in all groups. It suggested that they might play a vital role in the stability of the entire intestinal ecology.

From our differential correlation analysis, we obtained two main clusters the Bac\_Cluster and Fun\_Cluster. Our results showed that the inter-fungal correlations were weakened in CRC, while inter-bacterial correlations were enhanced. The disruption of the inter-fungal correlation may break the healthy intestinal environment and induce colorectal carcinogenesis. On the other hand, the increased bacterial correlations in CRC may potentially contribute to colorectal carcinogenesis. Interesting results were observed when comparing the of the fungal-bacterial interactions in two conditions (CRC vs Healthy controls). Our results showed that the fungal-bacterial correlations with smaller changes across two conditions (|z-score| < 2) 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 significantly 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 described11,67.

We discovered that most reported or potential probiotics were separated in the Bac\_Cluster, and *P. kudriavzevii* had multiple correlations with probiotics in this study. Supporting evidence from previous studies showed that *P. kudriavzevii* derived metabolites possess anticancer effects by inhibiting cell proliferation and inducing intrinsic and extrinsic apoptosis in colon cancer cells68. There were also strong correlations among *A. rambellii*, *F. nucleatum*35,69,70, and *P. micra*5, from which the latter two were the famous CRC-related pathogens. A previous study revealed that the altered trans-kingdom association between bacteria and virus are associated with CRC21. We proposed that the trans-kingdom interactions between bacteria and fungi are also important colorectal carcinogenesis. However, this discovery was explored only in metagenomic sequencing study. More experiments are needed to verify and prove this hypothesis.

for marker only,fecal fungi markers could be used together with bacterial markers to improve the accuracy of distinguishing CRC patients from tumor-free healthy individuals.

In conclusion,

References

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

4. Wong, S. H. & Yu, J. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nature Reviews Gastroenterology & Hepatology* **16**, 690–704 (2019).

5. 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).

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

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

8. 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).

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

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

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

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

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

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

15. 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).

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

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

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. Shannon, P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13**, 2498–2504 (2003).

26. Leone, M., Sumedha & Weigt, M. Clustering by soft-constraint affinity propagation: applications to gene-expression data. *Bioinformatics* **23**, 2708–2715 (2007).

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

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

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

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

31. 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).

32. 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).

33. 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).

34. 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).

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

36. 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).

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

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

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

40. 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).

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

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

43. 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).

44. 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).

45. 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).

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

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

48. 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).

49. 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).

50. 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).

51. 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).

52. 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).

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. Li, J. *et al.* Fungi in Gastrointestinal Tracts of Human and Mice: from Community to Functions. *Microb Ecol* **75**, 821–829 (2018).

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

56. MetaHIT consortium *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).

57. Manichanh, C. *et al.* Reduced diversity of faecal microbiota in Crohn’s disease revealed by a metagenomic approach. *Gut* **55**, 205–211 (2006).

58. 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).

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

60. 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).

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

62. 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).

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

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

65. 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.

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

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

68. 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).

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

70. 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).

**Figure legends**