Introduction

Colorectal cancer (CRC) is the third most common cancer globally and has the second-highest mortality rate after lung cancer1,2. It is estimated that CRC occurrence rate will increase by approximately 80% to over two million cases in the next two decades3. Interestingly, sporadic CRCs arise without known contribution from germline causes or significant family history of CRC, accounted for about 75% of CRCs, implying the importance of other environmental factors in CRC pathogenesis4. Trillions of symbiotic microbes reside in the gastrointestinal (GI) tract, providing 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 CRC 6,7. 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 CRC through targeting IncRNA ENO1-IT15. Meta-analysis with approximately 1,000 individuals from five cohorts have reported the microbial signatures specific or microbial genes 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 perturbed gut fungal composition had been descrbed to associate with inflammatory bowel disease8 and liver cirrhosis9. Micro-eukaryotes could influence the immunological responses of the host by dampening or promoting local inflammatory reactions8–12. For instance, the mammalian intestinal micro-eukaryotic community interacts with the immune system through the innate immune receptor Dectin-111. And the commensal fungi could protect from colitis-associated colon cancer with the SYK-CARD9 by prompting inflammasome activation and IL-18 maturation in the colon in AOM-DSS treatment mice13. With discovering the mechanism between the intestinal micro-eukaryotes and the host, it is apparent that micro-eukaryotes is playing a more significant role on CRC development than previously anticipated. Except for our previous study14 that disclosed the fungal biomarker in CRC among humans, no other research reported the related study. The role of micro-eukaryotes, remains largely unexplored in CRC pathogenesis, mainly due to their lower abundance and lack of well-characterized reference micro-eukaryotic genomes.

In this study, we performed a meta-analysis of eight publicly available datasets and one new cohort from Chinese. A total of 1,329 samples from night cohorts among four continents, including 525 healthy individuals, 350 adenoma patients, and 454 CRC pateients were analyzed. ~~First, we discovered altered micro-eukaryotic diversity in CRC compared with healthy control. Second, we revealed a list of micro-eukaryotes that played a significant difference in CRC. Moreover, we explored the performance of selective candidates in each cohort, and we obtained two outstanding micro-eukaryotes,~~ *~~Aspergillus rambellii~~* ~~and~~ *~~Aspergillus kawachii~~*~~. Third, the micro-eukaryotic interrelationships and the association between selective micro eukaryotes and bacteria in three stages were exhibited and compared. Interestingly,~~ ~~we identified that~~ *~~A. rambellii~~* ~~and two CRC-related pathogens,~~ *~~F. nucleatum~~*~~, and~~ *~~P. micra~~*~~, showed a significant difference between the CRC and healthy control. In the last, we validated the~~ *~~A. rambellii~~* ~~and its conditioned medium promoted the viability of colon cancer cells.~~

Results

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

Eight published fecal shotgun metagenomics cohorts and our recently completed cohort (unpublished) were included in this meta-analysis 6,7,15–20. All published datasets contained at least two groups, CRC patients and healthy individuals; five published encompass the adenoma patients7,16,18–20 (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. Even though a subset of samples from this patient collective was published previously14, we have added complete follow-up clinical information (see Supplementary Table 2 and Methods). These nine studies were organized from eight countries and various sampling procedures, sample storage, and DNA extraction protocols. all raw sequencing data were reprocessed using the KneadData, Kraken221, and Bracken22 for taxonomic profiling (see Methods). 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 micro-eukaryotic genome (figure 1a). The median ratio of micro-eukaryotes to bacteria was 10-2.80 (figure 1a), which is in consistent with previously reported23 that fungi make up of 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 rarefies micro-eukaryotes counts of each individual were defined as 10,000 in the downstream analysis. We applied the strict criteria to remove a few samples to enhance outcomes rigour further and reduce the outlier effect (figure 1c). Because of the mEuk containing a low proportion, deep enough sequencing and free PCR were compulsory. Notably, one cohort19 whole-metagenomic-library preparation was employed 12 cycles of limited-cycle PCR; moreover, its sequencing size was five to ten times smaller than others. Therefore, we didn't adopt this cohort. At last, after three main filters (figure 1c and Methods), a total of 1,329 samples (525 healthy control, 350 adenoma patients, and 454 CRC characters) were accepted for downstream analysis. It is consistent with previous study24 that approximately 70% of individuals could be detected micro-eukaryotes in all gastrointestinal segments.

The enteric micro-eukaryotic composition

Consistent with previous studies and as a validation for our analysis, we observed bacterial phyla Bacteroidetes and Fusobacteria were enriched in the CRC group compared with healthy control. Conversely, Firmicutes and Actinobacteria were reduced (see supplementary figure 1). Among the micro-eukaryotic taxa, the phylum Ascomycota dominated the microbiota. Each cohort showed a significant variation for the next most abundant species. example, the second-largest abundance in Yachida's cohort from Japan Asia was Mucoromycota instead of *Basidiomycota*. *Microsporidia* *contains* less proportion in Asians compared with non-Asians (figure 2b). In the downstream analysis, we normalized the data through healthy control median in each group and each feature to reduce these effects (see Methods). We also made the phylum comparison between CRC and healthy control. In bacteria phylum level, Fusobacteria performed significantly in 6 cohorts and enriched in all in CRC compared to healthy control (see supplementary figure 2). But none showed a steady trend or difference in each study like Fusobacteria in micro-eukaryotes phylum level (figure 2c). In agreement with the previous research showed distortion in microbiome diversity in the disease stage25, alpha diversity indices were reduced in patients with CRC compared to control individuals when compared all the samples together (figure 2d). Most cohorts showed diversity reduction by alpha diversity index, chao1. Even though the alteration in micro-eukaryotes is not as apparent as in bacteria level, it still offered some difference in CRC compared with healthy control. Although there was heterogeneity in different cohorts, overall, both micro-eukaryotes phyla composition and alpha diversity were significant differences in the CRC group compared with the healthy control.

Seventy-four micro-eukaryotic species were associated with CRC through univariate meta-analysis

As previously described, the factor 'cohort' has a predominant effect on species composition because of the different DNA extraction protocols, various races, etc. An analysis of microbial alpha diversity and beta diversity also revealed that cohort heterogeneity has a more significant effect on overall microbiome composition than CRC in our data (see supplementary figure 3 and supplementary figure 4). We normalized the data through the median for each species in different cohorts and accessed the relative median abundance. It could reduce the impact of various studies and enhance the influence of other factors. We filtered the rarefied micro-eukaryotes (relative abundance < 0.1% of all the microeukaryote) and accessed 296 features (figure 3a and supplementary table 3) from 592 aligned species (see supplementary table 4). To determine the potential enteric micro-eukaryotes shift in patients with CRC, we compared the selected 296 species relative median abundance between healthy control and CRC patients. We gained 74 and 33 candidates whose adjusted p-value is less than, respectively, 0.1 and 0.01 by the Mann-Whitney U test and Bonferroni adjustment (figure 3a). We identified the 74 candidates as the main set and 33 features as the core set. It's worth remarking that the difference between CRC and healthy control of *Aspergillus* *rambellii* (-log10FDR = 17.29) was much more significant than others. In the meantime, we made the same comparison between adenoma and CRC patients (see supplementary table 4). Only six features also significantly differed (FDR < 0.01) in CRC compared with adenoma, namely *Aspergillus* *rambellii*, *Erysiphe* *pulchra*, *Thielaviopsis* *punctulata*, *Moniliophthora* *perniciosa*, *Sphaerulina* *musiva*, *Aspergillus* *ochraceoroseus*. Except for *Moniliophthora* *perniciosa* belonging to the *Basidiomycota*, the other five belong to the *Ascomycota*, the dominant feature in the phylum level. In summary, thirty-three species were selected as core-set for the downstream analysis from seventy-four significant different species around 296 none-rarefied micro-eukaryotes.

Even though the main set of 74 micro-eukaryotes performed a significant difference between healthy control and CRC patients when combining all studies, we also wanted to realize their performance in each cohort. We analyze each study with the 74 candidates through SSTF and non-parameter tests (see Methods). Three of the main CRC-CTRL-associated species consisted of the same absolute trend in the comparison. *Aspergillus* *rambellii* and *Erysiphe* *pulchra* were enriched; simultaneously, *Trichophyton* *mentagrophytes* were decreased in CRC across eight cohorts (figure 3b and supplementary table 5). In addition, all these three belong to the core set. At seven from eight with the same trend, we identified the other five rose, and eleven reduced in CRC (see supplementary table 5). However, only *Aspergillus* *rambellii* was a significant difference (p-value < 0.05) in almost all the cohorts, excluding 2019\_Thomas research (figure 3d and supplementary table 6). Apart from the 2019\_Thomas and 2019\_Yachida, the other six cohorts performed the roughly synchronous trend, especially in the 33-core-set. Among the core set, ten species were enriched in CRC; meanwhile, the reduction was twenty-three (figure 3c). As for the cohort heterogeneous, we observed that the 2019\_Yachida research performance was dissimilar, and it seemed much cleaner. Excluding the *Aspergillus* *rambellii* and few species that have an apparent difference in fold change between CRC and healthy control, most features' variance was weak and small. One more study, 2019\_Thomas, also behaved outlier performance in another section; most of its features were rich or no difference in CRC compared with the healthy control. Our results found that most cohorts performed the same trend among core-set. Most selections had at least three cohorts that significantly differed between CRC and healthy control, but their cohort heterogeneity still existed.

*Aspergillus* *rambellii* and *Aspergillus* *kawachii* were the most apparent enrichment and reduction in CRC, respectively

We next increased the cutoff value to identify the most crucial candidate associated with CRC by filtering core-set using three strict criteria, FDR < 0.01, SSTF ≥ 6, and abs(log2FC) ≥ 1 (see Methods). After filtering, *Aspergillus* *rambellii* and *Aspergillus* *kawachii* were the only two micro-eukaryotes that meet these stricter criteria (figure 3c). A. rambellii was the only candidate with a significant difference among seven cohorts, excluding the 2019\_Thomas cohort (figure 3d). And *A. kawachii* was significantly different among 2014\_ZellerG, 2016\_VogtmannE, 2017\_JunY, and our indoor dataset (figure 3d). Although they belong to the same genus, *A. rambellii* was enriched, while *A. kawachii* was less in CRC. In previous research, *A. rambellii* has been acknowledged to accumulate aflatoxins (AF) and the aflatoxin precursor sterigmatocystin (ST)26. And AF and ST are the most carcinogenic natural products known27. In contrast, *A. kawachii* was usually reported with anticancer or cancer-curing herbs, such as Korean mistletoe28 and fermented silkworm larvae29. Collectively, our meta-analysis revealed the key microeukaryote, *A.* *rambellii*, and *A. kawachii*, that was significantly correlated with CRC among multiple metagenomic studies.

Alteration in the CRC Micro-eukaryotic Ecological Association

Due to the complexity and multifactorial nature of CRC, we investigated the potential alternations of polymicrobial ecological interactions in CRC by estimating multiple micro-eukaryotes and micro-eukaryotic-bacterial selections correlations. We observed that the correlations within the micro-eukaryotic core-set network were stronger in CRC than in healthy control (figure 4). There were only four strong positive (correlation index ≥ 0.5) and three negative interactions (correlation index ≤ -0.15) in healthy control (figure 4a); meanwhile, nine high positive and four negative associations exhibit in CRC (figure 4b). In addition, nine close positives and one negative interrelationship were executed in adenoma. *Aspergillus* *rambellii*, *Rhizophagus* *irregularis*, *Rhizophagus* *clarus*, *Phytopythium* *vexans*, and *Edhazardia* *aedis* appeared in all three stages (figure 4a and supplementary figure 5). Still, their correlation was not consistent in various stages. The further exciting discovery was that co-occurrence interactions were observed among micro-eukaryotes *Aspergillus* *rambellii*, *Eysiphe* *pulchra*, *Thielaviopsis* *punctulata*, and *Sphaerulina* *musiva* in CRC. They were seen as a clustering of micro-eukaryotes that cooperates and symbiotically (figure 4a). However, these close relationships were disappeared in the adenoma or healthy control group (figure 4a and supplementary figure 5). In adenoma or healthy control, *T. punctulata*, *S. musiva*, and *E. pulchra* didn't release the strong correspondence with any other candidates, but *A. rambellii* were a high positive connection with *Moniliophthora* *Pemiciosa* in these two stages. Thus, our analysis revealed that co-occurrence micro-eukaryotic relationships might be crucial for enteric homeostasis in a healthy gut. In contrast, micro-eukaryotic dysbiosis might break the balance and provide a suitable environment for the harmful micro-eukaryotes clustering developing, which might cause colorectal carcinogenesis.

Correlation between CRC-related bacteria and selected micro-eukaryotes was Perturbed in CRC

As we know, our research was the first study about micro-eukaryotic meta-analysis in CRC, and we want to know whether our methodologies were corrected or not. We tend to utilize these criteria to select the bacterial candidates. Compared with micro-eukaryotes, the difference between bacteria in healthy controls and CRC is more significant; we gained 31 features through two filters (see Methods). At least half of the bacterial candidates have informed cancer-related30–35,35–41 or the reported probiotics42–47, including some well-known cancer-related bacteria, such as *Fusobacterium* *nucleatum*, *Parvimonas* micra, and *Gemella* *morbillorum*; and some famous probiotics, such as *Roseburia* *intestinalis*, *Bifidobacterium* *bifidum*, and *Streptococcus* *thermophilus*. This result revealed that our discovery was validated the previous analysis selection were credible. And next, we aimed to explore the associations between the micro-eukaryotes and bacteria. We utilized the same method, DGCA48, as the internal correlation of micro-eukaryotes. We discovered the relationship of microeukaryote-bacteria was weaker than internal micro-eukaryotes. However, we explored that the associations in CRC were much more potent than in healthy control, same with internal micro-eukaryotes correlations (figure 4 and supplementary table 7).

Alternative enteric microbiome in CRC and *A. rambellii* interacted domains with *F. nucleatum* and *P. micra*

Our previous work only compared the distribution, counts, or value of correlation index in different groups. Still, we intended to use a more reliable method, DGCA48, to judge the differences in the correlation of enteric microbiome between CRC and healthy control (figure 4). DGCA identified the z-score to represent the relative strength of differential association (see methods). The positive z-score refers to the correlation in CRC is weaker than the healthy control, while the negative z-score means the correlation is more positive in CRC (figure 5b).

Our results revealed a marked difference in correlations among micro-eukaryotes and bacteria in CRC and healthy (figure 5a). Correlations among micro-eukaryotes were higher in healthy controls compared to CRC (figure 5a). In contrast, correlations among bacteria were increased in CRC (figure 5a). When assessing micro-eukaryotes-bacteria correlations, two peaks at -5 and 5 were observed, indicating the micro-eukaryotes-bacteria correlations didn't exist in the gentle relationship. Collectively, our results implicated those correlations among micro-eukaryotes were weakened in CRC, while bacteria-bacteria correlations were utterly opposite. And associations of micro-eukaryotes-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 micro-eukaryotes 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 micro-eukaryotes) were separated into six clusterings with affinity propagation cluster (figure 5d). Among these, two clusterings contained most of the candidates were identified. For the first main clustering, 22 candidates were involved, and most of them were in Eukaryota. We, therefore, named this clustering as mEuk\_Cluster. In this clustering, the correlation between the enriched micro-eukaryotes was enhanced in CRC compared. We identified that *E. pulchra* and *A. rambellii* were the core micro-eukaryotes in the eEuk\_cluster. We also identified three CRC enriched bacteria, including *F. nucleatum*, *F. periodonticum*, and *P. micra* had strong correlations with these micro-eukaryotes (figure 5d). We revealed that the most outstanding micro-eukaryotes, *A. rambellii*, and the most reported cancer-related pathogens, *F. nucleatum*, were in the same clustering. Its z-score was -5.95, and it belonged to the '+/+' case (see supplementary table 8), indicating that micro-eukaryotic *A. rambellii* and bacterial *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 8). In contrast, no direct correlation between another key microeukaryote *E. pulchra,* and these three CRC-enriched bacteria was identified.

The second clustering included twenty-one candidates, and most of them 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. 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 micro-eukaryotes-micro-eukaryotes, micro-eukaryotes-bacteria, and bacteria-bacteria were quite different in CRC and identified that *A. rambellii* and two CRC-related pathogens, *F. nucleatum*, and *P. micra*, showed a significant difference between the CRC and healthy control.

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

TBA

Discussion

Micro-eukaryotes are one of the crucial inhabitants of the human gastrointestinal tract49. In the previous, most of the research would like to force on the relationship between gut bacteria and the host. Micro-eukaryotes were ignored in the microbiota studies for an extended period because of their low proportion in the enteric microbiome50. During further research, some studies9,11,12,51 revealed that micro-eukaryotes also played an essential role in the associations with the host. Our study was the first to report the enteric micro-eukaryotes altered in CRC across multiple cohorts according to the acknowledge we know.

According to the alpha diversity rarefaction curve, our results indicated that the micro-eukaryotic plateau was 10,000. The previous study revealed that beyond 60 million and 80 million reads per sample were required to recover all bacterial classification52 and the full richness of different Antimicrobial resistance (AMR) gene families53 in the fecal sample. Our research could suggest that future studies on enteric micro-eukaryotic should be conducted at a sequencing depth of at least 10 thousand micro-eukaryotes in each case. We have supplemented the gaps in sequencing depth for micro-eukaryotes and provided reliable support for follow-up micro-eukaryotic 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 humans54. We observed a significant p-value for the micro-eukaryotic composition difference in different cohorts from the principal component analysis results. This observation is consistent with previous bacterial research55,56 exposing the effects of ethnicity and technical various on gut microbiota and highlights the compulsion for combined analysis. We showed that the enteric micro-eukaryotic 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 studies57 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 microeukaryotes and 31 bacteria that significantly differed in CRC across eight cohorts. Since few studies described the micro-eukaryota associated with CRC, the bacterial populations we screened could prove that the micro-eukaryotic 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. nucleatum58, P. micra15, G. morbillorum59,60, and A. hadrus61 have been reported previously to promote CRC development and occurrence. It also proved that there might be existed several potential probiotics or pathogenic micro-eukaryotic 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 micro-eukaryota, *A. rambellii* and *R. irregularis* were reported cancer related. *A. rambellii* was the most significantly different micro-eukaryotes in CRC-control, as well as in CRC-adenoma comparison. It would create carcinogenic products, aflatoxin and aflatoxin precursor sterigmatocystin26,27. Salvianolic acids, as a result of the *A. officinalis*-*R. irregularis* symbiosis, showed a significant impact on cancer treatment62,63. 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 activities64,65. *A. rambellii* and *A. kawachii* were the most enriched or depleted micro-eukaryotes 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. coloratum28 (Korean mistletoe; KM), a promising agent for treating colon cancer66, 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 cells29. Collectively, our results evidenced the correctness of micro-eukaryotes selection, and the micro-eukaryotes 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 development67–69. 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 altered25,70. Except for the transformed micro-eukaryotic abundance, alteration in micro-eukaryotic internal relationship and micro-eukaryotes-bacteria association could partially explain colorectal tumorigenesis. The micro-eukaryotic 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 micro-eukaryotic internal correlations were interrupted in CRC, while bacterial were enhance. The disruption from micro-eukaryotic 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 micro-eukaryotes and bacteria. Our results showed that the less changed differential micro-eukaryotic-bacterial correlations contained low proportions. The CRC strengthen and weaken correlations performed the primary and secondary ratios in micro-eukaryotic-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 micro-eukaryotes to tremble rapidly, which was not similar to the warm alteration of internal-kingdom relationships as previously described14,71.

In the present study, the candidates were automatically divided into six clusters. The two main clusters, named mEuk\_Cluster and Bac\_Cluster, had the highest proportions of bacteria and micro-eukaryotes. 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 cells72. 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*30,58,73, and *P. micra*15, 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 micro-eukaryotic internal network and micro-eukaryotic-bacterial relationship alterations in CRC, indicating that synergistic intra-micro-eukaryotes and micro-eukaryote-bacteria interaction might contribute to colorectal carcinogenesis. Several micro-eukaryotes, *R. rambellii*, *R. kawachii*, and *P. kudriavzevii,* play critical roles in promoting or inhibited CRC. [TBA].

Methodology

Study inclusion and data attainment

We used PubMed and Google scholar to search for CRC-related research containing at least CRC patients and healthy controls with faecal shotgun metagenomic data. And seven published studies and one of our previous researches were included. We downloaded six public faecal shotgun CRC datasets from European Nucleotide Archive (ENA) using the following ENA identifiers: ERP005534 for Zeller et al.16, ERP008729 for Feng et al.18, PRJEB12449 for Vogtmann et al.17, 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, DRA008156 for Yachida et al.20.,

Hong Kong study recruitment and sequencing

This clinical study performed here was approved by the relevant ethics committees (Ethics Committee of Prince of Wales Hospital, Hong Kong, China, protocol NO. \*\*\*). Inform consent was obtained from all participants.

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. Deep freezing at –80°C within 24 hours of stool collection was done for long-term storage. According to the manufacturer's instructions, DNA was extracted using Qiagen (Hilden, Germany) QIAamp DNA Stool Mini Kit. 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 research74.

Sample filter criteria

In the beginning, we included 2,052 individuals from eight countries and four continents among nine cohorts. We have three primary filtering sections (figure 1c). According to previously published meta-information, some outlier characters, such as history surgery patients, IBD patients, or other disease patients, would be filtered, and 1,986 samples were left after the first step. And then, 77 individuals would be disused because of the ambiguous stage. Free PCR is usual in the general whole-metagenomic-library preparation, but Hannigan's research19 applied the 12 cycles of limited-cycle PCR. The samples in this cohort would be filtered, and 1,837 cases would leave. One sample was filtered because of the low reads' alignment (alignment reads number < 1,000,000). In the second filtering section, we intended to exclude the suspected contamination and outlier samples. Following the previous research23, micro-eukaryotes account for around 0.1% of the total intestinal flora. So, we discarded 19 high-micro-eukaryotes-abundance (RelAbuneuk > 1%) and 78 low-micro-eukaryotes-abundance (RelAbuneuk < 0.01%) samples, respectively. We recognized the samples whose one species accounted for more than 50% were contaminated. Therefore, we reduced the 69 large proportion of micro-eukaryotes and 45 large proportion of bacterial cases. Collectively, 221 samples were filtered in this section. Through the rarefaction curve (figure 1b), we could know that all cohort samples have reached or exceeded the plateau at 10,000. In the last part, we abandoned the low micro-eukaryotes sequencing depth sample (RawReadseuk < 10,000), and 296 cases were filtered. In summary, we move 216 cases for the sample sequence quality in the first section, 211 cases for reducing the outlier and contamination samples effect, and 296 cases for removing the low-micro-eukaryotes sequencing depth samples.

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 micro-eukaryotes 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, and dividing the median of the control group of each feature in various cohorts with the script (https://github.com/ifanlyn95/multi-CRC-fungi).

Feature selections criteria

We had three criteria to select the potential candidates, whether it is bacteria or micro-eukaryotes. In the most beginning, we excluded the rarefied candidates with an average abundance of less than 0.1% in all micro-eukaryotes. 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 micro-eukaryotes 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.

Association calculation and comparison

Co-occurrence and co-exclusion relationships within micro-eukaryotes and between micro-eukaryotes and bacteria were estimated using the DGCA algorithm48, 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.

Additional validation experiments on cancer cell line

TBA

Reference