Signatures within esophageal microbiota with progression of esophageal squamous cell carcinoma

Minjuan Li[1], Dantong Shao[1], Jiachen Zhou[2], Jianhua Gu[1], Junjie Qin[3], Wen Chen[4], Wenqiang Wei[1]

[1]National Central Cancer Registry, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China; [2]Department of Epidemiology and Biostatistics, School of Public Health, Xi'an Jiaotong University Health Science Center, Xi'an 710061, China; [3]Promegene Translational Research Institute, Shenzhen 518000, China; [4]Department of Cancer Epidemiology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

Correspondence to: Wenqiang Wei, PhD. National Central Cancer Registry, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China. Email: weiwq@cicams.ac.cn.

Abstract

Objective: Esophageal squamous cell carcinoma (ESCC) is one of the dominant malignances worldwide, but currently there is less focus on the microbiota with ESCC and its precancerous lesions.

Methods: Paired esophageal biopsy and swab specimens were obtained from 236 participants in Linzhou, China. Data from 16S ribosomal RNA gene sequencing were processed using quantitative insights into microbial ecology (QIIME2) and R Studio to evaluate differences. The Wilcoxon rank sum test and Kruskal-Wallis rank sum test were used to compare diversity and characteristic genera by specimens and participant groups. Ordinal logistic regression model was used to build microbiol prediction model.

Results: Microbial diversity was similar between biopsy and swab specimens, including operational taxonomic unit (OTU) numbers and Shannon index. There were variations and similarities of esophageal microbiota among different pathological characteristics of ESCC. Top 10 relative abundance genera in all groups include *Streptococcus*, *Prevotella*, *Veillonella*, *Actinobacillus*, *Haemophilus*, *Neisseria*, *Alloprevotella*, *Rothia*, *Gemella* and *Porphyromonas*. Genus *Streptococcus*, *Haemophilus*, *Neisseria* and *Porphyromonas* showed significantly difference in disease groups when compared to normal control, whereas *Streptococcus* showed an increasing tendency with the progression of ESCC and others showed a decreasing tendency. About models based on all combinations of characteristic genera, only taken *Streptococcus* and *Neisseria* into model, the prediction performance was the ideal one, of which the area under the curve (AUC) was 0.738.

Conclusions: Esophageal biopsy and swab specimens could yield similar microbial characterization. The combination of *Streptococcus* and *Neisseria* has the potential to predict the progression of ESCC, which is needed to confirm by large-scale, prospective cohort studies.

Keywords: Esophageal squamous cell carcinoma; precancerous lesions; 16S rRNA; *Streptococcus*; *Neisseria*

Introduction

Esophageal cancer (EC) is one of the most usual malignancies in the world, ranking as the ninth and fifth leading causes of global cancer related morbidity and death (1). The two main types of EC are esophageal

adenocarcinoma (EAC) and esophageal squamous cell carcinoma (ESCC), and the latter is the predominant histologic subtype in China. The overall 5-year survival rate of ESCC is about 30.3%, which is mostly due to late stage at diagnosis and low early diagnosis rate (2). China has about half of all ESCC cases on earth for its large population of China and high rates. Given the heavy ESCC burden in China and worldwide, it is necessary to develop and advance early detection, early diagnosis and early treatment.

The exact cause of ESCC, which is still unclear, is the powerful basis of effective prevention. As the emerging of microbiome project and high-throughput sequencing method, there are more and more studies about the role of human microbiome in health and disease, which reveals the microbial characterization in carcinogenesis (3). Millions of microorganisms colonizing the gastrointestinal tract are vital to the health of the host and play a role in development, nutrition, immunization and multi-biological process (4-6). Once the microbiota get into unbalanced, alteration may occur to the microenvironment and then lead to carcinoma (7). Several studies (8-10) have showed the microbial diversity and composition of upper gastrointestinal tract (UGI), showing a possible association between the microbiota and carcinoma. And the richness, diversity, and exact composition of microbiota in different organs might contribute to the development of ESCC. However, compared with gastric diseases, less attention was paid to esophageal microbiota and ESCC let alone precancerous lesions.

To gradually explore the microbial signature of esophagus with the progression of ESCC, two key points are worthy of consideration. One point is the sampling method. About ESCC or precancerous lesions, esophageal biopsy is invasive but pathological diagnosis on esophageal biopsy is the "Gold standard". And mucosal swab sampling seems to be less invasive and with higher performance but not "Gold standard" (11,12). Consequently, biopsy (13,14) and swab specimens (15-17) are the main collection methods for esophageal tissues. However, more appropriate sampling methods are still to be approved for various esophageal diseases. Another point is that diverse and dynamic microbial characterization concerning the progression of ESCC is still unclear, including precancerous disease, precancerous lesions and ESCC. To answer the above two points, paired esophageal biopsy and swab specimens were collected from normal participants and participants with precancerous diseases, precancerous lesions and ESCC.

Materials and methods

Study participants

This study was based on the endoscopy screening of UGI cancer project in China. We retrospectively recruited 236 participants aged 40−69 years with pathologic diagnosis of normal, esophagitis, low-grade intraepithelial neoplasia (LGIN), high-grade intraepithelial neoplasia (HGIN) and ESCC at the Linzhou Cancer Hospital in Henan province in China. All participants were local residents who were screened for UGI cancer in August to September 2018 or January to March 2019. All participants in this study were clearly informed and signed written informed consent. This study was approved and overseen by the Institutional Review Board of the Cancer Hospital of Chinese Academy of Medical Sciences. Participant information, including participant's sex, age, smoking and alcohol drinking status, dietary habit, history of digestive system disease and oral condition, was collected by trained staff.

Sample collection

Paired biopsy and swab specimens from each participant were obtained from 236 participants under upper gastroenterology endoscopic examination during screening of UGI cancer. The mucosal specimens were collected using sterile swabs (Puritan, sterile polyester tipped applicators) prior to biopsies to prevent contamination of mucosal specimens with blood. If there were lesions, 5 looping brushes were taken at the lesion, or else 5 looping brushes were taken at the middle esophagus. After swab specimens were collected, the swab head was cut off into sterile tubes (Cryovial, cryogenic tube 3.0 mL) including 1.5 mL of cell preserving fluid (Hologic, ThinPrep, PreservCyt Solution, San Diego, USA). Biopsy specimens at the brush sampling site were macro-dissected by an experienced endoscopy doctor with sterile forceps under strict procedures, which were then placed into a sterile tube. All paired specimens were stored at −80 °C immediately after sampling and shipped to the laboratory on dry ice.

DNA extraction, amplification and sequencing

Esophageal biopsy and swab specimens were macro-collected and used for DNA extraction by the QIAGEN

PowerSoil DNA Isolation Kit 12888100 (QIAGEN, Germany). The extracted DNA was stored at −80 °C in Tris-EDTA buffer solution prior to other processes. To control reagent contamination, we included water as a negative control without DNA template during specimen processing. The V4 region of the 16S ribosomal RNA (rRNA) gene was amplified using the universal bacterial primer set to generate amplicons. The primer pairs for DNA amplification covering variable V4 region were generated using the following primers: 515 F (5'-GTGYCAG−CMGCCGCGGTAA-3') and 806 R (5'-GGACTACNVGG−GTWTCTAAT-3') (18,19), incorporating the barcode sequences for 16S rRNA gene. Polymerase chain reaction (PCR) mixtures contained 1 μL of forward and reverse primer (10 μmol/L), 1 μL of template DNA, 4 μL of deoxyribonucleoside-triphosphates (dNTPs) (2.5 mmol/L), 5 μL of 10× EasyPfu Buffer, 1 μL of EasyPfu DNA Polymerase (2.5 U/μL) and 1 μL of double distilled water into a 50 μL total reaction volume. The PCR thermal cycling consisted of an initial denaturation step at 95 °C for 5 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 60 °C for 30 s, and extension at 72 °C for 40 s, with a final extension step at 72 °C for 4 min. Amplicons from each specimen were run on an agarose gel to ensure consistent sequencing length. Expected band size for 515F to 806R is 300−350 bp. Amplicon quantification was performed using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific/Invitrogen catalog no. Q32854, Waltham, USA). The amplicon library for high-throughput sequencing was combined at equal concentrations and volumes and subsequently quantified (KAPA Library Quantification Kit KK4824, USA). Using Illumina V4 chemistry and paired-end 2×150 bp reads, sequencing was performed on the Illumina MiniSeq platform. All sequencing was performed in a single MiniSeq run. Original sequence data processing was performed by the Illumina MiniSeq Reporter to remove adapter and primer sequences and then sequence data were exported in the FASTQ format.

Sequence processing and taxonomic classification

Except one biopsy in esophagitis and HGIN, 470 specimens of the total 472 specimens collected were successfully amplified and sequenced. Sequencing data were processed using the Quantitative Insights into Microbial Ecology (QIIME2, https://qiime2.org/) platform (20). Raw sequences were under strict quality control and feature table construction using the Divisive Amplicon

Denoising Algorithm 2 (DADA2) algorithm (21). Possible phiX reads and chimeric sequences were removed, and the remaining reads were truncated from 0 to 140 bp (for both forward and reverse reads) to avoid including sequencing errors at the ends of the reads. Paired-end reads were matched at a maximum mismatch parameter of six bases, which indicates a minimum similarity threshold of 90% for the overlap of the forward and reverse reads. The representative sequences (named "features" in QIIME2 nomenclature) were then generated by removing the redundant and low occurrence (n<5 within all samples) sequences. Data from all samples were rarefied to 1,000 reads for both diversity and relative abundance to avoid bias due to different sampling depths. The taxonomic assignment of the sequence variants (99% similarity) was assigned using the trained Naive Bayes classifier [trained on the Greengenes 13_8 (22)] through the q2-feature-classifier plugin, and the taxonomic composition at the phylum and genus level was generated based on operational taxonomic units (OTU) annotation.

Statistical analysis

All statistical analyses were performed with software program R Studio (Version 1.1.456; R Foundation for Statistical Computing, Vienna, Austria). Differences in observed OTUs and Shannon index were analyzed between paired esophageal biopsy and swab specimens in each participant group by the Wilcoxon rank-sum test. And the observed OTUs and Shannon index among five participant groups were compared by the Kruskal-Wallis test in esophageal biopsy and swab specimens, respectively. Principal coordinates analysis (PCoA) and the Adonis test were used to find discrepancy among the independent β diversity matrices. The relative abundance was calculated at the phylum and genus level for each participant group in each type of specimens. High relative abundance (≥0.01) genera were compared between normal and other four participant groups by the Wilcoxon rank-sum test and were compared among five participant groups by the Kruskal-Wallis test in swab or tissue specimens, respectively. The relative abundance of significant genus was brought into ordinal logistic regression model to draw receiver operating characteristic (ROC) curve. Area under the curve (AUC), specificity, sensitivity and akaike information criterion (AIC) were calculated for each model. The 10-fold cross-validation was used as the internal validation method, and normalized mean square error (NMSE) was calculated. In

758

Li et al. Esophageal microbiota with esophageal cancer

this study, a conventional level of significance (P<0.05) was used for rejecting the null hypothesis.

Results

Baseline characteristics of participants

According to the pathological diagnosis, a total of 236 cases enrolled in this study were assigned to the normal group (70 cases), esophagitis group (all 70 cases were with mild esophagitis), LGIN group (62 mild dysplasia cases and 8 moderate dysplasia cases), HGIN group (9 severe dysplasia cases and 10 carcinoma *in situ* cases) and ESCC group (7 cases). The average age was significant different among groups (P<0.01) and there was an increasing tendency with the progression of ESCC (Mean age $_{normal}$ =51.89 years, Mean age $_{ESCC}$ =59.20 years). Education background was statistical different among groups (P<0.05). Different from more normal and esophagitis participants with secondary or above education, participants with LGIN, HGIN and ESCC were mostly in primary or below education. There were no significant differences among groups in other personal information, life style and dietary habits (*Table 1*).

Microbial diversity between two esophageal specimens

This study examined the microbial diversity of paired esophageal biopsy and swab specimens to explore the collection methods of the two types of esophageal samples. There was no statistical difference of microbial diversity between swab and biopsy specimens (OTUs, P=0.71; Shannon index, P=0.49) (*Table 2*). Similarly, by comparing paired specimens of participants in each group respectively, no statistical difference was shown both for OTUs and Shannon index. However, on the whole, the microbial diversity was higher in swab samples than in biopsy samples, including OTUs (Biopsy=116.00, Swab=115.50) and Shannon index (Biopsy=4.39, Swab=4.34) (*Table 2*). About β diversity, significant clustering was detected for the weighted UniFrac distance (P<0.05, *Figure 1A*) and unweighted UniFrac distance (P<0.05, *Figure 1B*).

Microbial diversity of five participant groups

On the basis of comparison between two sampling methods, this study examined the microbial diversity by five participant groups to explore the dynamic changing during progression of ESCC. Within all biopsy samples, no statistical difference was detected among five groups (OTUs, P=0.13; Shannon index, P=0.09) but there was a

decreasing tendency of OTUs from normal to ESCC. Within all swab samples, there was significant difference of OTUs (P<0.01) among groups but not Shannon index (P=0.26) (*Figure 2*). And partly similar to OTUs of biopsy samples, there was an inconspicuous decreasing of observed OTUs in swab samples from normal to HGIN and an inverse increasing tendency to ESCC. About β diversity, significant clustering was detected for the weighted UniFrac distance (P<0.05, *Figure 1C*) and unweighted UniFrac distance (P<0.05, *Figure 1D*).

Characteristic phylum and genus with progression of ESCC

The top 5 phyla in biopsy and swab specimens include *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, *Actinobacteria* and *Fusobacteria* (*Figure 3*). No significant statistical difference was discovered for any dominant phylum among groups in both collections. In swab specimens, there was statistical difference of *Firmicutes* and *Fusobacteria* in ESCC comparing with the normal group. And in both biopsy and swab specimens, there was a decreasing tendency of *Firmicutes* with the progression of ESCC, while the changing tendency of *Fusobacteria* was opposite. At genus level, the top 10 genera in biopsy and swab specimens include *Streptococcus*, *Prevotella*, *Veillonella*, *Actinobacillus*, *Haemophilus*, *Neisseria*, *Alloprevotella*, *Rothia*, *Gemella* and *Porphyromonas* (*Figure 3*). In biopsy samples, *Neisseria* was significant different among groups with an increasing tendency from normal to ESCC. Besides, *Streptococcus* (Normal=0.1951, ESCC=0.0878, P<0.05), *Haemophilus* (Normal=0.0541, LGIN=0.0675, P<0.05) and *Porphyromonas* (Normal=0.0090, Esophagitis=0.0093, P<0.05) were meaningful in biopsy specimens. In swab samples, *Streptococcus* and *Neisseria* were significant different among groups, with a decreasing and increasing tendency from normal to ESCC, respectively. And there was statistical difference of *Haemophilus* in swab (Normal=0.0487, ESCC=0.0881, P<0.05) (*Figure 4*). The relative abundance of *Prevotella* (P<0.05), *Neisseria* (P<0.05), *Rothia* (P<0.01) and *Porphyromonas* (P<0.05) was higher in swab than in tissues of normal group. In LGIN group, *Prevotella* (P<0.001), *Veillonella* (P<0.05) and *Rothia* (P<0.05) were all higher in swab than in tissues but *Gemella* (P<0.05) was opposite.

Microbial prediction model of ESCC and its precancerous lesions

According to the results of characteristic genera with the

Table 1 Baseline characteristics of participants

Variables	n (%)						P
	Normal (N=70)	Esophagitis (N=70)	LGIN (N=70)	HGIN (N=19)	ESCC (N=7)	Total (N=236)	
Age (year) ($\bar{x}\pm s$)	51.89±7.66	54.80±7.16	56.21±6.43	56.89±7.15	59.29±6.13	54.66±7.30	<0.01
Male	21 (30.00)	29 (41.43)	24 (34.29)	11 (57.89)	4 (57.14)	89 (37.71)	0.14
Married	65 (92.86)	61 (87.14)	59 (84.29)	16 (84.21)	7 (100)	208 (88.14)	0.45
Average annual income <50,000 CNY	48 (68.58)	50 (71.43)	49 (70.00)	14 (73.68)	5 (71.43)	166 (70.34)	1.00
Current smoking*	9 (12.86)	14 (20.00)	8 (11.43)	6 (33.33)	3 (42.86)	40 (17.02)	0.05
Current drinking*	7 (10.00)	9 (12.86)	7 (10.00)	1 (5.56)	1 (14.29)	25 (10.64)	0.87
Family history of cancer	45 (64.29)	44 (62.86)	46 (65.71)	14 (73.68)	3 (42.86)	152 (64.41)	0.70
Education							<0.05
Primary school or below	31 (44.29)	20 (28.57)	43 (61.43)	10 (52.63)	4 (57.14)	108 (45.76)	
Secondary education or more	39 (55.71)	50 (71.43)	27 (38.57)	9 (47.37)	3 (42.86)	128 (54.24)	
Teeth loss*							0.12
≤3 teeth	36 (51.43)	26 (37.14)	30 (42.86)	6 (33.33)	1 (14.29)	99 (42.13)	
≤11 teeth	23 (32.86)	20 (28.57)	17 (24.29)	6 (33.33)	2 (28.57)	68 (28.94)	
≥12 teeth	11 (15.71)	24 (34.29)	23 (32.86)	6 (33.33)	4 (57.14)	68 (28.94)	
Rice (1−3 d/week)*,#	53 (75.71)	55 (78.57)	55 (78.57)	11 (61.11)	5 (71.43)	179 (76.17)	0.57
Wheaten food (4−6 d/week)*,#	47 (67.14)	50 (71.43)	47 (67.14)	10 (55.56)	2 (28.57)	156 (66.38)	0.19
Coarse grains (Every day)*,#	50 (71.43)	57 (81.43)	56 (80.00)	10 (55.56)	4 (57.14)	177 (75.32)	0.09
Meat (1−3 d/month)#	30 (42.86)	24 (34.29)	22 (31.43)	8 (42.11)	4 (57.14)	88 (37.29)	0.45
Poultry (Never)#	69 (98.57)	68 (97.14)	70 (100)	19 (100)	7 (100)	233 (98.73)	0.84
Sea food (Never)#	70 (100)	68 (97.14)	70 (100)	19 (100)	7 (100)	234 (99.15)	0.47
Eggs (1−3 d/week)#	29 (41.43)	25 (35.71)	25 (35.71)	10 (52.63)	2 (28.57)	91 (38.56)	0.64
Vegetables (Every day)#	66 (94.29)	69 (98.57)	70 (100)	18 (94.74)	7 (100)	230 (97.46)	0.18
Fruits (1−3 d/month)#	31 (44.29)	27 (38.57)	22 (31.43)	9 (47.37)	5 (71.43)	94 (39.83)	0.20
Bean products (1−3 d/week)#	36 (51.43)	35 (50.00)	39 (55.71)	10 (52.63)	5 (71.43)	125 (52.97)	0.85
Salty degree (Medium)*,#	51 (72.86)	53 (75.71)	54 (77.14)	15 (83.33)	4 (57.14)	177 (75.32)	0.70
Spicy food (Never)*,#	55 (78.57)	53 (75.71)	59 (84.29)	17 (94.44)	5 (71.43)	189 (80.43)	0.70
Hot food (Never)#	68 (97.14)	69 (98.57)	66 (94.29)	18 (94.74)	6 (85.71)	227 (96.19)	0.24

*, Due to the lack of information on this variable in one participant with HGIN, the number of participants with HGIN and all participants are 18 and 235 respectively in this variable; #, All diet questions except spicy food are about the dietary frequency in the past year. Spicy food is about the dietary frequency in the past month. Salty degree includes salty, medium and light. Frequency options include every day, 4−6 d/week, 1−3 d/week, 1−3 d/month and never. LGIN, low-grade intraepithelial neoplasia; HGIN, high-grade intraepithelial neoplasia; ESCC, esophageal squamous cell carcinoma.

progression of ESCC in two specimens, *Streptococcus*, *Neisseria*, *Haemophilus* and *Porphyromonas* were taken into the microbial prediction model of ESCC and its precancerous lesions. To explore a better one, this study carried out all combinations of the above four characteristic genera. Only taken one genus into model, the AUC from high to low was as follows: 0.732 (*Neisseria*), 0.712 (*Haemophilus*), 0.697 (*Streptococcus*) and 0.696 (*Porphyromonas*). About 6 combinations of any two genera, the AUC of *Neisseria* and *Streptococcus* was the best

(0.738), and the AUC of *Neisseria* and *Porphyromonas* or *Haemophilus* was 0.733 and 0.732, respectively. Bring *Neisseria*, *Streptococcus* and *Porphyromonas* or *Haemophilus* into model, the AUC was the same (0.738) and higher than the other two combinations of three genera. Taken all genera into model, the AUC was 0.738. Of all the combinations, the AUC of four combinations was the highest and same, which was 0.738 (*Figure 5*). Synthesizing several evaluation indexes of the four models, including sensitivity, specificity and AIC, the combination of

Table 2 α diversity between esophageal biopsy and swab specimen

Variables	Observed OTUs			Shannon index		
	Biopsy	Swab	P	Biopsy	Swab	P
Normal	122.00	125.00	0.34	4.63	4.48	0.15
Esophagitis	119.00	112.00	0.48	4.62	4.33	0.35
LGIN	112.50	113.50	0.36	4.09	4.34	0.11
HGIN	112.00	103.00	0.24	4.25	3.96	0.24
ESCC	95.00	114.00	0.80	4.58	4.78	0.71
Total	116.00	115.50	0.71	4.39	4.34	0.49

The number in the column of biopsy and swab means the median of α diversity in each group. OTU, operational taxonomic unit; LGIN, low-grade intraepithelial neoplasia; HGIN, high-grade intraepithelial neoplasia; ESCC, esophageal squamous cell carcinoma.
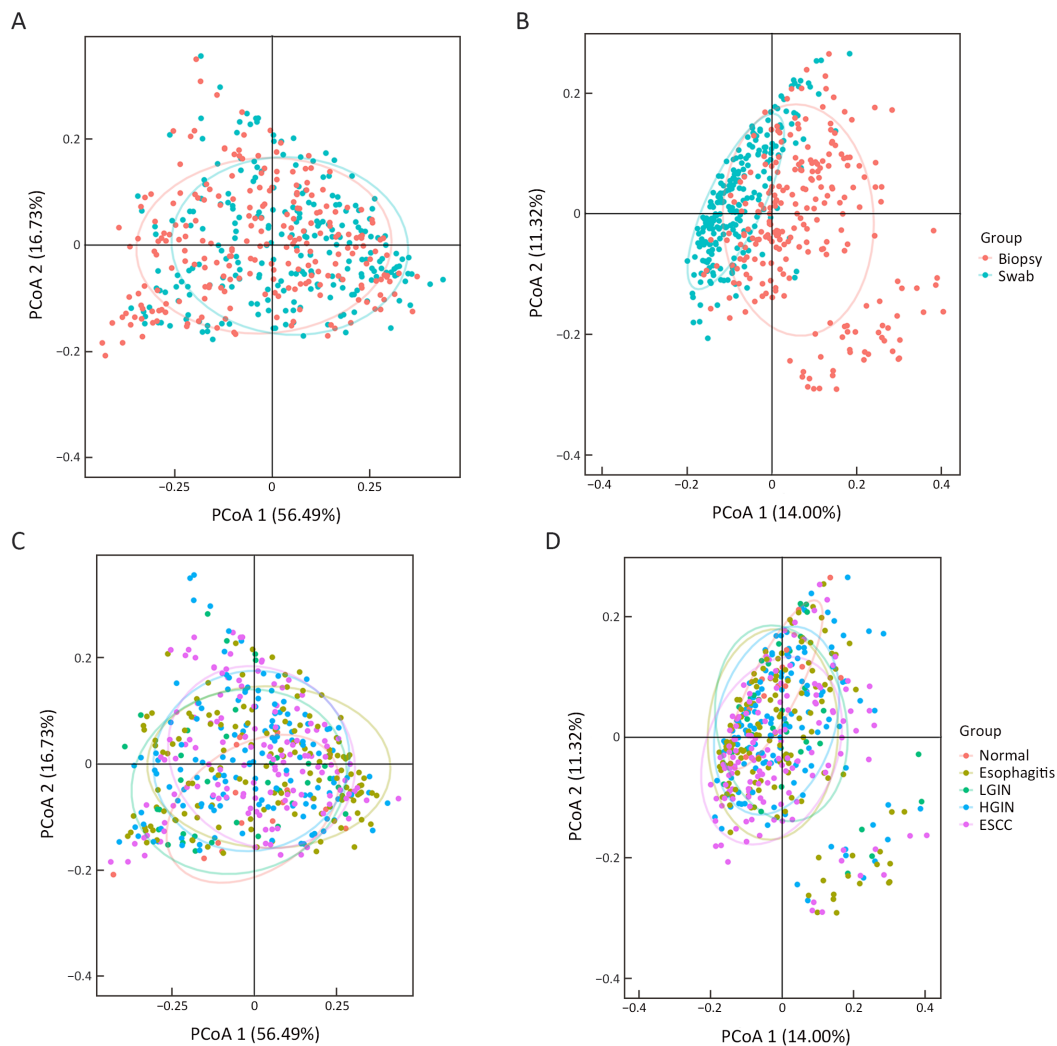
Figure 1 Microbial comparison for β diversity. β diversity based on (A) the weighted UniFrac distance ($P<0.05$) and (B) the unweighted UniFrac distance ($P<0.05$) was compared with PCoA for swab and biopsy specimens. β diversity based on (C) the weighted UniFrac distance ($P<0.05$) and (D) the unweighted UniFrac distance ($P<0.05$) was compared with PCoA for participant groups. PCoA, principal coordinates analysis; ESCC, esophageal squamous cell carcinoma; HGIN, high-grade intraepithelial neoplasia; LGIN, low-grade intraepithelial neoplasia.
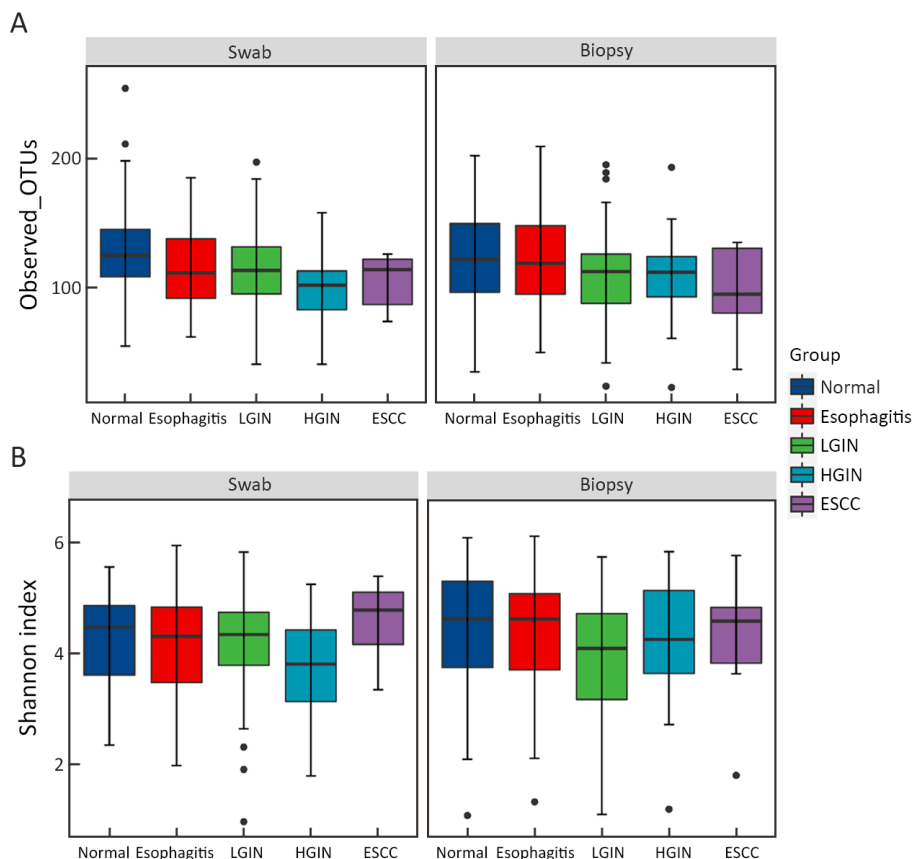
Figure 2 Comparison for α diversity of participants with different diagnosis in swab and biopsy specimens. α diversity was assessed by observed OTUs (Swab, P<0.01; Biopsy, P=0.13) (A) and Shannon index (Swab, P=0.26; Biopsy, P=0.09) (B). OTU, operational taxonomic unit; LGIN, low-grade intraepithelial neoplasia; HGIN, high-grade intraepithelial neoplasia; ESCC, esophageal squamous cell carcinoma.

Streptococcus and *Neisseria* was a more appropriate microbial prediction model of ESCC and its precancerous lesions. And we utilized the 10-fold cross-validation as the internal validation method, the NMSE was 1.575.

Discussion

The esophageal microbiota includes various passenger microbes transiting from the oral cavity to the stomach and resident bacteria closely associated with the esophageal epithelium. In this study, we have observed and compared the esophageal microbiota that was closely associated with the esophageal mucosa under distinct pathological conditions to investigate the role of esophageal microbiota in the development of ESCC.

Combining endoscopic iodine staining and indicative biopsy is the gold diagnosis standard for ESCC and squamous dysplasia in China. Given the invasion of endoscopic biopsy and emerging of cytology, to enrich and verify previous studies, we flexibly compared the microbial

composition of esophageal swab and biopsy specimens from normal, esophagitis, LGIN, HGIN to ESCC. In this study, α diversity was not significantly different between two specimens in any group, while mucosal swab specimens had a better performance than biopsy, which was similar to the findings conducted by Liu and Gall (11,23). Gall *et al.* (23) found a significant higher diversity in swab specimens compared with biopsies in esophagus. Another meaningful study (24) indicated that mucosal biopsies should remain the gold standard for microflora analysis in the esophagus, however, the swab specimens were from uvula and endoscope. Similar study in intestinal tract (12) revealed both swab and biopsy collection methods provided indifferent assessments of the microbial community and recommended mucosal brushing. Additionally, no significant difference was shown in microbial community in esophageal biopsy and swab specimens. Taken findings from this study and literatures together, both collection methods in microbial studies have pros and cons, which
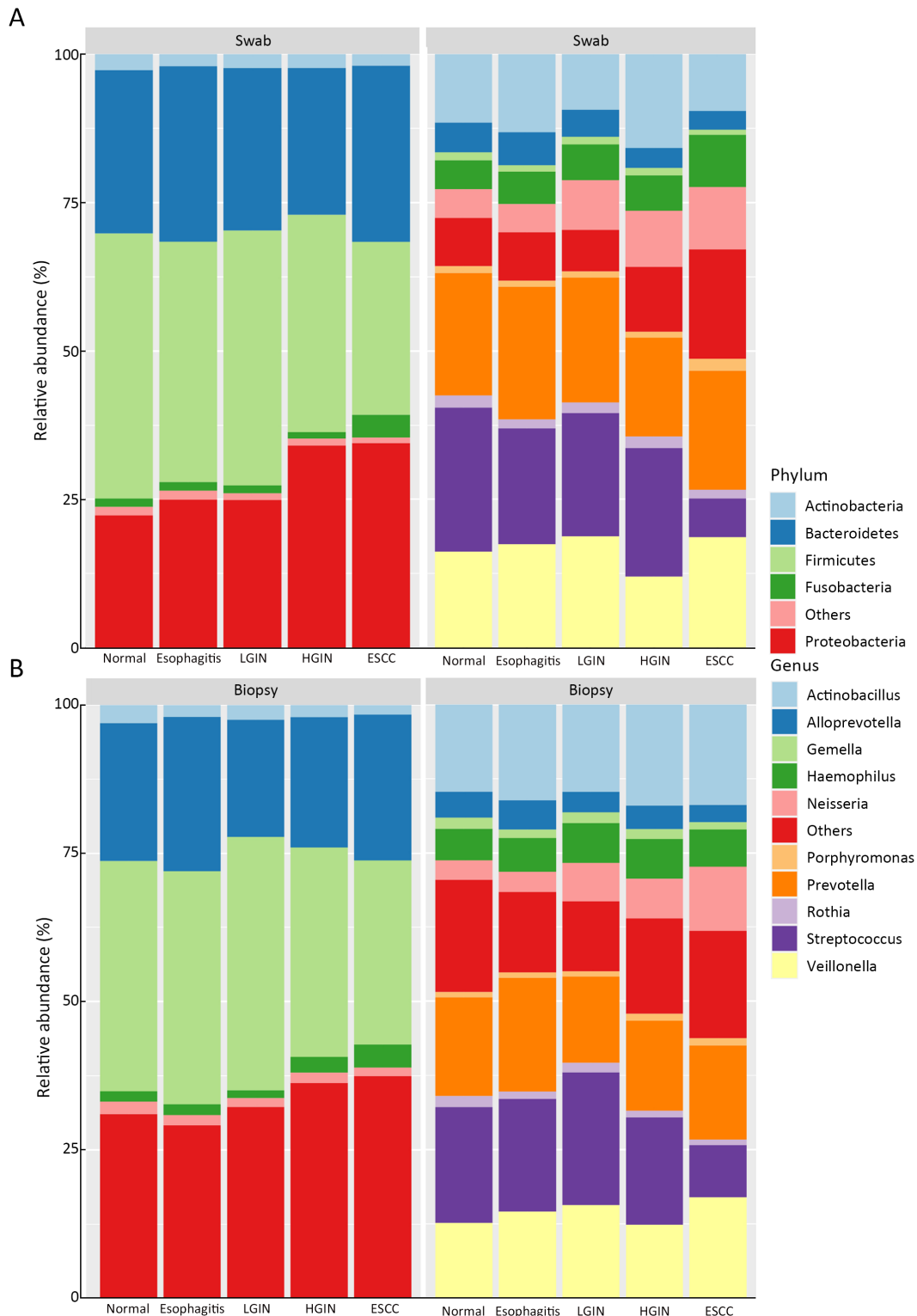
www.cjcrcn.org   *Chin J Cancer Res* 2020;32(6):755-767

Figure 3 Microbial composition at phylum and genus level for participants in different participant groups in swab and biopsy specimens respectively. (A) Microbial composition for all participants in swab specimens at the phylum and genus level; (B) Microbial composition for all participants in biopsy specimens at the phylum and genus level. LGIN, low-grade intraepithelial neoplasia; HGIN, high-grade intraepithelial neoplasia; ESCC, esophageal squamous cell carcinoma.
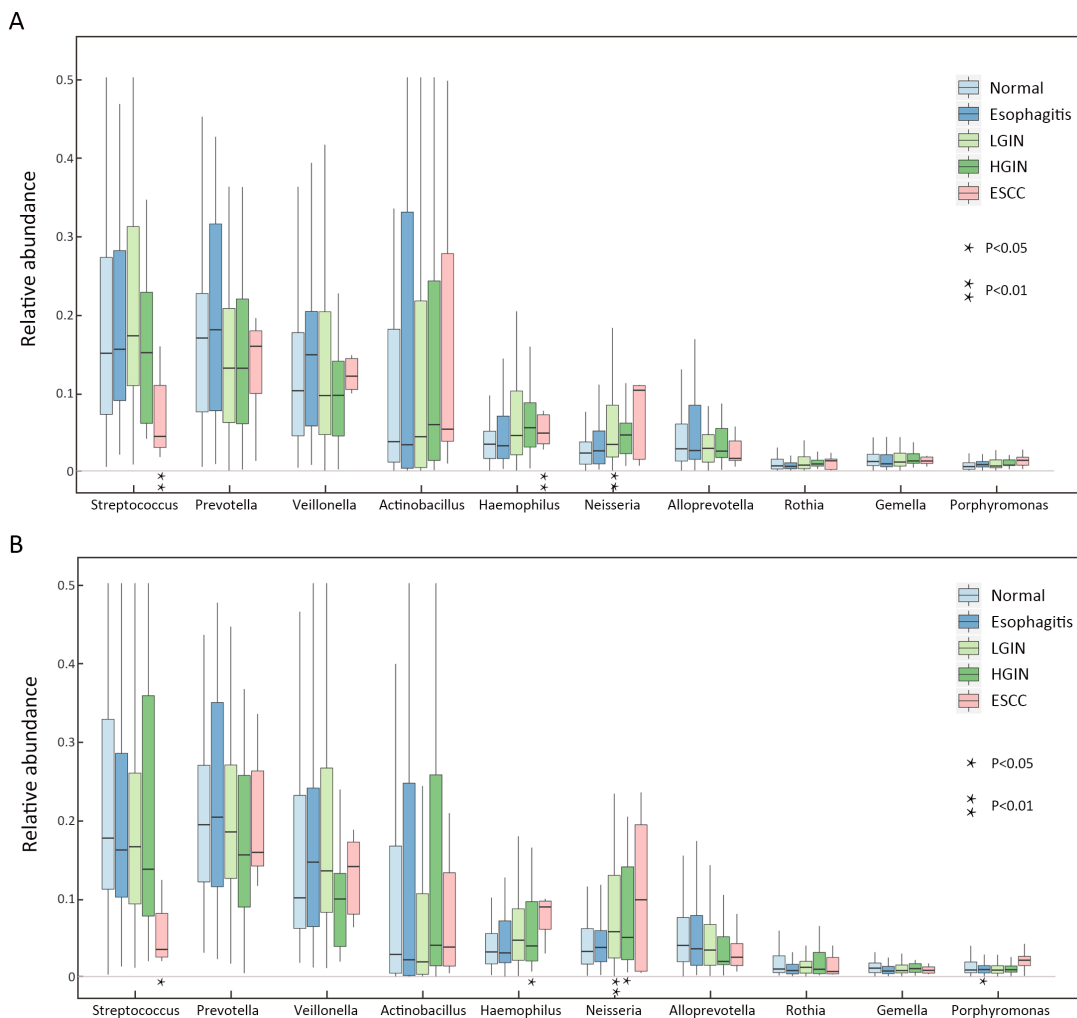
A

B

Figure 4 Comparison of relative abundance of microbial genera among participant groups. (A) Mean relative abundance of top 10 high-relative abundance genera in swab specimens; (B) Mean relative abundance of top 10 high-relative abundance genera in biopsy specimens. LGIN, low-grade intraepithelial neoplasia; HGIN, high-grade intraepithelial neoplasia; ESCC, esophageal squamous cell carcinoma; *, $P<0.05$; **, $P<0.01$.

means appropriate collection methods are needed to depend on situation.

High-relative abundance genera including *Streptococcus*, *Prevotella*, *Veillonella*, *Actinobacillus*, *Haemophilus*, *Neisseria*, *Alloprevotella*, *Rothia*, *Gemella* and *Porphyromonas* were commonly seen in all samples from all participant groups, which was not exactly similar to findings of Dong and Pei (13,25). About the stable dominant genera community, there were various changes in relative abundance not only between two specimens but among participant groups. Comparing with healthy control, individuals with esophageal dysplasia had a significantly lower number of genera in esophagus in Yu's finding (26). However, due to unbalanced sample size in

this study, the decreasing tendency was inconspicuous and insignificant comparing with that of Yu (26). Yang *et al.* (27) proposed there was a shift from gram-positive microbiota to gram-negative microbiota in the microenvironment of the histologically abnormal distal esophagus. Only *Streptococcus* and *Gemella* were gram-positive microorganisms among top 10 genera in our study. And *Gemella* was one of the high-relative abundance genera in UGI (28) but not mentioned in Dong's finding (25).

All esophageal samples were taken from middle or lower esophagus in this study, which had no influence on the microbial preference (25). Both in biopsy and swab specimens, there was no difference in α diversity between
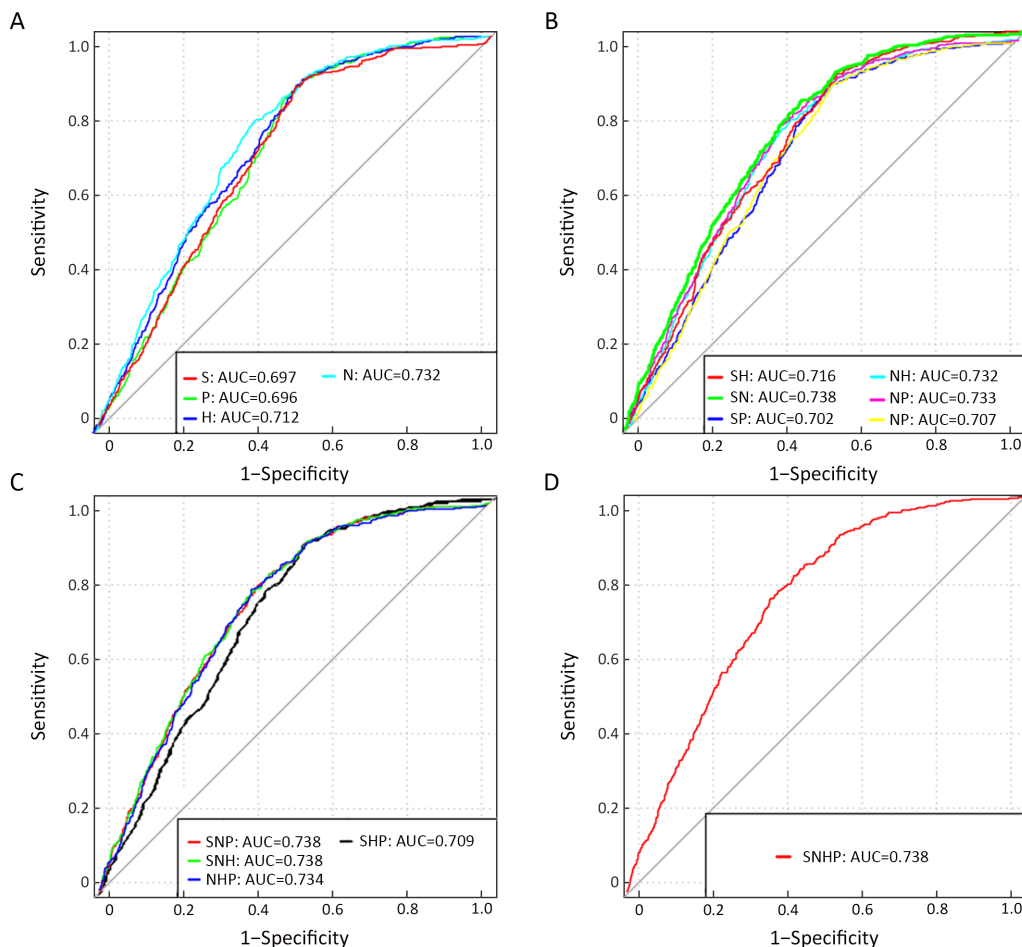
Figure 5 Microbial models based on characteristic genera. (A) Each characteristic genus was tested by a ROC analysis; (B) Combinations of any two genera were tested by ROC analysis; (C) Combinations of any three genera were tested by ROC analysis; (D) Combination of four genera was tested by ROC analysis. S, *Streptococcus*; N, *Neisseria*; P, *Porphyromonas*; H, *Haemophilus*; ROC, receiver operating characteristic; AUC, area under the curve.

normal control and the other groups but relative abundance was slightly fluctuating. A previous study about the UGI microbiota has shown the lower microbial richness was significantly associated with the presence of esophageal squamous dysplasia (26). Shao *et al.* (8) found there was no significant difference in α diversity for ESCC tumor and non-tumor tissues. Through various comparisons, the differences in *Streptococcus*, *Neisseria*, *Haemophilus* and *Porphyromonas* were detected in swab or tissue specimens. *Streptococcus*, gram-positive genus, was one of the dominant genera in health esophagus, while the other three gram-negative genera were more likely to be dominant in esophagitis and Barrett's esophagus (13,27). Among four significant genera in this paper, interestingly, only *Streptococcus* had a decreasing tendency from normal to ESCC, while the other genera had an increasing

tendency. The dynamic changing of *Streptococcus* in this study was a comprehensive complement of another microbial characteristic study that the relative abundance of *Streptococcus* decreased with more advanced ESCC (8). And this study firstly found there was a significant increasing tendency of *Haemophilus* and *Neisseria* with the progression of ESCC. Previously, *Haemophilus* or *Neisseria* enrichening in eosinophilic esophagitis (EoE) has been found in findings conducted by Harris and Benitez (29,30).

Taken several evaluation indexes into consideration, combining *Streptococcus* and *Neisseria* into microbial prediction model of ESCC and its precancerous lesions had an ideal outcome. Interesting, genus *Neisseria* of the phylum *Proteobacteria* was more abundant in the oral cavity than in esophagus, which was opposite to genus *Streptococcus* of the phylum *Firmicutes* (25). However, not

Neisseria but *Prevotella* has been discussed in limited bacterial model studies. A prognosis study of ESCC revealed that combined *Streptococcus* and *Prevotella* abundance was an independent prognosis biomarker in ESCC patients (31), which did not discuss about squamous dysplasia. Analysis of esophageal brush samples from EAC participants has shown distinct community cluster of *Streptococcus* and *Prevotella* (32). A study (8) about gastric cardia adenocarcinoma has shown the relative abundance of *Streptococcus* decreased and *Prevotella* increased with a more advanced tumor stage even these findings were not statistically significant. *Prevotella* was associated with carbohydrates and simple sugars (33). Nobel *et al*. (34) found low fiber intake was associated with increased relative abundance of several gram-negative bacteria, including *Prevotella*, *Neisseria* and so on. All participants in this study were from the same county and had a similar dietary pattern. And as the progression of ESCC, the diet was easier to be affected. Boldly postulating, participants with higher disease degree, dietary intake may be reduced, and the relative abundance of *Neisseria* associated with low fiber intake may be higher, which needs to be tested and verified in the future. And this study is an attempt to discuss the microbial model about the progression of ESCC. The AUC of *Streptococcus* and *Neisseria* in this study was 0.738 which was lower than that in the study of Hsieh *et al*. (10), but not comparable. Hsieh *et al*. (10) figured out increased abundance of *Clostridium* and *Fusobacterium*, which exhibited a diagnostic ability for gastric cancer, and the AUC was 0.875. And unlike *Streptococcus*, *Prevotella* and *Fusobacteria*, to our knowledge, this is one of the few highlighting the potential prediction of *Neisseria*, which enriches previous findings and provides new direction for future studies (10,31,35).

Due to difficulty in sampling, few studies have explored the signatures within esophageal microbiome with the progression of ESCC. And this is an attempt to dynamically depict the microbial composition covering the whole progression of ESCC. However, there are still several limitations in this study. First, all participants in this study were from the natural population in Linzhou. Due to the low detection rate of precancerous lesions and cancer in natural population and limited screening time, there was an unbalanced but reasonable sample size for each group from normal to ESCC. Second, all swab and biopsy samples were collected by the same technological process that swab specimens were preserved in cell preserving fluid. This study did not extract DNA from the cell preserving fluid to

further exclude the potential influence. Third, interaction among diet, microbiota and health is complicated and unclear (36). Age may contribute to but not fully explain the different esophageal microbial community types. For example, age is positively correlated with the relative abundance of *Streptococcus* (32). In this study, all individuals were with a similar dietary pattern, but there was an aging tendency in different groups as the progression of ESCC, which was hard to diminish or exclude this interaction.

Conclusions

Returning to the two points posed at the beginning of this study. Esophageal swab specimens from participants with different histopathologic diagnosis have similar microbial signatures as corresponding esophageal biopsy, which provides sampling suggestion for future microbial studies. With the progression of ESCC, dynamic changing of high-relative abundance genera was informative and complicated. *Streptococcus* and *Neisseria* may predict ESCC and its precancerous lesions but it is hard to ensure the causality between microbiota and ESCC progression. In the future, more attention needs to be paid to low-relative abundance genera to enrich and perfect conclusions found by this study. And more large-scale case-control studies and prospective cohort studies are needed to further explore the dynamic bacterial changing in ESCC progression.

Acknowledgements

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1.   Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.

2. Zeng H, Chen W, Zheng R, et al. Changing cancer survival in China during 2003-15: a pooled analysis of 17 population-based cancer registries. Lancet Glob Health 2018;6:e555-e67.

3. Vogtmann E, Goedert JJ. Epidemiologic studies of the human microbiome and cancer. Br J Cancer 2016;114:237-42.

4. Requena T, Martínez-Cuesta MC, Peláez C. Diet and microbiota linked in health and disease. Food Funct 2018;9:688-704.

5. Belkaid Y, Harrison OJ. Homeostatic immunity and the microbiota. Immunity 2017;46:562-76.

6. Dominguez-Bello MG, Godoy-Vitorino F, Knight R, et al. Role of the microbiome in human development. Gut 2019;68:1108-14.

7. Garrett WS. Cancer and the microbiota. Science 2015;348:80-6.

8. Shao D, Vogtmann E, Liu A, et al. Microbial characterization of esophageal squamous cell carcinoma and gastric cardia adenocarcinoma from a high-risk region of China. Cancer 2019;125:3993-4002.

9. Kageyama S, Takeshita T, Takeuchi K, et al. Characteristics of the salivary microbiota in patients with various digestive tract cancers. Front Microbiol 2019;10:1780.

10. Hsieh YY, Tung SY, Pan HY, et al. Increased abundance of clostridium and fusobacterium in gastric microbiota of patients with gastric cancer in Taiwan. Sci Rep 2018;8:158.

11. Liu AQ, Vogtmann E, Shao DT, et al. A comparison of biopsy and mucosal swab specimens for examining the microbiota of upper gastrointestinal carcinoma. Cancer Epidemiol Biomarkers Prev 2019;28:2030-7.

12. Huse SM, Young VB, Morrison HG, et al. Comparison of brush and biopsy sampling methods of the ileal pouch for assessment of mucosa-associated microbiota of human subjects. Microbiome 2014;2:5.

13. Pei Z, Bini EJ, Yang L, et al. Bacterial biota in the human distal esophagus. Proc Natl Acad Sci U S A 2004;101:4250-5.

14. Pei Z, Yang L, Peek RM, et al. Bacterial biota in reflux esophagitis and Barrett's esophagus. World J Gastroenterol 2005;11:7277-83.

15. Macfarlane S, Furrie E, Macfarlane GT, et al. Microbial colonization of the upper gastrointestinal tract in patients with Barrett's esophagus. Clin Infect Dis 2007;45:29-38.

16. Fillon SA, Harris JK, Wagner BD, et al. Novel device to sample the esophageal microbiome — the esophageal string test. PLoS One 2012;7:e42938.

17. Elliott DRF, Walker AW, O'Donovan M, et al. A non-endoscopic device to sample the oesophageal microbiota: a case-control study. Lancet Gastroenterol Hepatol 2017;2:32-42.

18. Walters WA, Caporaso JG, Lauber CL, et al. PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. Bioinformatics 2011;27:1159-61.

19. Caporaso JG, Lauber CL, Walters WA, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 2012;6:1621-4.

20. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 2019;37:852-7.

21. Callahan BJ, McMurdie PJ, Rosen MJ, et al. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 2016;13:581-3.

22. Desantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 2006;72:5069-72.

23. Gall A, Fero J, McCoy C, et al. Bacterial composition of the human upper gastrointestinal tract microbiome is dynamic and associated with genomic instability in a Barrett's esophagus cohort. PloS One 2015;10:e0129055.

24. Okereke IC, Miller AL, Hamilton CF, et al. Microbiota of the oropharynx and endoscope compared to the esophagus. Sci Rep 2019;9:10201.

25. Dong L, Yin J, Zhao J, et al. Microbial similarity and preference for specific sites in healthy oral cavity and esophagus. Front Microbiol 2018;9:1603.

26. Yu G, Gail MH, Shi J, et al. Association between upper digestive tract microbiota and cancer-predisposing states in the esophagus and stomach. Cancer Epidemiol Biomarkers Prev 2014;23:735-41.

27. Yang L, Lu X, Nossa CW, et al. Inflammation and intestinal metaplasia of the distal esophagus are associated with alterations in the microbiome. Gastroenterology 2009;137:588-97.

28. Andersson AF, Lindberg M, Jakobsson H, et al. Comparative analysis of human gut microbiota by barcoded pyrosequencing. PLoS One 2008;3:e2836.

29. Harris JK, Fang R, Wagner BD, et al. Esophageal microbiome in eosinophilic esophagitis. PLoS One 2015;10:e0128346.

30. Benitez AJ, Hoffmann C, Muir AB, et al. Inflammation-associated microbiota in pediatric eosinophilic esophagitis. Microbiome 2015;3:23.

31. Liu Y, Lin Z, Lin Y, et al. Streptococcus and Prevotella are associated with the prognosis of oesophageal squamous cell carcinoma. J Med Microbiol 2018;67:1058-68.

32. Deshpande NP, Riordan SM, Castaño-Rodríguez N, et al. Signatures within the esophageal microbiome

are associated with host genetics, age, and disease. Microbiome 2018;6:227.

33. Wu GD, Chen J, Hoffmann C, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science 2011;334:105-8.

34. Nobel YR, Snider EJ, Compres G, et al. Increasing dietary fiber intake is associated with a distinct esophageal microbiome. Clin Transl Gastroenterol 2018;9:199.

35. Yamamura K, Baba Y, Nakagawa S, et al. Human microbiome fusobacterium nucleatum in esophageal cancer tissue is associated with prognosis. Clin Cancer Res 2016;22:5574-81.

36. O'Toole PW, Jeffery IB. Gut microbiota and aging. Science 2015;350:1214-5.