

# Gut microbiota-based algorithms in the prediction of metachronous adenoma in colorectal cancer patients following surgery

**Yang Liu**

First affiliated hospital of Harbin medical university

**Rui Geng**

First affiliated hospital of Harbin medical university

**Lujia Liu**

First affiliated hospital of Harbin medical university

**Xiangren Jin**

First affiliated hospital of Harbin medical university

**Wei Yan**

First affiliated hospital of Harbin medical university

**Lei Zhao**

First affiliated hospital of Harbin medical university

**Fuya Zhao**

First affiliated hospital of Harbin medical university

**Shuang Wang**

First affiliated hospital of Harbin medical university

**Xiao Guo**

First affiliated hospital of Harbin medical university

**Yunwei Wei** (✉ [hydwyw11@hotmail.com](mailto:hydwyw11@hotmail.com))

The First Affiliated Hospital of Harbin Medical University <https://orcid.org/0000-0001-6482-4680>

---

## Research

**Keywords:** Colorectal cancer, metachronous cancer, colorectal adenoma, gut microbiota, random forest

**Posted Date:** November 21st, 2019

**DOI:** <https://doi.org/10.21203/rs.2.17564/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background/Aims:** The occurrence of metachronous colorectal cancer (CRC) in patients after surgery has been well documented. Despite high risk, there is a lack of reliable factors or models to be used for predicting the development of metachronous CRC. The objectives of this study were to examine mucosal and fecal microbiota, and to assess their values in the prediction of metachronous colorectal adenoma among CRC patients who underwent surgical removal of their primary CRC.

**Methods:** A cohort of CRC patients was prospectively enrolled, and their mucosal and fecal samples were used for analysis of gut microbiota by sequencing the 16S rRNA genes. The relatively predominant gut microbial populations, in combination with clinical risk factors, were utilized to generate Random-Forest (RF) algorithms for the predication of metachronous adenoma.

**Results:** Patients with metachronous adenoma in the MA group exhibited significantly lower mucosal microbial alpha-diversity compared to those individuals in the nMA group. Linear discriminant analysis of effect size (LEfSe) identified 10 predominant bacterial genera, some of which were identified as independent risk factors for metachronous adenoma. The microbiota-based RF model was established utilizing specific members of predominant gut microbiota and independent clinical risk factors (high body mass index (BMI) and the status of synchronous adenoma) in combination. The RF model had an AUC of 0.885 for predicting metachronous adenoma. The RF model performed well on fecal and off-tumor samples with the AUC of 0.835 and 0.889, respectively. Further, we generated a RF model by including specific bacteria taxa for differential prediction of metachronous adenoma from liver metastasis, which showed good performance with AUC of 0.86. Finally, we introduced a risk score for potential clinical application using the four independent predictive factors, and the scoring system had an AUC of 0.94. The presence of two or more risk factors for metachronous adenoma had a sensitivity and specificity of 90.9% and 89.5%, respectively.

**Conclusions:** The findings have demonstrated that the microbiota-based models and scoring system have good ability to predict the risk for developing metachronous adenoma after surgical resection. The newly established algorithms may hold potential to guide individual postoperative surveillance plan for CRC patients.

## Introduction

Colorectal cancer (CRC) is among the leading cause of cancer-related deaths worldwide. Despite substantial progress in the early diagnosis and treatment of CRC, and the fact that more than two-thirds of CRC patients received surgical resection and adjuvant therapy, a large proportion of these patients developed CRC recurrence, including local recurrence and distant metastasis. Notably, local recurrence or metachronous cancer has been shown to be associated with a high risk of cancer-related death[1]. It has been well documented that patients with a history of CRC are at an increased risk of developing metachronous CRC following surgical resection and perioperative clearing[2, 3]. As such, postoperative

colonoscopy is highly recommend for patients after surgical resection of CRC to improve survival via diagnosing metachronous CRC at an early stage, or to prevent the occurrence of metachronous CRC via detecting and removing of the precancerous colorectal polyps[4]. According to the major guidelines, an initial full colonoscopy is recommended at the time of diagnosis, or within 3–6 months following surgical intervention for, detection of synchronous lesions, while further colonoscopies should be carried out >6 months, generally 1 year after the surgical resection, followed by colonoscopies every 3–5 years for detection of metachronous cancer[5]. There is no level one evidence in support of the optimal total duration of surveillance after treatment for CRC[1].

Although high risk, there has been a lack of reliable factors to be used for predicting metachronous CRC in patients who were undertaken surgical treatment. Thus, life-long colonoscopy surveillance is needed. Currently, several factors have been shown to be associated with an increased risk of metachronous CRC, including age, previous or synchronous adenomas or history of CRC, right-sided index tumors, and microsatellite instability (MSI), many of these reported risk factors were inconsistent in the previous studies[3, 6–11]. Identification of individuals at high risk for the development of metachronous colorectal cancer is necessary to increase the efficiency of surveillance and to improve prognosis.

Recent studies have suggested that the community of microbes inhabiting the gastrointestinal tract plays an important role in the development and progression of CRC[12–14]. In fact, gut microbiota dysbiosis was found in patients with colorectal adenoma, and the disturbance became more apparent during the progression of adenoma into CRC[15]. It has been of note that gut bacteria may exert a role in tumorigenesis, and in turn may have potential as useful biomarkers for the early detection of disease[16]. A previous study has indicated that gut microbiota could be used to quantify the risk of recurrence[17]. Until now, it remains unknown if gut microbiota could hold a value in an assessment of risk for metachronous CRC or precancerous lesions such as colorectal adenoma, given the pathogenesis of CRC.

As CRC develops gradually from premalignant adenomatous, accurate prediction and early detection polyp provides an opportunity to halt this process. Our previous study found that colorectal cancer patients who developed metachronous adenoma postoperatively showed distinct fecal microbiota, which can be potentially used for diagnosis for metachronous adenoma[18]. But this difference already existed before operation or formed postoperatively is still unknown. Could gut microbiota be used as a tool to predict the risk for postoperative metachronous adenoma?

In this study, a cohort of CRC patients was prospectively enrolled and the mucosal and fecal samples were used for analysis of gut microbiota by sequencing the 16S rRNA genes. We aimed to test the hypothesis that the gut microbiota composition before surgery could be associated with the risk of developing metachronous adenoma, and thus could be used, together with other independent risk factors, to generate new algorithms for better predicting metachronous adenoma.

## Results

### *Characteristics of the study patients*

A total of 41 patients were included in this study, of which 22 patients developed metachronous adenoma [metachronous adenoma (MA) group], and the remaining 19 patients did not have any signs of metachronous adenoma [non-metachronous adenoma group (nMA)]. Demographic and clinical features between the two groups were summarized in Table 1. Body mass index (BMI) in the MA group was significantly greater than that of the nMA group (25.25 vs. 23.0,  $P < 0.05$ ). Notably, the incidence of synchronous adenoma was significantly higher in the MA versus nMA groups (15/22 vs. 7/19,  $P < 0.05$ ). No other significant differences between the two groups were observed. In addition, eight patients with liver metastasis were included for metastasis study. Information for every participant were supplied in Table S1.

### *Mucosal microbial diversity is significantly associated with metachronous adenoma*

We initially examined the correlation between mucosal microbial diversity and the development of metachronous adenoma. As shown in supplementary Fig.S1, the 16S rRNA gene sequencing reads and depths were adequate. An analysis of the mucosal microbial diversity with two methodologies (Shannon and Simpson-reciprocal indices) showed that alpha-diversity of the mucosal microbiome was significantly higher in the nMA group compared with the MA group ( $P < 0.05$  for each index) (Figs. 1A&B). A principal coordinate analysis (PCoA) on genus level with Bary-Curtis metric distance was performed for comparison of beta-diversity between the two groups. As shown in Fig. 1C, a clear clustering between the MA and nMA groups was revealed, suggesting that the mucosa microbial communities exhibited phylogenetic closeness within each group ( $P < 0.05$ ). Importantly, we excluded the possibility of any other potential contributors to the microbial diversity, such as clinical-pathological features, synchronous adenoma, BMI, sex, adjuvant therapies (Fig. S2).

### *Mucosal microbial composition and function in the MA group differs significantly from those in the nMA group*

We next determined if there were differences in the mucosal microbial composition between the MA and nMA patients using linear discriminant analysis of effect size (LEfSe). After bacterial taxa with relative abundance  $< 0.5\%$  were excluded for comparison, 10 taxa showed differentiated distribution with LAD score  $> 4.0$  on genus level. The MA group exhibited a predominance of *Escherichia-Shigella* and *Roseburia*, while the nMA group had a predominance of *Prevotella\_9*, *Herbaspirillum*, *unclassified\_k\_norank\_d\_Bacteria*, *Acinetobacter*, *Blautia*, *Faecalibacterium*, *Rhodococcus* and *Ruminococcus\_torques\_group* (Fig. 1D). We then examined the potential interactions among these 10 taxa with spearman rank test. As a result, *Escherichia-Shigella* was almost negatively correlated (Red dots) with others taxa, while the genera enriched in nMA group (Green text) positively correlated (Blue dots) each other (Fig. 1E).

Further analysis showed there were 4 taxa on phylum and 6 taxa on family level that predominated in the two groups with LAD score  $> 4.0$  (Fig. S3). We then interrogated whether the mucosal microbiome can be segregated using BMI or synchronous adenoma as grouping variables. Only one and two predominate genera with LAD score  $> 4.0$  were found respectively (Fig. S4), indicating that metachronous adenoma

rather than BMI and synchronous adenoma was the main explanation to the different microbiota composition between the two groups.

The functions of the gut microbiota were predicted using the PICRUSt analysis. 16S rRNA gene sequencing data were categorized into 328 KEGG functional pathways, pathways present in <10% of participants were removed, leaving 284 KEGG pathways for comparison. Fifty-five pathways were differentially enriched between the two groups ( $P_{\text{fdr}} < 0.1$ ) (Fig. S5). We observed significant upregulation of *bacterial invasion of epithelial cells pathway* and *Lipopolysaccharide biosynthesis protein pathway* in the MA group compared with the nMA group ( $P_{\text{fdr}} < 0.1$ ). On the contrary, *p53 signal pathway* was downregulated in the MA group ( $P_{\text{fdr}} < 0.1$ ) (Figs. 1F&H). Specifically, the potential pathogenic bacteria *Escherichia-Shigella* was positively correlated with *bacterial invasion of epithelial cells pathway* ( $r = 0.89$ ,  $P < 0.01$ ) (Fig. 1I).

#### *Microbiota profiles of the mucosal and fecal samples on and off the tumor site*

The mucosal and fecal samples were collected on/off the tumor site, referred to as on-tumor/off-tumor. Bar plots of the class taxonomic levels showed Gammaproteobacteria and Clostridia as the top two classes with higher relative abundance in all samples. The microbiota composition was similar between on-tumor and off-tumor mucosal samples, whereas fecal samples showed independent features without detecting of *unclassified\_k\_\_norank\_d\_\_Bacteria* and *Fusobacteriia* (Fig. 2A). Despite the collective differences between subjects with MA and nMA, the microbiota associated with on-tumor and off-tumor tissues in the same individual ( $n = 12$ ) did not differ significantly in PCoA (Fig. 2B). Hierarchical-Clustering analysis with Bary-Curtis distance indicated no apparent difference between the paired On/Off mucosal samples in the same individual (Fig. S6). On the contrary, fecal and mucosal samples in the same individual showed obviously different in PCoA (Fig. 2C), paired fecal and mucosa samples within the same individual did not close to each other (Fig. S7).

Next, we assessed whether fecal microbiota profiles could reflect the difference between MA ( $n = 11$ ) and nMA ( $n = 8$ ). As expected, fecal microbiota profiles in the MA and nMA patients differed significantly in PCoA analysis (Table S2, Fig. S8). The microbiota of the fecal samples in LEfSe analysis produced five genera with LAD score >4.0, with *Escherichia-Shigella*, *Blautia* and *unclassified\_k\_\_norank\_d\_\_Bacteria* were in consistent with the findings of the mucosal profiling (Fig. S9). These results indicated that even though fecal microbiota not corresponded to mucosa microbiota and only partially reflects the microbiota at the mucus layer, differences due to disease status are still evident.

#### *Gut microbiota variation of MA may still exist to some degree in patients after surgery*

Our previous cross-sectional study showed significant difference in post-operative fecal microbiota between patients with and without metachronous adenoma, and the alterations in the gut microbiota was associated with the disease progression in health-adenoma-carcinoma sequence[15], indicating that patients with occurrence of metachronous had more “carcinoma-like” gut microbiota compared to clear intestine patients. Intrigued by these pervious findings, we

examined if there was an association between preoperational fecal microbiota and metachronous adenoma using conjoint analysis by import our previous 16S rRNA gene sequence data of fecal samples. The overall alpha-diversity of post-operation patients (n = 47) was higher than that of pre-operation patients (n = 19) (data did not show). Similarly, alpha-diversity of the fecal samples were higher in the nMA patients ( $P < 0.05$  for both Shannon and Simpson reciprocal). For post-operation patients, alpha-diversity was higher in the nMA patients, whereas the difference was not statistically significant ( $P > 0.05$  for both Shannon and Simpson-reciprocal) (Fig. 3A-B). Next, *Escherichia-Shigella* was selected, as it was highly enriched and relatively abundant in both the mucosal and fecal samples in the MA patients ( $P < 0.05$ ). In addition, this difference was also found in post-operation patients without reaching statistical significance (Fig. 3C). Bar plots of the class taxonomic levels showed difference in the microbiota composition between the MA and nMA patients, as well as between the Post-MA and Post-nMA patients. It was worth noticing that the microbiota composition of the MA patients was similar to that of the Post-MA patients, while that of nMA was more similar to Post-nMA (Fig. 3D).

Analysis of similarity (ANOSIM) was performed to determine the beta-diversity between groups, in which ANOSIM gave a P value (i.e. significance levels) and a R value (i.e. the strength of the factors on the samples). As a result, R value between the MA and nMA groups was 0.204 ( $P = 0.033$ ), while R value between the Post-MA and Post-nMA groups was 0.045 ( $P = 0.068$ ), indicating that the discrepancy between patients with and without metachronous adenoma was less obvious in patients undergone surgery compared to untreated patients. R value between Post-nMA and MA or nMA ( $R = 0.707$  or  $R = 0.301$   $P = 0.001$  or  $P = 0.01$ ) were higher than that between Post-MA and MA or nMA ( $R = 0.392$  or  $R = 0.112$ ,  $P = 0.001$  or  $P = 0.13$ ) (Fig 3E), suggesting that gut microbiota of post-operation patients without metachronous adenoma more differed from CRC patients, especially from CRC patients who develop metachronous adenoma. Collectively, these results indicated that gut microbiota-based discrepancy between patients with and without metachronous adenoma remained in post-operation patients.

#### *Mucosal microbiota of MA patients shared common features with that of liver metastasis patients*

Among the studied CRC patients, eight patients developed liver metastasis as diagnosed during the follow-up and they were assigned as the MT group. We next characterized gut mucosal microbiota between MT and MA/nMA patients. More advanced TNM stage tumor in MT group compare to nMA ( $P < 0.05$ ) (Table S3). Mucosal microbiota differed between MT and nMA patients, whereas the separation was not obvious between the MT and MA patients as shown by Bary-Curtis distance based PCoA (Fig. 4A). Shannon and Simpson-reciprocal indices were lower in the MT patients compared to the nMA patients ( $P < 0.05$  for each alpha-diversity index), whereas there was no significant statistical difference between MT and MA (Figs. 4B&C). Bar plots of the class taxonomic levels showed that the global composition of microbiota in the MT group was similar to that in the MA group (Fig. 4D). LEfSe analysis identified three predominated taxa between the MT and nMA groups on genus level with LAD score  $> 4.0$ , all these three genera were consisted LEfSe analysis between the MA and nMA groups (Fig. S10). It was noted that *Escherichia-Shigella* had higher relative abundance in the MT group compared to the nMA group ( $P < 0.05$ ), whereas no significant difference was observed between the MT and MA groups (Fig 4E).

These results were confirmed by ANOSIM analysis, R value between the MT and MA groups was as low as 0.013, whereas R value between the MT and nMA groups was 0.214, nearly equal to R value of 0.243 between the MA and nMA groups. These results suggested that mucosal microbiota composition of the MA patients may share common features with that of the MT patients (Fig. 4F).

#### *Preoperative gut microbiota-based Random-Forest algorithms and scoring system in the prediction of metachronous adenoma in CRC patients after surgery*

Firstly, seven of 10 predominance bacterial genera in MA and nMA identified by LEfSe analysis, together with BMI, and synchronous adenoma were applied to logistic regression. *Herbaspirillum*, *Rhodococcus*, and *Prevotella\_9* were excluded as they were not detectable in more than 5 patients. All these variables were identified as significant risk factors for metachronous adenoma by univariate logistic regression ( $P \leq 0.1$ ) (Table 2), then multivariate logistic regression analysis was applied for independent risk factor validation. As shown in Table 3, the predominant bacterial genera, including *Escherichia-Shigella* and *Acinetobacter*, as well as BMI were identified as independent risk factors for metachronous adenoma ( $P < 0.05$ ), with a good ability for differentiating MA from nMA (AUC, 0.935).

Next, we constructed a Random-Forest (RF) algorithm using the relative abundance of the gut microbial populations with or without the clinical risk factors to predict MA. To determine the potential of bacterial taxa in discriminating MA, we aim to identify a minimal set of bacterial genera that maximally differentiated nMA from MA. Firstly, 10 predominant bacterial genera produced by LEfSe were initially screened, and a combination of *Escherichia-Shigella* and *Acinetobacter* optimized the performance of RF model (Fig S11), and thus were used to generate a new model. Ten-times and ten-fold cross validation were conducted to optimize the model in case of over-fitting. As shown in Fig. 5, the AUC for the model was 0.809 in predicting MA. Considering the potential value of some clinical factors in the prediction of MA, we hypothesized that the predominant bacterial populations and clinical factors in combination could generate a more precise RF model. To test the hypothesis, the independent clinical risk factors, including synchronous adenoma and BMI (Fig. S12), together with the predominant bacterial populations, *Escherichia-Shigella* and *Acinetobacter*, were used to build a new RF model. The AUC for the RF model was 0.885, which was greater than the AUC for the RF model using predominant bacterial populations alone (Fig. 5). This result indicated that in addition to gut microbiota, clinical features of patients possessed additional predictive ability on MA. The RF model were further tested on fecal and off-tumor samples, the AUC was 0.835 and 0.889, respectively (Figs. S13&S14), suggesting that fecal and off-tumor mucosal samples can be used for metachronous adenoma prediction as well. However, the AUC for the RF model was 0.61 on post-operational fecal samples (Fig. S15).

After having developed the RF algorithm for more accurately differentiating MA from nMA (AUC, 0.914, Fig. S16), we next attempted to build a better RF model for distinguishing MA from MT using predominant bacterial populations and clinical risk factors in combination. As identified in LEfSe, *Ralstonia* was predominated in MA, while *Rhodococcus* was predominated in MT. We therefore integrated these two genera into the present RF model to build a RF algorithm (RF1 model) for distinguishing MA

from MT. As *Ralstonia* was not a predictor in the RF1 model, *Ralstonia* was removed from the RF1 model (Fig. S17). The RF1 model had AUC of 0.89 for MA prediction and 0.90 for MT prediction (Figs. S18&S19). The optimized RF1 model had AUC of 0.86 for differentiating MA from MT (Fig. S20)

Finally, we developed a risk score for MA, which utilized the two predominant bacterial populations and the two clinical features. *Escherichia-Shigella*, BMI and synchronous-adenoma were risk factors, and the presence of each one was assigned one point, while the absence of beneficial factor, *Acinetobacter*, was scored one point. As a result, the total risk scores ranged from zero to four points and the risk score showed an AUC of 0.94 for the prediction of MA. Further, the presence of two or more risk factors had a sensitivity and specificity of 90.9% and 89.5%, respectively (Table 4). Hierarchical clustering (Bary-Curtis distance) based on these four predictive factors was generated, showing, a good differential prediction between MA and nMA (Fig. 6). On the basis of the main findings, the risk for MA can be simply stratified into the following categories: (1) The absence of negative-risk genus *Acinetobacter* combined with the presence of positive-risk genus *Escherichia-Shigella* may predict the development of metachronous adenoma; (2) The presence of synchronous adenoma together with high BMI may predict occurrence of metachronous adenoma; (3) The absence of positive-risk genus *Escherichia-Shigella*, the presence of negative-risk genus *Acinetobacter*, and normal BMI or absence of synchronous adenoma seemed to guarantee a normal outcome.

## Discussion

We conducted the first study, to the best of our knowledge, to assess the correlation between preoperational gut microbiota and metachronous adenoma among Chinese CRC patients after surgery and to develop novel microbiota-based predictive models. The novel findings were summarized as follows: (1) There was a significant correlation between preoperative gut microbiota and the development of metachronous adenoma among CRC patients after surgery; (2) Specific members of the predominant gut microbiota, including *Escherichia-Shigella* and *Acinetobacter*, was identified as independent risk factors for metachronous adenoma; (3) The microbiota-based RF model was established utilizing these specific members of predominant gut microbiota combined with independent clinical risk factors (BMI) and the status of synchronous adenoma, showing a good performance (AUC, 0.885) to predict metachronous adenoma among CRC patients after surgery; (4) The microbiota-based RF model exhibited good ability in the prediction of metachronous adenoma using fecal and off-tumor samples (AUC, 0.835 and 0.889, respectively); (5) On the basis of the microbiota-based RF model, a modified RF algorithm was constructed to differentiate metachronous adenoma from liver metastasis among CRC patients after surgery with AUC of 0.86; (6) A risk scoring system was proposed after assigning the four independent predictive factors with quantitative points, and the presence of two or more risk factors for the development of metachronous adenoma had AUC of 0.94 and high sensitivity and specificity (90.9% and 89.5%, respectively).

In the present study, the incidence of metachronous adenoma after colectomy was 53.7%, much higher than 25% as reported previously [2, 3]. The findings demonstrated that the presence of synchronous



adenoma and high BMI were positively associated with metachronous adenoma. It is generally accepted that obesity is associated with reduced microbial diversity and a compositional change in the gut microbiota[19]. Additionally, obesity is an established risk factor for CRC[20]. The presence of synchronous CRC was identified as a significant risk factor for developing metachronous colorectal neoplasms[3]. However, it was of note in this study that microbiota cannot be separated obviously by BMI or synchronous adenoma as classifier. In addition, we assessed for potential contributors to microbial diversity and were not able to find any significant association. This is likely because that metachronous adenoma was truly the main explanation to the microbiota variations.

Coloscopy mucosa biopsies were used rather than intraoperation specimen because we thought the microbiota of samples from resected tumor after operation may be disturbed by clinical intervention, such as the preventive antibiotics application before operation. A clear clustering between the MA and nMA patients was observed. Alpha-diversity of the mucosal and fecal samples were both lower in the MA patients compared with the nMA patients. For post-operation patients, alpha-diversity of fecal samples microbiota remained lower in the MA patients without reaching statistical significance. As low diversity microbiota indicated unstable ecosystem, one theme that has emerged from many large surveys of gut microbial communities is that low microbial diversity is almost invariably associated with disease[21].

It was noticed that in our study there were predominated bacterial taxa in both MA and nMA respectively. Specifically, we found *Escherichia-Shigella* enriched in MA were almost negatively correlated with others genera while the genera enriched in nMA group positively correlated each other. This co-abundance groups (CAG) of bacterial taxa resembled the previously formulated concept of enterotypes. The bacterial taxa belonged to a CAG may relate to each other not only quantitatively but also functionally[22].

*Escherichia-Shigella* was identified as the most abundant genus in the MA patients in this study. *Escherichia* is a genus of bacteria, consisting of eight species including the well-known *Escherichia coli* (*E. coli*). Although *Shigella* is technically a separate genus with four species, which are inseparable from *E. coli* in terms of DNA, for which they are commonly bracketed together and named *Escherichia-Shigella* in 16S rRNA gene based microbiota study. All these species belong to the Enterobacteriaceae family, which was highly enriched in the MA patients as well. *Escherichia-Shigella* has been shown to produce Colibactin, which is encoded by polyketide synthase (*pks*) genotoxicity island[23] and possesses the capacity to damage DNA and lead to CRC development [24, 25]. Mucosa-associated *E. coli* has been found to be significantly more prevalent in CRC tissue and correlate with tumor stage and prognosis [26].

*E. coli* and *Shigella* have been shown to increase intestinal permeability in this intestinal disorder, likely due to down-regulation of tight junction proteins [27]. Alterations in barrier function may allow luminal bacteria access to the epithelium, where direct contact with host cells enables the bacteria to transfer or to deliver specific toxins. Our study demonstrated that *Escherichia-Shigella* was positively correlated with *bacterial invasion of epithelial cells* pathway, which was also enriched in the MA patients as identified by PICRUST method. The *bacterial invasion of epithelial cells* pathway indicates that the potential pathogens such as *Escherichia-Shigella*, Enterococcus and a member in Enterobacteriaceae

could adhere the surface of host cells, cross host epithelial barriers, and get access to internal tissues, thereby promoting their dissemination inside the host [28].

We also observed significant upregulation of *Lipopolysaccharide biosynthesis protein* pathway and downregulation of *p53 signal* pathway in the MA patients. Alterations in signaling of TLR4, which is the major receptor for LPS, have been linked to the progression of CRC [29]. It has been well accepted that the p53 signaling plays an important role in the co-ordination of the cellular response to different types of stress such as DNA damage and hypoxia. Inactivation of p53 has been implicated in the CRC progression[30]. Thus, we have reasons to propose that CRC patients who developed a metachronous adenoma possessed more serious and CRC related gut microbiota.

It was striking that there was high similarity in the mucosal microbiota of paired on-off tumor samples with regard to overall composition of the microbiota. In contrast, paired fecal and mucosal samples had lower similarity. These findings were consistent with a previous study[31]. We found that microbiota in the fecal samples can be also separated between the MA and nMA groups. Specifically, differed individual taxa of the fecal samples between the MA and nMA patients were largely in agreement with that of the mucosal samples. For example, CRC-related genus, *Escherichia-Shigella*, showed relatively higher abundance in MA fecal samples as well. As such, even though fecal microbiota differed from and may only partially reflect the microbiota at the mucus layer, differences due to disease status are still evident. Unlike mucosal samples which mainly reflected the local microbiota, the fecal samples may be a representative for the whole gut environment. It is possible that except for the lesion site, the normal sites of the colon may also possess more CRC-related bacteria in the MA patients compare to the nMA patients, and that maybe contribute to the development of metachronous adenoma.

Our previous cross-sectional study showed differences in post-operative fecal microbiota between patients with and without metachronous adenoma[18]. We wonder whether such difference could exist in the preoperational fecal samples, and if there could be relationship between pre- and post-colectomy gut microbiota profile. As observed in our study, similar to pre-operation CRC patients, lower microbiota diversity and higher abundance CRC-related bacterial taxa were characteristics for MA in the post-operative CRC patients, whereas they were not obvious as pre-operation patients. ANOSIM results also showed the distance value between MA and nMA was high in preoperative CRC patients. In addition, the fecal microbiota of post-operative MA patients were more closed to preoperative CRC patients. Collectively, these findings suggest residual microbiota features for metachronous adenoma in post-operation patients.

In addition to metachronous neoplasm, cancer metastasis is another poor prognosis for post-operative patients. We examined common and different features between MA and MT. MT patients showed totally different mucosa microbiota from nMA patients. Similarly, MT patients showed lower Alpha-diversity and higher CRC related bacterial genus. ANOSIM analysis showed more closed distance between MT and MA. Both MA and MT possess relatively abundant *Escherichia-Shigella*, which were significantly more prevalent in CRC tissue and correlated with poor prognosis (tumor-node-metastasis stage)[32].

*Escherichia-Shigella* belongs to Enterobacteriaceae, which is the main resource of LPS. LPS was found to promote CRC metastasis by stimulating TLR4 signaling and increasing integrin mediated cell adhesion[33]. We found *lipopolysaccharide biosynthesis protein* pathway was more active in MT compared to nMA without reaching statistical significance. It remains unclear whether metachronous adenoma could be correlated with metastasis.

In this study, we identified novel microbiome biomarkers for prediction of the metachronous adenoma. It is important to highlight that metachronous adenoma is a complex disease that occurs as a combination of microbial colonization, patient genetic background and other environment factors. Given that, we established the RF model utilizing the gut microbiota together with the clinical risk factors to predict MA. As a result, a combination of *Escherichia-Shigella*, *Acinetobacter* and BMI improved the performance of the RF model without integrating BMI. We observed that the key predictor was *Escherichia-Shigella* in this model which was in agreement with logistic regression result, showing that *Escherichia-Shigella* was an independent risk factors with an overt OR value of 53.245. Although synchronous adenoma was not included in the RF model, in view of it as a risk factor for metachronous adenoma and in order to translate our result to clinical application, we developed a risk score based on presence of the negative prognostic genus *Escherichia-Shigella*, absence of the positive prognostic genus *Acinetobacter*, together with BMI and the traditionally accepted risk factors, synchronous adenoma.

As mentioned above, paired on-off tumor mucosa microbiota were similar to each other, while fecal microbiota were partly reflected the mucosa microbiota but can still be separated according to MA status. We also applied the newly established RF model to these samples. As expected, the RF model performed well for off-tumor mucosal and fecal samples. In addition, this RF model could predict MT, this may because MT shared some common feature with MA. We further optimized the RF model, showing good performance in predicting MA, and distinguishing MA from MT. It is possible that the difference between the post-operative fecal microbiota of patients developed metachronous adenoma or not is subtle. This is supported by the reduced ability of our models to correctly classify patients with or without metachronous adenoma by post-operative fecal microbiota.

The development of metachronous adenoma is complex, involving microbial colonization, host immune response, and other CRC-related factors, all of which can be further confounded by environment and socioeconomic status[34]. The microbiome-based biomarkers described in this study enable predicting metachronous adenoma and metastasis, and this may highlight the significance of gut microbiota for evaluation of CRC prognosis. It may merit attention that the risk score can help to identify patients at high risk of developing metachronous adenoma within one year after surgery. Although this clinical condition is an excellent model for investigating whether dysbiosis precedes metachronous adenoma, we are not be able to draw conclusions regarding the causality on the basis of our data. If CRC patients at high risk for metachronous adenoma can be identified preoperatively, individual post-operative surveillance plan can be made to prevent the occurrence of metachronous CRC.

Our study may have a number of limitations. Firstly, patients were followed up but mucosal or fecal samples were not collected after surgery, for which we can not make before-after analysis in the same cohort of patients. But we made conjoint analysis with previous data of another cohort patients. Secondly, the sample size was relatively small, and the predict potential of the selected biomarkers should be evaluated in an independent cohort. Although no external cross-validation was achieved in this study, but sufficient internal cross-validation with different samples was made in this study. Thirdly, the patients were followed up for 12 months, and no metachronous CRC was observed for further study. We will continue to follow these patients and to study the predictive microbiota for metachronous CRC in the future.

## **Conclusion**

The findings have demonstrated that specific members of the dominant gut microbiota as non-invasive biomarkers for prediction of metachronous adenoma or CRC after surgical resection. The newly established RF algorithm and the risk scoring system have good ability in predicting the development of metachronous adenoma after surgical resection, and therefore the novel approaches hold potential to guide individual postoperative surveillance plan for CRC patients in future clinical application.

## **Methods**

### **Study patients**

A cohort of CRC patients was prospectively enrolled at the First Affiliated Hospital of Harbin Medical University during the period between September 2017 and April 2019. All the patients were diagnosed with primary colorectal adenocarcinoma and underwent surgical resection of CRC. During the enrollment, the patients who had the following conditions were excluded from this study: 1) taking antibiotics in one month prior to colonoscopy examination; 2) previous diagnosis of CRC, IBD or IBS; 3) medical history of surgery, radiation, or chemotherapy. A total of 41 CRC patients were enrolled in this study. Of these, 28 patients had colonoscopic mucosal samples, 13 patients had both fecal and colonoscopic mucosal samples, 16 patients failed to accomplish initial full colonoscopy at the time of diagnosis due to tumor obstruction.

### **Sample collection**

Cold biopsy forceps were used for collection of coloscopic mucosal biopsies from CRC tissues and adjacent, cancer-free tissues (at least 5 cm away from lesions) respectively. Fecal samples were taken before colonoscopic examination. All the samples were snap-frozen in cryovial immediately following collection and stored at  $-80\text{ }^{\circ}\text{C}$  until DNA extraction.

### **Follow up**

All the study patients were followed up for 12 months, and at the end of follow-up they were scheduled to undergo surveillance colonoscopy. For the patients with a malignant bowel obstruction (MBO), they were asked for colonoscopy within 2–4 months after surgery to detect synchronous lesions, followed by repeat colonoscopy at 1 year to detect metachronous cancer according to the guidelines[5]. For patients with synchronous adenoma detected before surgery, endoscopic mucosa resection (EMR) was performed to remove the lesion prior to colon resection. The primary endpoint was metachronous adenoma, which was defined as metachronous adenoma detected during coloscopic examination one-year after surgery. As we intended to assess whether gut microbiota has common features among patients with poor prognosis such as liver metastasis, the second endpoint was set as hepatic metastasis post operation.

## DNA extraction and 16S rRNA gene sequencing for bacterial identification

The fecal and mucosal samples as described in the sample collection were used for DNA extraction. In brief, microbial DNA was extracted using a DNA kit (Bio-Tek, GA, USA) according to the manufacturer's instructions and used for an amplification of the hypervariable regions (V3-V4) of the bacterial 16S rRNA gene. The resulting amplicons were purified and pooled in equimolar concentrations, followed by paired-end sequencing (2 × 300) on an Illumina MiSeq platform (Illumina, San Diego, USA), which was performed by Majorbio Bio-Pharm Technology (Shanghai, China). After the raw reads were filtered and quality control was conducted, OTUs were clustered with a 97% similarity cutoff using UPARSE (version 7.1, <http://drive5.com/uparse/>), following which the identified chimeric sequences were removed using UCHIME. With the RDP Classifier algorithm, taxonomic assignments for the 16S rRNA gene sequences were made (<http://rdp.cme.msu.edu/>) with the GreenGene 16S rRNA gene database at a confidence threshold of 70%.

## Bioinformatics and statistical analysis

Both  $\alpha$ -diversity (Simpson reciprocal and Shannon index) and  $\beta$ -diversity (Bray-Curtis distance) were examined using QIIME (Version 1.7.0), in which  $\beta$ -diversity was observed using PCoA to reduce the dimension of the original variables with the Vegan and ggplot2 packages in R. While ANOSIM of the distance matrices in the vegan package in R was used to quantize the similarity between groups[35]. Hierarchical clustering on the basis of similarities in the combination of variables was carried out using Pvcust in R. The microbiota were characterized using the Linear discriminant analysis Effect Size (LEfSe) method for representative taxa discovery, emphasizing both significance and biological relevance[36]. Functional composition of the gut metagenomes were predicted and profiled in accordance with the 16S rRNA gene sequences using PICRUST with level III KEGG database pathways[37]. Both PICRUST and LEfSe were accomplished online (<http://huttenhower.sph.harvard.edu/galaxy>). A heatmap was created to express the results with the heatmap package in R. The microbiota features were further analyzed as categorical variables using an univariate logistic regression to screen risk factors. The

optimal cut-off for each bacterial group was determined by ROC analysis. Variables with a P value <0.1 on the univariate analysis were selected for further forward stepwise multivariate logistic regression to identify independent predictors. Odds ratios (ORs) were calculated with a 95% confidence interval (CI). The random forest algorithm was used to create the classification models. The optimal number of variables was determined by maximizing the AUC (area under the curve of the receiver operator characteristic) with AUCRF package, then caret (v6.0.76) and random forest R package to build model. To avoid over-fitting of the data in the model, 10-time and 10-fold cross validation was made. The resulting model was subsequently used for testing the samples.

All categorical data were presented as number of cases and percentages, while continuous data were shown as median with range. Categorical variables were compared by the Pearson's chi-square ( $\chi^2$ ) test, and continuous variables by Mann–Whitney U test where appropriate. Statistical analysis of the data was performed using SPSS (SPSS version 19, La Jolla, CA, USA). Wilcoxon rank sum test and Multiple hypothesis tests were used for analysis of continuous and categorical data and adjusted using the Benjamini and Hochberg FDR. The results with an FDR threshold lower than 0.1 were considered significant differences. Spearman's rank test was used for correlation analysis and a P value less than 0.05 was considered statistically significant.

## Abbreviations

CRC: Colorectal cancer; MA: metachronous adenoma; nMA: non-metachronous adenoma; LEfSe: Linear discriminant analysis of effect size; AUC: area under the curve; MSI: microsatellite instability; IBD: Inflammatory bowel disease; IBS: Irritable Bowel Syndrome; EMR: endoscopic mucosa resection; OTUs: Operational taxonomic units; PCoA: Principal coordinate analysis; PICRUST: Phylogenetic Investigation of Communities by Reconstruction of Unobserved States; KEGG: Kyoto Encyclopedia of Genes and Genomes; ROC: receiver operating characteristic; ORs: Odds ratios; BMI: Body Mass Index; ANOSIM: Analysis of similarity; MT: metastasis; RF: Random Forest; CAG: co-abundance groups.

## Declarations

## Ethics approval and consent to participate

The study protocol was reviewed and approved by the Research Ethics Committee of the First Affiliated Hospital of Harbin Medical University. Each patient had provided a written informed consent. The study involving human subjects was strictly performed according to international guidelines regarding the conduct of clinical trials. This study was registered at [ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT03667495) (NCT03667495)

## Consent for publication

Not applicable

# Availability of data and material

The raw sequences have been deposited in the NCBI Sequence Read Archive (No. [SUB6297611](https://www.ncbi.nlm.nih.gov/Traces/study/)), and the necessary metadata can be found at <https://www.ncbi.nlm.nih.gov/Traces/study/> by searching the respective SRA study accession.

# Competing interests

The authors declare that they have no competing interests

# Funding

This work was supported by grant 81970466 from National Natural Science Foundation of China.

# Author contribution

YL and YW W conceived the study design, RG and LJ L recruited and followed up the patients,

XR J coordinated with patients transported patient samples, L Z and FY Z maintained patient records. YL performed the sequencing analysis, W Y and S W contributed to the data analyses, YL and YW W drafted the manuscript, and all authors read and approved its final version.

# Acknowledgements

We thank the patients enrolled in the study for their commitment with the project. Sequence processing and analysis were performed using the resources of the Majorbio Bio-Pharm Technology (Shanghai, China).

# References

- 1.van der Stok EP, Spaander MCW, Grünhagen DJ, Verhoef C, Kuipers EJ: *Surveillance after curative treatment for colorectal cancer. Nature Reviews Clinical Oncology* 2016, *14*(5):297–315.
- 2.Mulder SA, Kranse R, Damhuis RA, Ouwendijk RJ, Kuipers EJ, van Leerdam ME: *The incidence and risk factors of metachronous colorectal cancer: an indication for follow-up. Dis Colon Rectum* 2012, *55*(5):522–531.
- 3.Balleste B, Bessa X, Pinol V, Castellvi-Bel S, Castells A, Alenda C, Paya A, Jover R, Xicola RM, Pons E *et al: Detection of metachronous neoplasms in colorectal cancer patients: identification of risk factors. Dis Colon Rectum* 2007, *50*(7):971–980.

4. Kahi CJ, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, Kaltenbach T, Lieberman D, Levin TR, Robertson DJ, Rex DK *et al*: *Colonoscopy Surveillance After Colorectal Cancer Resection: Recommendations of the US Multi-Society Task Force on Colorectal Cancer*. *Gastroenterology* 2016, *150*(3):758–768 e711.
5. Meyerhardt JA, Mangu PB, Flynn PJ, Korde L, Loprinzi CL, Minsky BD, Petrelli NJ, Ryan K, Schrag DH, Wong SL *et al*: *Follow-up care, surveillance protocol, and secondary prevention measures for survivors of colorectal cancer: American Society of Clinical Oncology clinical practice guideline endorsement*. *J Clin Oncol* 2013, *31*(35):4465–4470.
6. Shureiqi I, Cooksley CD, Morris J, Soliman AS, Levin B, Lippman SM: *Effect of age on risk of second primary colorectal cancer*. *J Natl Cancer Inst* 2001, *93*(16):1264–1266.
7. Shitoh K, Konishi F, Miyakura Y, Togashi K, Okamoto T, Nagai H: *Microsatellite instability as a marker in predicting metachronous multiple colorectal carcinomas after surgery: a cohort-like study*. *Dis Colon Rectum* 2002, *45*(3):329–333.
8. Park IJ, Yu CS, Kim HC, Jung YH, Han KR, Kim JC: *Metachronous colorectal cancer*. *Colorectal Dis* 2006, *8*(4):323–327.
9. Gervaz P, Bucher P, Neyroud-Caspar I, Soravia C, Morel P: *Proximal location of colon cancer is a risk factor for development of metachronous colorectal cancer: a population-based study*. *Dis Colon Rectum* 2005, *48*(2):227–232.
10. Bouvier AM, Latournerie M, Jooste V, Lepage C, Cottet V, Faivre J: *The lifelong risk of metachronous colorectal cancer justifies long-term colonoscopic follow-up*. *Eur J Cancer* 2008, *44*(4):522–527.
11. Yamazaki T, Takii Y, Okamoto H, Sakai Y, Hatakeyama K: *What is the risk factor for metachronous colorectal carcinoma?* *Dis Colon Rectum* 1997, *40*(8):935–938.
12. Wu S, Rhee KJ, Albesiano E, Rabizadeh S, Wu X, Yen HR, Huso DL, Brancati FL, Wick E, McAllister F *et al*: *A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses*. *Nat Med* 2009, *15*(9):1016–1022.
13. Arthur JC, Perez-Chanona E, Muhlbauer M, Tomkovich S, Uronis JM, Fan TJ, Campbell BJ, Abujamel T, Dogan B, Rogers AB *et al*: *Intestinal inflammation targets cancer-inducing activity of the microbiota*. *Science* 2012, *338*(6103):120–123.
14. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, Clancy TE, Chung DC, Lochhead P, Hold GL *et al*: *Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment*. *Cell Host Microbe* 2013, *14*(2):207–215.
15. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, Zhang D, Xia H, Xu X, Jie Z *et al*: *Gut microbiome development along the colorectal adenoma-carcinoma sequence*. *Nat Commun* 2015, *6*:6528.



16. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Bohm J, Brunetti F, Habermann N *et al*: *Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol* 2014, *10*:766.
17. Sze MA, Baxter NT, Ruffin MT, Rogers MAM, Schloss PD: *Normalization of the microbiota in patients after treatment for colonic lesions. Microbiome* 2017, *5*(1):150.
18. Jin Y, Liu Y, Zhao L, Zhao F, Feng J, Li S, Chen H, Sun J, Zhu B, Geng R *et al*: *Gut microbiota in patients after surgical treatment for colorectal cancer. Environ Microbiol* 2019, *21*(2):772–783.
19. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto JM, Kennedy S *et al*: *Richness of human gut microbiome correlates with metabolic markers. Nature* 2013, *500*(7464):541–546.
20. Kerr J, Anderson C, Lippman SM: *Physical activity, sedentary behaviour, diet, and cancer: an update and emerging new evidence. The Lancet Oncology* 2017, *18*(8):e457-e471.
21. Round JL, Palm NW: *Causal effects of the microbiota on immune-mediated diseases. Sci Immunol* 2018, *3*(20).
22. Flemer B, Lynch DB, Brown JM, Jeffery IB, Ryan FJ, Claesson MJ, O'Riordain M, Shanahan F, O'Toole PW: *Tumour-associated and non-tumour-associated microbiota in colorectal cancer. Gut* 2017, *66*(4):633–643.
23. Nougayrède JP, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, Gottschalk G, Buchrieser C, Hacker J, Dobrindt U, Oswald E: *Escherichia coli induces DNA double-strand breaks in eukaryotic cells. Science (New York, NY)* 2006, *313*(5788):848–851.
24. Wu S, Rhee KJ, Albesiano E, Rabizadeh S, Wu X, Yen HR, Huso DL, Brancati FL, Wick E, McAllister F *et al*: *A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. Nature medicine* 2009, *15*(9):1016–1022.
25. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan TJ, Campbell BJ, Abujamel T, Dogan B, Rogers AB *et al*: *Intestinal inflammation targets cancer-inducing activity of the microbiota. Science (New York, NY)* 2012, *338*(6103):120–123.
26. Bonnet M, Buc E, Sauvanet P, Darcha C, Dubois D, Pereira B, Déchelotte P, Bonnet R, Pezet D, Darfeuille-Michaud A: *Colonization of the human gut by E. coli and colorectal cancer risk. Clinical cancer research: an official journal of the American Association for Cancer Research* 2014, *20*(4):859–867.
27. Cinova J, De Palma G, Stepankova R, Kofronova O, Kverka M, Sanz Y, Tuckova L: *Role of intestinal bacteria in gliadin-induced changes in intestinal mucosa: study in germ-free rats. PLoS One* 2011, *6*(1):e16169.

28. Ribet D, Cossart P: *How bacterial pathogens colonize their hosts and invade deeper tissues. Microbes and infection* 2015, *17*(3):173–183.
29. Neal MD, Sodhi CP, Jia H, Dyer M, Egan CE, Yazji I, Good M, Afrazi A, Marino R, Slagle D *et al*: *Toll-like receptor 4 is expressed on intestinal stem cells and regulates their proliferation and apoptosis via the p53 up-regulated modulator of apoptosis. The Journal of biological chemistry* 2012, *287*(44):37296–37308.
30. Levine AJ, Oren M: *The first 30 years of p53: growing ever more complex. Nat Rev Cancer* 2009, *9*(10):749–758.
31. Nakatsu G, Li X, Zhou H, Sheng J, Wong SH, Wu WK, Ng SC, Tsoi H, Dong Y, Zhang N *et al*: *Gut mucosal microbiome across stages of colorectal carcinogenesis. Nat Commun* 2015, *6*:8727.
32. Bonnet M, Buc E, Sauvanet P, Darcha C, Dubois D, Pereira B, Dechelotte P, Bonnet R, Pezet D, Darfeuille-Michaud A: *Colonization of the human gut by E. coli and colorectal cancer risk. Clin Cancer Res* 2014, *20*(4):859–867.
33. Hsu RY, Chan CH, Spicer JD, Rousseau MC, Giannias B, Rousseau S, Ferri LE: *LPS-induced TLR4 signaling in human colorectal cancer cells increases beta1 integrin-mediated cell adhesion and liver metastasis. Cancer Res* 2011, *71*(5):1989–1998.
34. Brenner H, Kloor M, Pox CP: *Colorectal cancer. Lancet* 2014, *383*(9927):1490–1502.
35. Buttigieg PL, Ramette A: *A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. FEMS Microbiol Ecol* 2014, *90*(3):543–550.
36. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C: *Metagenomic biomarker discovery and explanation. Genome Biol* 2011, *12*(6):R60.
37. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R *et al*: *Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol* 2013, *31*(9):814–821.

## Tables

**Table 1. Clinico-pathological Characteristics of Patients.**

	MA(n=22)	nMA(n=19)	P-value
<b>Gender</b>			
Female	12	6	0.139
Male	10	13	
<b>Age (years)<sup>a</sup></b>			
	63[58.5-68.75]	61.3[53-68.5]	0.619
<b>BMI<sup>a</sup></b>			
	25.25[22.75-27.98]	23.0[21.74-23.7]	0.011*
<b>Synchronous adenoma</b>			
Yes	15	7	0.045*
No	7	12	
<b>Bowel obstruction<sup>d</sup></b>			
Yes	10	6	0.364
No	12	13	
<b>Hematochezia</b>			
Yes	11	11	0.613
No	11	8	
<b>Tumor size<sup>ac</sup></b>			
	4[3.6-4.2]	4[3.1-4.75]	0.854
<b>Tumor location<sup>b</sup></b>			
Left hemi-colon	7	2	0.171
Right hemi-colon	3	6	
Rectum	12	11	
<b>CEA<sup>a</sup></b>			
	6.725[2.38-14.30]	3.97[2.37-12.83]	0.896
<b>CA 19-9<sup>a</sup></b>			
	12.31[7.15-65.44]	12.55[10.99-20.06]	0.744
<b>Adjuvant therapy</b>			
Yes	13	12	0.790
No	9	7	
<b>TNM-Stage</b>			
I	2	2	0.537
IIA	17	11	
IIIA	0	1	
IIIB	3	5	

\*P<0.05, different from controls by Wilcoxon rank-sum test or Chi-squared test for continuous or categorical variables, respectively.

<sup>a</sup> Data shown as median (1st and 3rd quartile).

<sup>b</sup>Tumor location: splenic flexure, descending, sigmoid, rectosigmoid were classified as left hemi-colon; ileocecal, ascending, hepatic flexure, transverse were classified as right hemi-colon.

<sup>c</sup>Tumor size definition: maximum diameter.

<sup>d</sup>Bowel obstruction was defined when colonoscopy cannot pass through the tumor obstruction.

CEA: carcinoembryonic antigen.

CA19-9: carbohydrate antigen 19-9.

**Table 2. Univariate logistic regression predicting MA.**

	Cut-off value	OR	95%CI	P value
erichia-Shigella	564.5	10.000	2.350 to 42.547	0.002*
classified_k_norank_d_Bacteria	147	0.206	0.037 to 1.131	0.069
alibacterium	608.5	0.172	0.044 to 0.672	0.011*
inococcus]_torques_group	10.5	0.097	0.011 to 0.871	0.037*
ia	732.5	0.065	0.007 to 0.593	0.015*
ctobacter	45	0.056	0.006 to 0.492	0.009*
buria	55	0.172	0.044 to 0.672	0.011*
chronous adenoma		3.673	1.007 to 13.395	0.049*
		1.396	1.069 to 1.824	0.014*

\*P<0.05. MA: metachronous adenoma. BMI: Body Mass Index.

**Table 3. Multivariable logistic regression model predicting MA**

	OR	95%CI	P value
Escherichia-Shigella	53.254	3.338 to 849.676	0.005*
Acinetobacter	0.026	0.001 to 0.477	0.014*
BMI	1.684	0.993 to 2.855	0.053

\*P<0.05. MA: metachronous adenoma. BMI: Body Mass Index. The final model had good calibration (goodness of fit p=0.936) and discrimination (area under the receiver operating characteristic (ROC) curve=0.935).

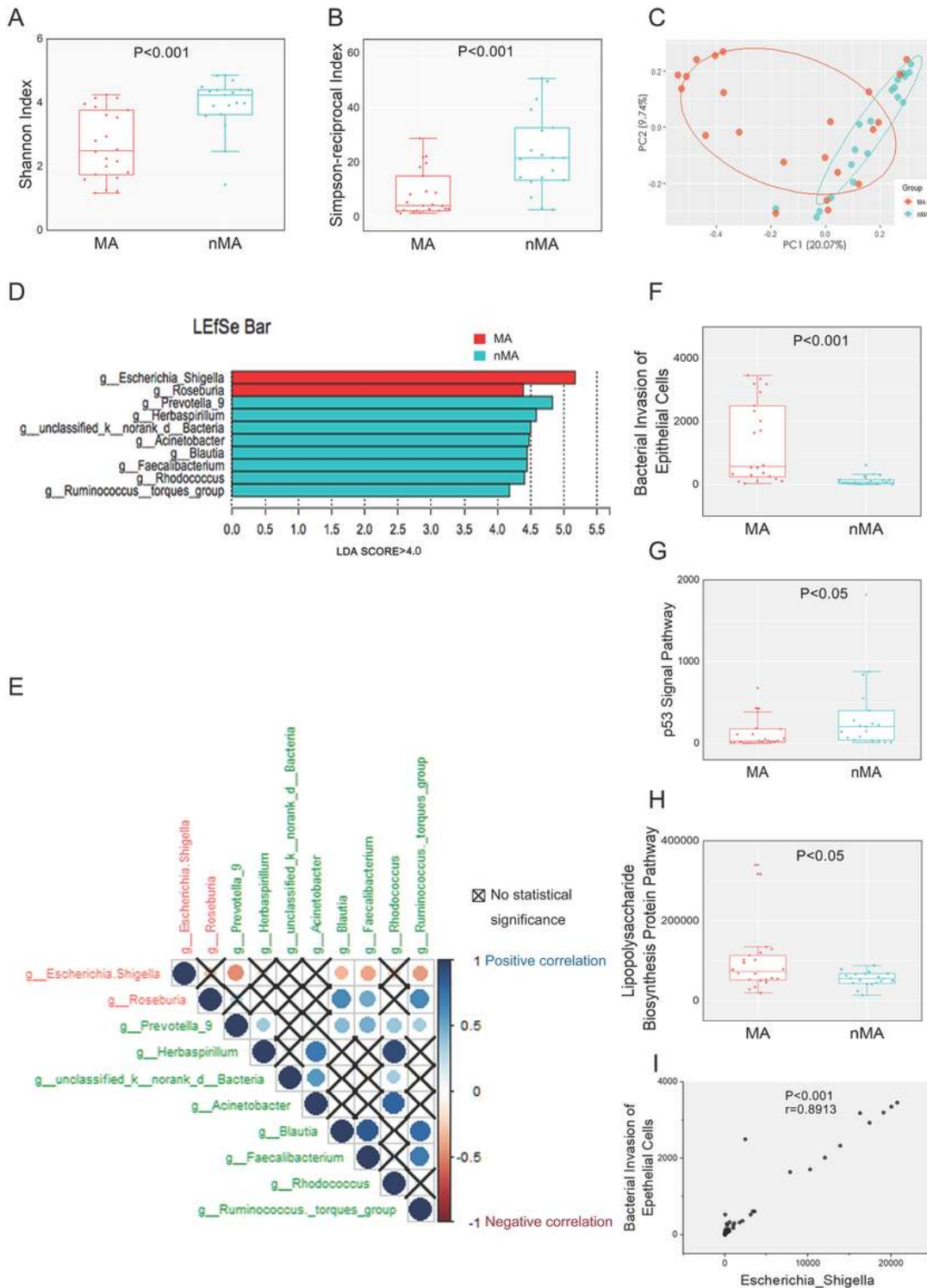
**Table 4. Sensitivity, specificity, PPV and NPV of the risk score based on predominant presence of the risk factors.**

Risk score	Sensitivity (%)	Specificity (%)	PPV	NPV	MA rate*
0	100	0	53.7	/	0(0/6)
1	100	31.6	62.9	100	15.38(2/13)
2	90.9	89.5	90.9	89.5	83.3(10/12)
3	45.5	100	100	61.3	100(5/5)
4	22.7	100	100	52.8	100(5/5)

\*The percentage of patients developing MA after 1 year of follow-up (number of patients showing MA/total number of patients, for each risk score).

NPV, negative predictive value; PPV, positive predictive value.

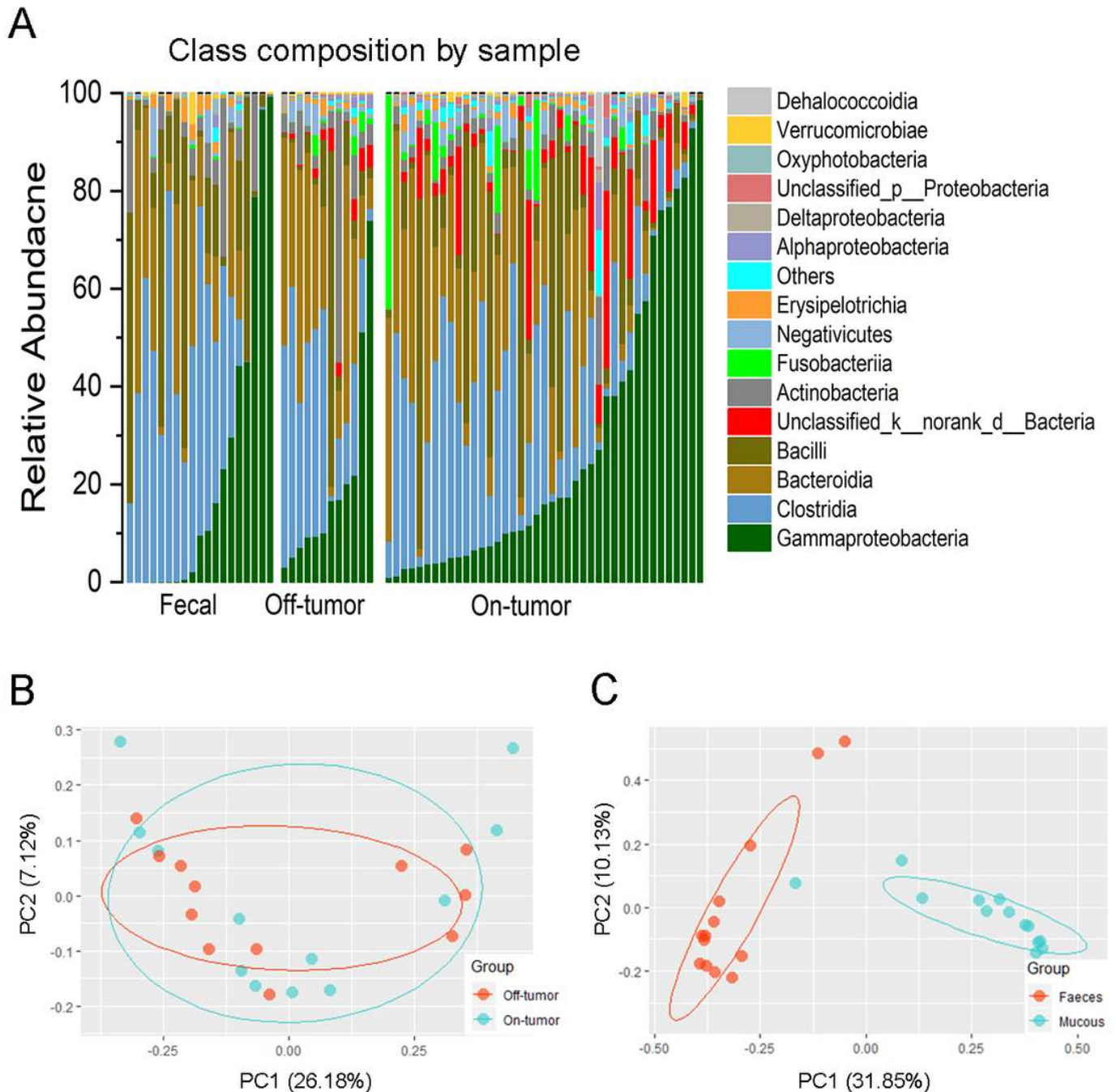
## Figures



**Figure 1**

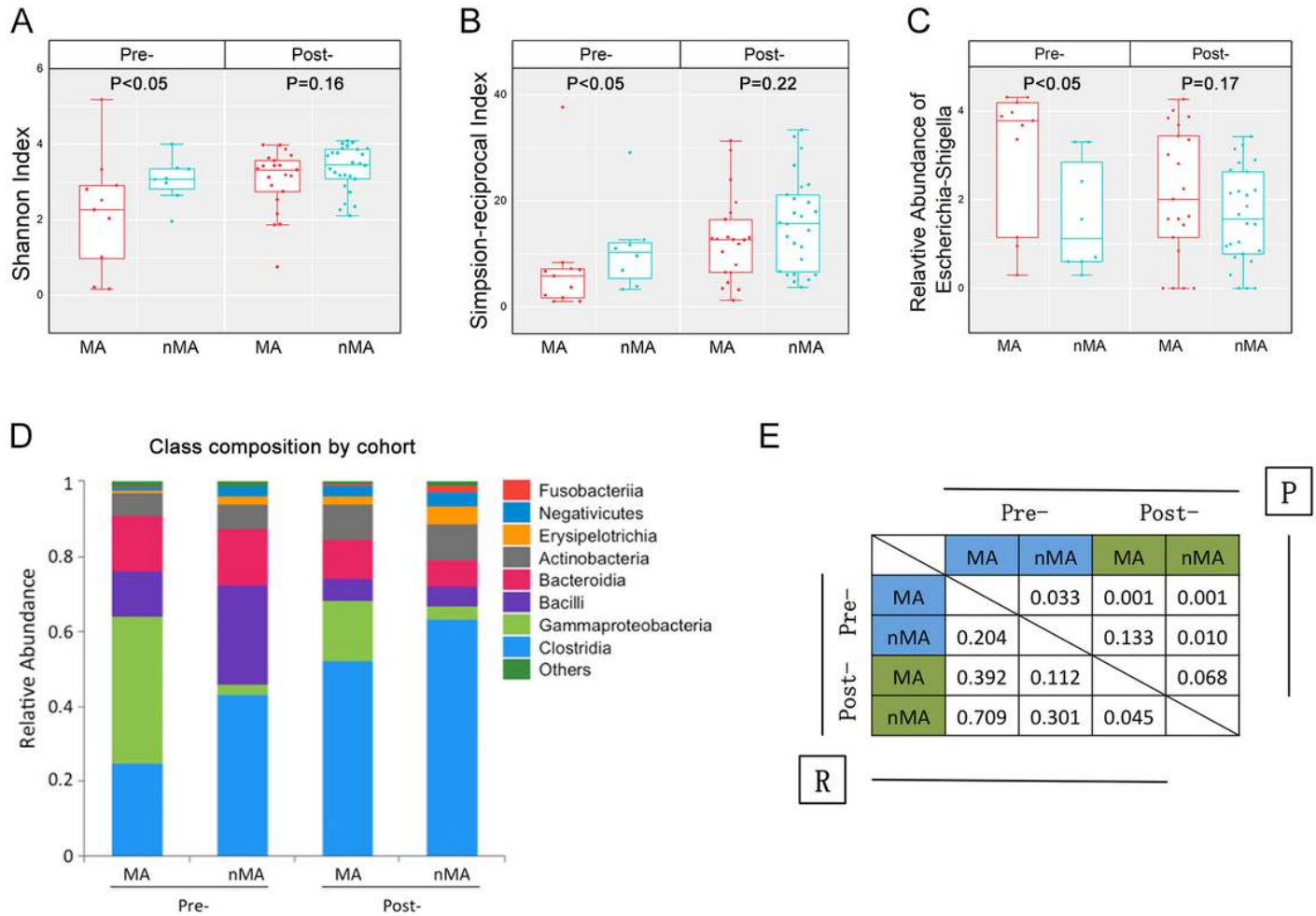
Mucosal microbiome diversity and communities are significantly different between MA and nMA. (A-B) Alpha diversity boxplot (Shannon and Simpson reciprocal) of mucosal samples in MA and nMA groups. Boxes represented the 25 to 75th percentile of the distribution; the median was shown as a thick line in the middle of the box; whiskers extend to values with 1.5 times the difference between the 25th and 75th percentiles. (C) PCoA using Bary-Curtis of beta diversity in MA and nMA groups. (D) LDA score computed

from features differentially abundant between MA and nMA in mucosal samples. The criteria for feature selection was log LDA score > 4. (E) Spearman correlations among 2 MA enriched (red) and 8 nMA enriched (green) genus taxa in mucosal samples of CRC patients. Red dots indicated negative correlation, blue dots indicated positive correlation, cross indicated no significance ( $P > 0.05$ ). (F-H) Boxplot of bacterial invasion of epithelial cells pathway, Lipopolysaccharide biosynthesis protein pathway and p53 signal pathway between MA and nMA. p values were adjusted using the FDR correction. (J) Spearman correlation between bacterial invasion of epithelial cells pathway and relative abundance of *Escherichia-Shigella*.



**Figure 2**

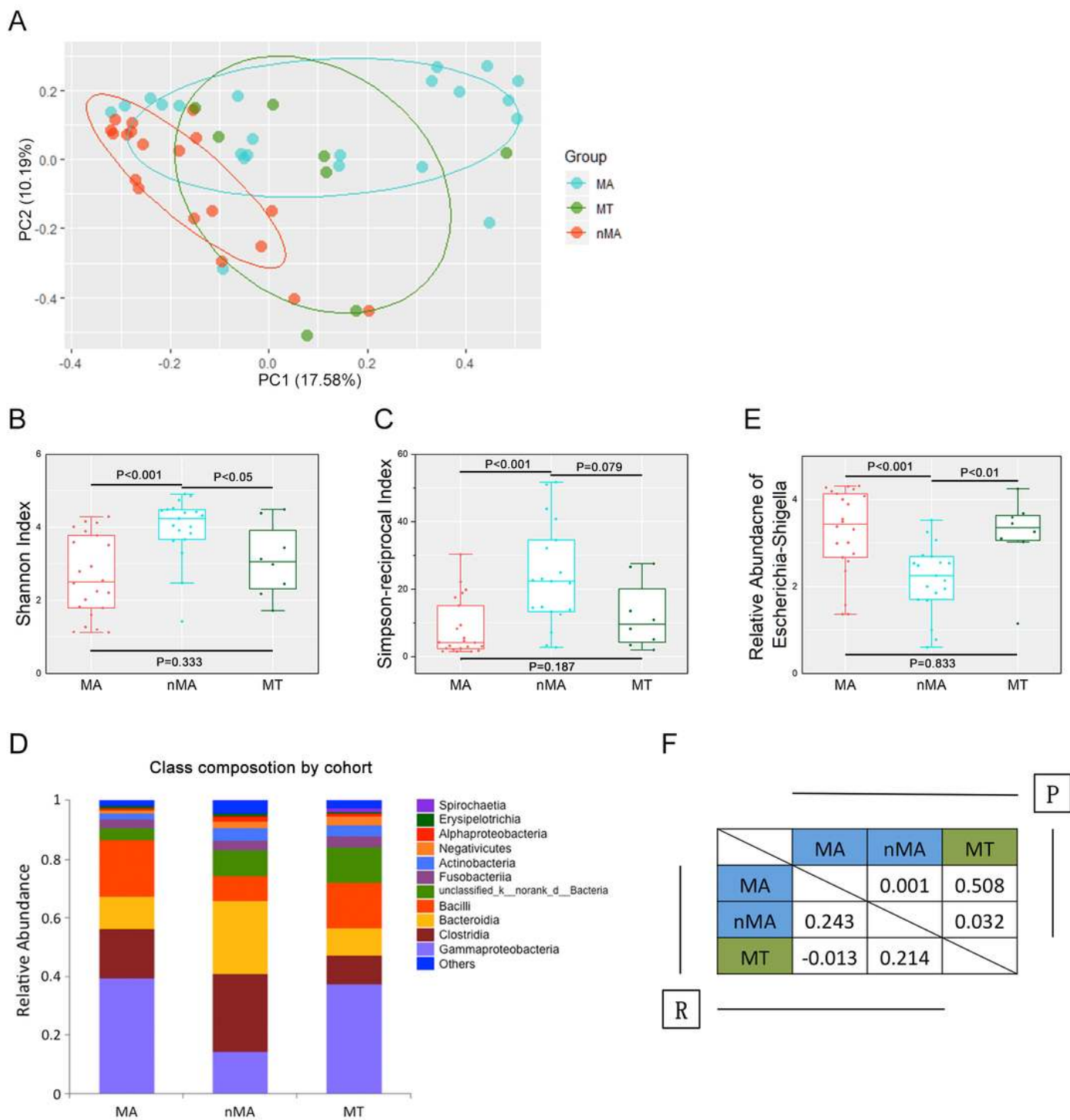
Fecal and off-tumor samples. (A) Bar plots of the class taxonomic levels of microbiota in Fecal, Off-tumor and On-tumor samples. Relative abundance is plotted for each samples. (B) PCoA using Bary-Curtis of beta diversity between On and Off tumor mucosal samples. (C) PCoA using Bary-Curtis of beta diversity between fecal and mucosal samples.



**Figure 3**

Fecal microbiota in CRC patients and CRC patients after surgical therapy. (A) Alpha diversity boxplot (Shannon and Simpson reciprocal) of fecal samples. (B) Box plots of relative abundance of fecal Escherichia-Shigella, box plot illustration was provided in Fig. 1. (C) Bar plots of the class taxonomic levels of fecal microbiota. Relative abundance is plotted for each group. (D) ANOSIM result between fecal samples of groups. R value indicated the strength of the factors on the samples, while give P value indicated the significance levels

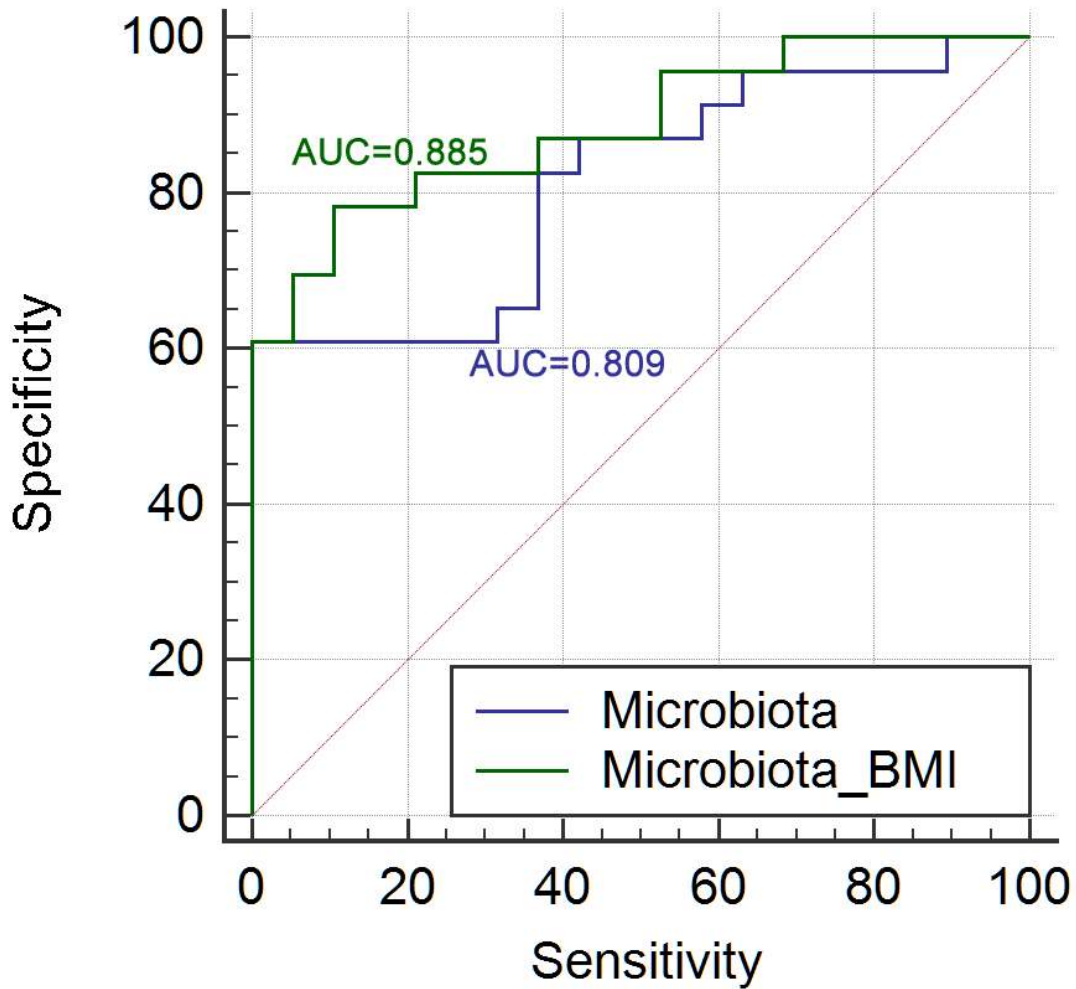




**Figure 4**

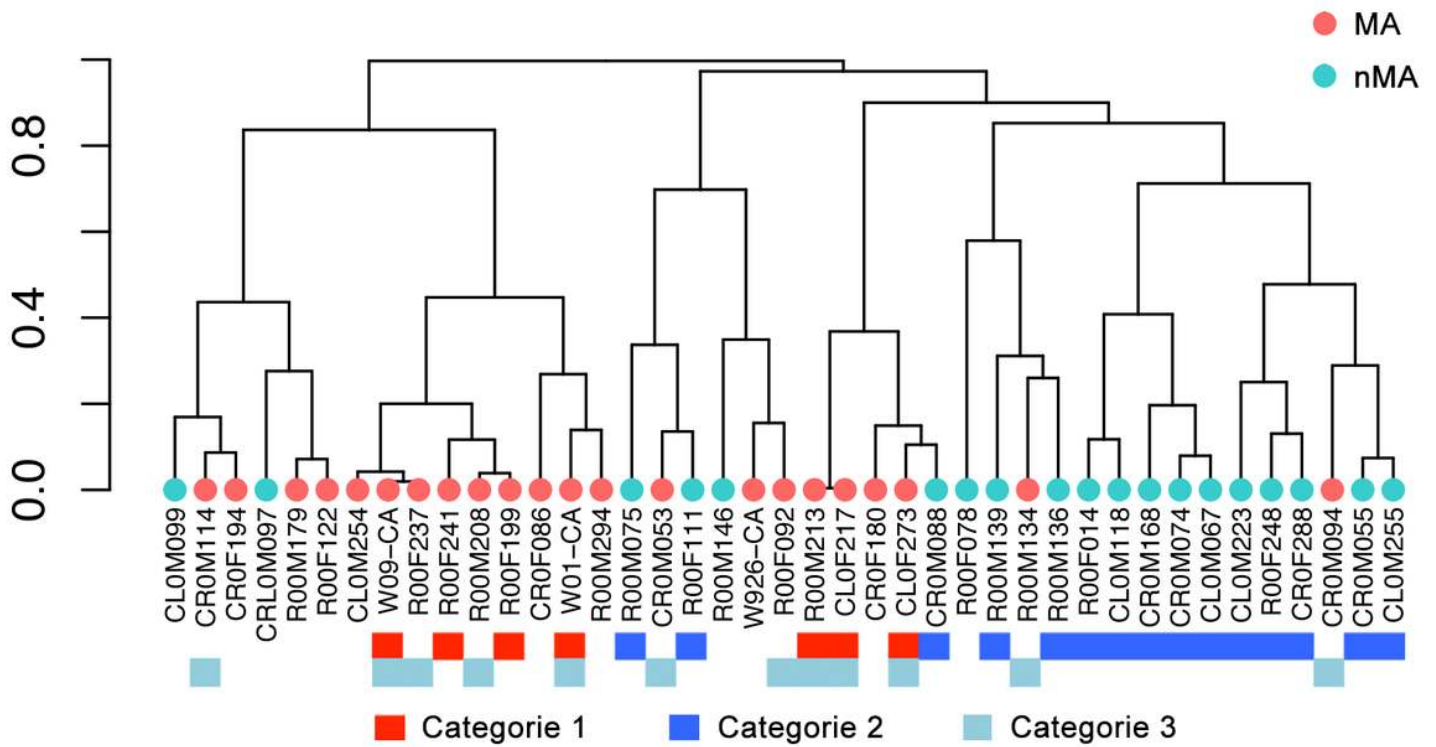
Mucosal microbiota in MA, nMA and MT patients. (A) Principal coordinate analysis (PCoA) using Bary-Curtis of beta diversity between MA, nMA and MT mucosal samples. (B-C) Alpha diversity boxplot (Shannon and Simpson reciprocal) between MA, nMA and MT mucosal samples. (D) Box plots of relative abundance of fecal *Escherichia-Shigella* abundance between MA, nMA and MT mucosal samples, box plot illustration was provided in Fig. 1. (E) Bar plots of the class taxonomic levels of mucosal microbiota

between MA, nMA and MT mucosal samples. Relative abundance is plotted for each group. (F) ANOSIM result between MA, nMA and MT mucosal samples. Meaning of R and P value were provided in Figure 3.



**Figure 5**

Gut microbiota signature can be used to discriminate between MA patients from nMA patients. ROC analysis of microbiota based RF model with (green) and without (blue) BMI.



**Figure 6**

MA can be predicted by combination of four variables. Hierarchical clustering for 41 participants with MA (red dots) and nMA (blue dots) using Bary-Curtis distance, based on absence or presence of *Escherichia-Shigella*, high BMI, Synchronous-adenoma and *Acinetobacter*. Three clusters were found and each cluster was represented a category.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarytable1.xlsx](#)
- [Supplementarytable23.docx](#)
- [SupplementaryFigures.pdf](#)
- [Rscript.docx](#)