Fusobacterium Nucleatum Predicts a Risk for Esophageal Squamous Cell Carcinoma

Chao Shi

Affiliated Tumor Hospital of Zhengzhou University: Henan Cancer Hospital

Zhen Li

Henan Provincial People's Hospital

Jiawen Zheng

Affiliated Tumor Hospital of Zhengzhou University: Henan Cancer Hospital

Weimin Kong

Affiliated Tumor Hospital of Zhengzhou University: Henan Cancer Hospital

Taibing Fan

Central China Fuwai Hospital

Dongdong Jian

Central Chian Fuwai Hospital

Xiaolei Cheng

Central China Fuwai Hospital

Huan Zhao

First Affiliated Hospital of Zhengzhou University

Jie Ma

Affiliated Tumor Hospital of Zhengzhou University: Henan Cancer Hospital

Hao Tang

Central China Fuwai Hospital

Yongjun Guo (✉ yongjunguo@hotmail.com)

Henan Cancer Hospital

Research

Keywords: ESCC, *Fusobacterium nucleatum*, Tumor Mutation Burden, Tumor progression.

Posted Date: November 5th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-101163/v1>

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

[Read Full License](#)

Abstract

Background: Esophageal squamous cell carcinoma (ESCC) is the major type of esophageal cancer in China. The role of the bacteria present in Esophageal squamous cell carcinoma (ESCC) tissue in neoplastic progression has not been fully elucidated. This study aimed to uncover different bacterial communities in ESCC tissues and examine the correlation between the abundance of the esophageal flora and clinical characteristics of ESCC.

Method: Samples from total of 111 ESCC patients were collected, and 21 pairs of ESCC samples were analyzed by 16S rRNA sequencing. qPCR was used to verify the abundances of specific bacteria. Gene Ontology and protein domain analyses were used to analyze the whole-exome sequencing data of ESCC specimens to analyze the relationship between the abundance of the esophageal flora and signaling pathways.

Results: Microorganisms in tumors and normal tissues showed obvious clustering characteristics. The abundance of *Fusobacterium* ($P=0.0052$) was increased in tumor tissues. The high level of *Fusobacterium nucleatum* was significantly associated with pT stage ($P=0.039$) and clinical stage ($P=0.0039$). The WES data showed that *COL22A1*, *TRBV10-1*, *CSMD3*, *SCN7A* and *PSG11* were present in only the *F. nucleatum*-positive ESCC samples. GO and protein domain enrichment results suggested that epidermal growth factor might be involved in the regulation of cell apoptosis in *F. nucleatum*-positive ESCC. Both a higher mutational burden and *F. nucleatum*-positive was observed in tumors with metastasis than in tumors without metastasis.

Conclusion: *F. nucleatum* is closely related to the pT stage and clinical stage of ESCC. The abundance of *F. nucleatum* and tumor mutation burden may be used in combination as a marker to predict metastasis in ESCC.

Background

Esophageal cancer is the main malignant tumor of the upper digestive tract worldwide. There are approximately 570 thousand new cases of esophageal cancer globally every year, of which more than 80% are new cases in developing countries, half of which are in China(1). Esophageal squamous cell carcinoma (ESCC) is the major type of esophageal cancer in China, in contrast to western countries. Henan Province is a typical region with a high incidence and mortality of esophageal cancer in China. Despite the development of multimodal therapies, including surgery and chemotherapy radiotherapy, the prognosis of patients, including those who undergo complete resection, remains poor. The five-year survival rate of this disease has remained only approximately 10%-25%(2). Therefore, further studies are needed to clarify the pathogenesis of esophageal cancer and to explore new diagnostic and therapeutic possibilities.

Many studies have demonstrated that the composition of the microbiota in human skin and various mucous membranes (e.g., mouth, upper respiratory tract, lower genitourinary tract, gastrointestinal tract)

is closely related to health and disease(3). Digestive tract microorganisms are very important in all parts of the human body because they can directly or indirectly regulate the digestive system, immune system, nervous system, circulatory system, brain and other organ functions(4). In recent years, reports have paid attention to the microbiome as the key player triggering tumorigenesis. *Helicobacter pylori* infection is the strongest known risk factor for gastric adenocarcinoma (5), and it is also a risk factor for liver cancer, esophageal cancer and colorectal cancer; *Fusobacterium nucleatum* is the main cause of colorectal cancer and pancreatic cancer (6, 7); *Porphyromonas gingivalis* has been proven to be closely related to the occurrence of oral squamous cell cancer, gastrointestinal cancer or pancreatic cancer(8, 9); and *Enterobacter*, *Bacteroides* and *Enterococcus* are also responsible for multiple organ cancers throughout the body. Recently, a microbiome analysis of seven human tumor microenvironments showed that bacteria are widely found in cancer cells and immune cells in various tumors, and the bacterial content is related to the type of tumor(10). The crucial event in carcinogenesis triggered by the microbiome seems to be chronic inflammation influencing the genomic stability of host cells and activating immune mechanisms (11–13).

Due to the limitation of sampling methods and esophageal environmental dynamics, the study of the microbiome in the esophagus is still in its infancy; in particular, the relationship between ESCC and the microflora needs to be further studied. Recent studies have shown that the diversity of the microflora in esophageal cancer tissues is significantly different from that in normal esophageal tissues(14). Recent studies have shown that the microbiota of esophageal carcinoma is associated with prognostic survival. *F. nucleatum* in esophageal cancer tissues is associated with shorter survival, suggesting a potential role as a prognostic biomarker(15, 16). *F. nucleatum* might also contribute to aggressive tumor behavior through activation of chemokines, such as CCL20 (15). However, little information is available concerning the association between microorganisms and pathological characteristics in ESCC.

In this study, 41 pairs of tumors and normal tissues of ESCC patients were used for 16S rRNA or whole-exome sequencing (WES) sequencing. We also quantified *F. nucleatum* DNA in 98 samples of ESCC by qPCR and detected the relationship between *F. nucleatum*, tumor gene mutations and clinical characteristics of ESCC. Moreover, the combined mutational burden and the content of *F. nucleatum* in tumors could predict postoperative tumor metastasis in ESCC. In brief, we focused on the relationship between the specific microbiome and pathological characteristics and explored its application prospects as a biomarker for the pathogenesis and progression of ESCC.

Materials And Methods

Ethical statement

The Ethics Committee of Henan Cancer Hospital affiliated with Zhengzhou University approved this study (2017407). All surgical tissue specimens were collected after obtaining informed and written consents from the patients.

Sample collection

We retrospectively collected 152 surgical specimens from 111 individuals (including 82 tumor and paired nontumor tissue samples from 41 cases, and 70 individual ESCC tumor tissue samples) with a diagnosis of primary ESCC by surgery, confirmed by postoperative pathology, at Henan Cancer Hospital affiliated with Zhengzhou University from 2016 to 2017. Exclusion criteria include those individuals who have received pre-operative radiation, chemotherapy treatment or had a past history of ESCC or inflammatory bowel disease. We collected retrospective clinic-pathological data, including gender, age, drinking, smoking, tumor location, histopathological grading, clinical pT stage, lymph node metastasis (N stage) and clinical tumor stage, according to the 7th edition of the Union for International Cancer Control (UICC) /TNM. Detailed demography and characteristics of the ESCC tissue of the study patients are as summarized in Table 1.

DNA extraction and 16S ribosomal RNA sequencing

Total DNA were extracted from each specimen using the QIAamp DNA FFPE Tissue kit (QIAGEN, Germany) according to the manufacturer's instructions. The DNA was finally eluted in approximately 50 µL sterile distilled water and was stored at -20°C until use. We quantified extracted DNA using High Sensitivity Qubit (Thermofisher, USA) and amplified 30 ng of DNA using specific barcoded primers (336F: GTACTCCTACGGGAGGCAGCA; 806R: GTGGACTACHVGGGTWTCTAAT) around the V3-V4 region of 16S rRNA gene. The amplified samples were sequenced through using the 16S rRNA high-throughput next generation Illumina MiSeq (250 PE, USA) platform in Allwegene company (Beijing, China).

Analysis of 16S rRNA amplicon sequence data

Sequences analyses were performed using quantitative insights into microbial ecology (QIIME) (17). Operational taxonomic units (OTUs) at 97% similarity were selected and cross-referenced to the Greengenes and silva128 database(18), which we trimmed to span only the 16S rRNA region flanked by our sequencing primers. In addition, Magic-BLAST was used to eliminate human genome noise from the host DNA contamination. Trimmomatic (V0.32) was used to trim FASTQ data, quality check(19) and Cutadapt removed adapter sequences and primers. Dada2 was performed for relative abundance, alpha-/beta-diversity and sample cluttering heatmap analyses(20). Bray-Curtis principal coordinate analysis was performed using QIIME script "beta_diversity_through_plost.py".

Quantitative polymerase chain reaction (qPCR) assays

The amount of *F. nucleatum* DNA was quantified by use of a real-time qPCR assay. Each reaction contained 40 ng of genomic DNA and was assayed in triplicate in 10 µL reactions containing 1 × Power SYBR Green PCR Master Mix (Applied Biosystems, Carlsbad, CA, USA), 0.4 µM each primer and was placed in a 96-well optical PCR plate(21). The primers of *nusG* gene of *F. nucleatum* and the reference human gene *SLCO2A1* as described previously (15,22): *nusG* forward primer, 5'-TGGTGTCATTCTTCCAAAAATATCA-3'; *nusG* reverse primer, 5'-AGATCAAGAAGGACAAGTTGCTGAA-3'; *SLCO2A1* forward primer, 5'-ATCCCCAAAGCACCTGGTTT-3'; *SLCO2A1* reverse primer, 5'-AGAGGCCAAGATAGTCCTGGTAA-3'; Amplification and detection of DNA was performed with the

StepOnePlus real-time PCR Systems (Applied Biosystems) using the following reaction conditions: 10 min at 95°C, 45 cycles of 15 s at 95°C and 1 minute at 60°C. The amount of tissue *F. nucleatum* DNA in each specimen was calculated as a relative unitless value normalised with *SLC02A1* using the $2^{-\Delta Ct}$ method (where $\Delta Ct = \text{the mean Ct value of } F. \text{ nucleatum} - \text{the mean Ct value of } SLC02A1$) (6). The reliability of qPCR results was evaluated by the dissolution curve, and the Ct value (inflection point of the expanded dynamic curve) was taken. Any sample with a Ct less than 37 was considered *F. nucleatum*-positive. Specimens were considered negative when the Ct value of the specimen was greater than 37.

FISH analysis

FISH analysis was performed on the FFPE of ESCC, culturing with 100 to 120 ng of each probe (Sangon, Shanghai, China) in a hybridization buffer. Cell nuclei were stained with DAPI. An Oregon-Green 488-conjugated “universal bacterial” probe (EUB338, pB-00159, green) binding 16S rRNA gene at bacterial conserved regions and a Cy3conjugated *Fusobacterium* probe (FUSO, pB-2634, red) binding 16S rRNA gene at *Fusobacterium* specific regions were applied. The sequences of the probes were referred to probeBase (<http://www.microbialecolgy.net/probebase>) (Loy, 2008 #229) (23,24). FISH analysis was performed as described previously (25). Five random × 40 fields were chosen for evaluation by two pathologists blind to tumor/normal status. The selection criteria of mucosal tissue depth were used, and a minimum of five bacteria visualized by the EUB338 probe per field was required.

Whole-Exome Sequencing (WES)

Tumor and normal genomic DNA was isolated from paraffin tissue blocks using the Qiagen Allprep DNA/RNA Micro Kit (Qiagen, Hilden, Germany). DNA was quantified using the Qubit quantitation system with standard curve as per the supplier protocol (Thermo Fisher, Waltham, MA, USA). After extraction, the nucleic acids of 21 pairs samples were subjected to quantitative and qualitative evaluation, using the NanoDrop instrument, and the Agilent Bioanalyzer (Agilent Technologies, Santa Clara USA), respectively. All the final DNA library were subsequently sequencing on the MGISEQ-2000 high-throughput platform. DNA nanoballs (DNBs) were generated with the ssDNA circle by rolling circle replication (RCR) to enlarge the fluorescent signals at the sequencing process. The DNBs were loaded into the patterned nanoarrays and pair-end read of 100 bp were read through on the MGISEQ-2000 platform for the following data analysis study. For this step, we applied the advanced combinatorial Probe-Anchor Synthesis (cPAS) technology and DNA Nanoballs (DNB) sequencing technology.

WES sequence data analysis

After removal of terminal adaptor sequences and low-quality data, reads were mapped to the reference human genome (hg19) and aligned using BWA version 0.5.9 (Broad Institute). Sequencing data from paired tumor-normal samples were used to identify somatic mutations. After annotation, the variants were cross referenced with those in the 1000 Genomes Project, GAD, dbSNP, and the ExAC. Variants with an allele prevalence >1% in the 1000 genomes project, 1000 genomes project EAS, ExAC, ExAC EAS were excluded. Single-nucleotide variants were identified using MuTect (version 1.1.4) and NChot. Small

insertions and deletions were determined using GATK. Copy number variations were detected using the CONTRA tool (2.0.8). An in-house algorithm was used to identify split reads and discordant read pairs to identify gene fusion. At least five supporting reads were required for true fusion. All final candidate variants were manually verified with the integrative genomics viewer browser.

Statistical analysis

The biological functions of the related genes were explored by GO and pathway enrichment analysis in DAVID Bioinformatics Resources 6.8 (26). Statistical analysis and drawing were performed using the GraphPad Prism 8 (GraphPad, San Diego, California, USA). One-way ANOVA was used to determine significant among groups. Unpaired t test was performed within patient tumor and normal tissues. Principle coordinate analysis, tree clustering analyses, cluster dendrogram with AU/BP values were used for clustering between tumor and normal groups. All *P*-values were two-sided, and *P* < 0.05 was considered statistically significant.

Results

Sample collection and clinical information

In this study, we collected esophageal tissue wax blocks from 111 individuals, of which 41 included tumors and paired nontumor tissues and the other 70 were only tumor tissues (Fig. 1). Forty-one patients with ESCC were included, and microbiome samples from both tumor and normal tissues were collected. The clinical information of the 111 patients is shown in Table 1.

Figure 1. Diagram of study enrollment and eligible paired and total samples. ESCC indicates esophageal squamous cell carcinoma.

Microbiota diversity in ESCC

We generated 13,671,987 quality-filtered sequence reads, with an average of 48,311 reads per sample. Sequence reads were mapped to the bacteria in the SILVA database. In general, the 16S rRNA gene sequencing results of paired tumor and normal tissue microbial profiles show partial differences. For alpha-diversity, the Shannon index (4.93 vs 5.05, *P*=0.6233), PD_whole_tree (9.35 vs 10.73, *P*=0.2645) and the observed species (46.97 vs 52.29, *P*=0.1893) were lower in the ESCC tumor tissues than in the normal tissues, but these differences were not statistically significant; however, the overall alpha-diversity Chao 1 index (132.06 vs 185.86, *P*=0.0091) in tumor tissues decreased dramatically by approximately 25% compared with that in the matched normal tissues (Fig. 2A). For the microbial community composition (beta-diversity), significant clustering was detected for the weighted UniFrac distance between paired ESCC tumor and nontumor tissues (Fig. 2B).

Relative abundances of bacteria at the phylum level in both the normal and tumor groups of ESCC patients

To further determine the signature of microbial profiles in both the normal and tumor groups, the relative abundances of bacteria at the phylum level were determined, and the results are shown in Fig. 2C (Class level was shown in Fig. S1). The top 6 phyla across all samples were *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Fusobacteria*, *Proteobacteria* and *Spirochaetes*. Differential abundance analysis revealed a similar conclusion: *Bacteroidetes* ($P=0.03149$), *Fusobacteria* ($P=0.01796$) and *Spirochaetae* ($P=0.004155$) were the top 3 significantly enriched taxa in the tumor group compared with the matched normal group samples (Fig. 2D). To identify the signature microbiota existing in tumor tissues, which potentially play roles during the carcinogenesis of ESCC, 16S rRNA sequencing data were annotated at the genus level. *Butyrivibrio* ($P=0.0078$) and *Lactobacillus* ($P=0.00052$) were closely associated with sugar and fiber fermentation and were found at significantly lower relative abundances in the tumor group than in the normal group. *Streptococcus* ($P=0.0013$) was discovered increased in tumor group. In addition, *Fusobacterium* is anaerobic, gram-negative bacteria and normally treated as pathogen, which discovered as increasing relative abundance in tumor groups ($P=0.0052$) (Fig. S2). The full list of signature microbiota profiles (genus level) is summarized in Table S1.

Figure 2. Microbial comparison of paired tumor and paired-normal tissues from patients with ESCC for α (A) and β diversity (B). The α diversity of 21 paired ESCC tumor and nontumor tissues was assessed with the Chao1 index, Shannon index, observed species and PD_whole_tree. The β diversity, based on the weighted UniFrac distance, was compared with PCoA plots for 21 paired tumor and nontumor tissues from patients with ESCC. PCoA, principal coordinate analysis. (C) The relative abundance of bacteria at the phylum level in both the normal and tumor groups of ESCC. (D) Differential abundance analysis revealed that *Bacteroidetes* ($P=0.03149$), *Fusobacteria* ($P=0.01796$) and *Spirochaetae* ($P=0.004155$) were the top 3 significantly enriched taxa in the tumor group compared with the matched normal group samples. ** P -value<0.01.

Microbiota associated with clinical characters in ESCC

To explore whether there are regular differences among the compositions of microorganisms in ESCC, we grouped 21 pairs of samples according to their respective clinical stages and clustered trees based on the unweighted UniFrac distance matrix and UPGMA method and integrated the clustering results with the relative abundances of each sample species. The results of the histogram (Fig. 3A) and heatmap showed that the relative abundances of *Fusobacterium* ($P=0.039$) and *Prevotella* ($P=0.0379$) were correlated with clinical stage in ESCC, where they were higher in tumors than in the corresponding normal tissues, not with the IA stage.

To further explore the microbiome alterations that occur during the progression of esophageal cancer and identify signature species as diagnostic biomarkers, we performed differential abundance analyses on 16S rRNA sequencing data of tumor tissues only using DESeq2(27), and cross-referenced with both patient clinical tumor classification of malignant tumor.

Subjects in the pT1-T2 classes were designated Group a, and those in pT3-T4 were designated Group b. In the pT classification, α diversity measures indicated a decreased diversity in Group b (Fig. 3B).

Although there was no significant difference in the Chao1 (1094.03 vs 843.77, $P=0.0659$) index, PD_whole_tree (493.81 vs 450.37, $P=0.0771$) and the observed species (790.09 vs 626.27, $P=0.0819$), the Shannon index (5.858 vs 5.219, $P=0.0242$) decreased significantly in Group b. We found that the relative abundances of the phyla *Fusobacteria* ($P=0.0048$) and *Bacteroidetes* ($P=0.0035$) increased significantly in Group b (Fig. S4). In addition, multiple comparisons between every pT stage were performed, and bacterial taxa with significant changes in abundance are summarized in Table S2. Many bacterial genera were significantly changed, including *Lactobacillus*, members of which show preventive effects in bacterial infection and inflammation (28). Moreover, data from clinical stage I-II subjects were grouped (Group c) and then compared with stage III subject tumor microbiome profiles (Group d). As illustrated in the figure (Fig. 3C), there were no statistically significant differences between groups c and d in α diversity as measured by the Chao1 index (1011.25 vs 984.73, $P=0.743$), Shannon index (5.67 vs 5.64, $P=0.7048$), PD_whole_tree (480.52 vs 464.56, $P=0.2355$) and the observed species (741.26 vs 720.91, $P=0.624$). The data showed that the relative abundance in stage III was higher than that in stage I-II; however, there were no statistically significant differences. In addition, the current analysis shows that the distribution of microflora diversity in tumor tissue is not related to the patient's age, gender, smoking history, drinking history, tumor location, lymph node metastasis or G stage (Table S3).

Figure 3. Microbiota associated with clinical tumor stage in ESCC. (A) Twenty-one pairs of samples according to their respective clinical stages and clustered trees based on the unweighted UniFrac distance matrix and UPGMA method; the clustering results were integrated with the relative abundance of each sample species. The α diversity at the clinical pT stage (B) and clinical tumor stage (C) was assessed with the Chao1 index, Shannon index, observed species and PD_whole_tree.

FISH detection

F. nucleatum was detected in 20 specimens of ESCC tissue by FISH. *F. nucleatum* stained with the FUSO probe (in red) was found to be enriched within the esophageal tumor tissue (median, 23; range, 5 to 69) (Figure 4).

Figure 4. Enrichment of *Fusobacterium nucleatum* in ESCC detected by FISH. Graph A: ESCC specimens stained with DAPI. Cell nuclei were stained in blue; B: ESCC specimens stained with both DAPI and a universal bacterial probe (EUB338). Bacterial conserved regions were stained in green; C: ESCC specimen stained with both DAPI and a *Fusobacterium*-specific probe (FUSO). The *F. nucleatum*-specific regions were stained in red; D: ESCC specimens triply stained with DAPI, FUSO and EUB338. Cell nuclei (in blue), bacterial conserved regions (in green) and *F. nucleatum*-specific regions (in red) were clearly visible (all $\times 400$).

qPCR to verify the level of *Fusobacterium nucleatum* in ESCC

To verify the results of 16S rRNA gene sequencing, we collected more samples to check the relationship between *F. nucleatum* and clinical characteristics of ESCC. To evaluate the relative abundance of *F. nucleatum* in tumor tissues, the specific *nusG* gene of *F. nucleatum* was quantitatively analyzed by qPCR

in samples from 98 ESCC patients. The results showed that of the 98 tumor tissue samples, 69.4% (68/98) were positive for *F. nucleatum*, and the relative abundance of *F. nucleatum* in tumor tissues was significantly higher than that in paired normal tissues ($P=0.0262$) (Fig. 5A). By univariate analysis, it was found that the abundance of *F. nucleatum* was highly correlated with the pT and clinical stages. The relative *F. nucleatum* levels in the pT3-4 stage were significantly higher than those in the pT1-2 stage ($P=0.039$), and there was greater enrichment in the III stage than in the I-II stage ($P=0.0039$) (Fig. 5B and C). This is consistent with the above analysis results showing that the relative *F. nucleatum* DNA levels in ESCC tumor tissues were not significantly changed patient age, gender, smoking history, drinking history, lymph node metastasis and G stage (Table 1 and Fig. S5).

Figure 5. Distribution of *F. nucleatum* in ESCC. (A) The relative abundance of *F. nucleatum* in tumor tissues was significantly higher than that in paired normal tissue samples in 21 patients with ESCC ($P=0.0262$). (B) The relative *F. nucleatum* levels in the pT3-4 stage were significantly higher than those in the pT1-2 stage ($P=0.039$), and the enrichment was higher in the III stage than in the I-II stage ($P=0.0039$).

Mechanism by which *F. nucleatum* affects the progression of ESCC

According to the qPCR results, we selected 20 pairs of ESCC samples for whole-exome sequencing (WES). Thirteen tumor tissues contained *F. nucleatum*, and the other 7 were *F. nucleatum* negative. GO functional enrichment analysis of all the high-risk mutant genes of the 13 *F. nucleatum*-positive samples was conducted, and the most significantly enriched functions are shown in Fig. 6A. The most significantly enriched functions included cell cycle ($P=0.00071$), positive regulation of apoptotic process ($P=0.0063$) and positive regulation of transcription, DNA-templated ($P=0.0049$). Moreover, the involved proteins were further classified by protein domain, and it was found that the epidermal growth factor-related domain (EGF-like domain) was significantly enriched ($P=0.00029$) (Fig. 6B). The enriched GO terms for high-risk mutant genes in *F. nucleatum*-positive specimens were significantly different from those for the 7 *F. nucleatum*-negative samples (Fig. S6).

***F. nucleatum* significantly mutated genes in ESCC**

MutSigCV analysis of 20 ESCC paired tissues revealed 20 significantly mutated genes (Fig. 6C). ESCC samples were divided into two groups according to the abundance of *F. nucleatum*. Any sample with a Ct less than 37 was considered *F. nucleatum*-positive. Specimens were considered negative when the Ct value of the specimen was greater than 37 or no melt curve could be generated. *TP53* mutations were present in all *F. nucleatum*-positive ESCCs. Most *TP53* mutations were in the hotspot exons 4-8. *TP53* is involved in the regulation of cell proliferation and apoptosis. Five of the 13 *F. nucleatum*-positive ESCCs harboring *TP53* mutations had concomitant mutations in *COL22A1*. The frequency of *F. nucleatum*-negative ESCC harboring *RBMXL3* and *HDGFRP2* mutations was 57.14% and 42.85%, respectively, which are different from the values for the *F. nucleatum*-positive group. Notably, we also identified a number of genes as occurring in only the *F. nucleatum*-positive group, including *COL22A1* (38.46%, 5/13), *TRBV10-1* (23.07%, 3/13), *CSMD3* (30.76%, 4/13), *SCN7A* (15.38%, 2/13) and *PSG11* (7.69%, 1/13) (Fig. 6C).

Relationship between *F. nucleatum* and the mutational burden in ESCC

Of the 20 patients who provided samples for sequencing, 11 had distant metastasis within 6 months after the operation. There was no metastasis during 2 years of follow-up in the other 9 patients. Among these 11 patients, 72.8% (8/11) were *F. nucleatum* positive. A higher (mean $\pm$ SE) mutational burden was observed in the tumors of 11 patients with metastasis than those of patients without metastasis (141.5 ± 16.94 vs. 124 ± 16.31 , $P = 0.467$), but there was no significant difference (Fig. 6D). The mutational burden per primary tumor was identified (median, 133.6; range, 33 to 256). The high-mutational-burden group was defined as patients with 120 or more mutations, and the low-mutational-burden group was defined as patients with less than 120 mutations. The results indicated that there was a significant correlation between the combination of the mutational burden and the *F. nucleatum* content in tumors and metastasis in ESCC. In *F. nucleatum*-positive ESCC, there were more patients with lymphatic and distal metastasis in the high-mutational-burden group than in the low-mutational-burden group.

Figure 6. (A) Gene Ontology (GO) enrichment results; (B) protein domain enrichment analysis results; (C) mutation frequencies and signatures, and significantly mutated genes in 20 ESCCs. The number of somatic mutations of each examined case (top) and the significantly mutated genes (SMGs) colored by the types of mutations and their mutational frequency (bottom). Columns, examined cases; rows, genes; N, *F. nucleatum*-negative group; P, *F. nucleatum*-positive group. (D) Combination of the mutational burden and the *F. nucleatum* content to predict metastasis in ESCC. The red horizontal line indicates the cutoff score for mutational burden. Clinical and pathological features that include the presence or absence of distant metastases in the surgical specimen are annotated for each patient. A sample was considered *F. nucleatum* positive if the qPCR Ct value was <37 and the melt curve could be generated; otherwise, it was considered negative.

Discussion

The esophageal mucosa is among the sites colonized by the human microbiota, the complex microbial ecosystem that colonizes various body surfaces and is increasingly recognized to play roles in several physiological and pathological processes. To investigate the role of the microbiota of ESCC in different pathological stages, we performed 16S rRNA high-throughput sequencing in esophageal cancer tissues and normal tissues from 21 ESCC patients. The six major bacterial phyla (e.g., *Actinobacteria*, *Bacteroides*, *Firmicutes*, *Fusobacteria*, *Proteobacteria* and *Spirochaetes*) in these tissues are similar to the high-abundance bacterial types found in esophageal mucosa in previous studies (29). *Proteobacteria*, *Firmicutes*, and *Actinobacteria* were the top 3 taxa that were significantly depleted in tumor tissues. In addition, specific bacteria are closely related to the progression of tumors in the pT stage and clinical stage, especially *Fusobacterium* and *Streptococcus*, which were also discovered to be enriched in the tumor group and may be among the important factors for the progression of ESCC. Previous studies on the diversity of bacterial flora between tumor and normal tissues have not been conclusive (30). Bacterial diversity differs significantly according to cancer type. Some studies suggested that there were differences in the relative abundances of bacteria in oral squamous cell carcinoma (OSCC) patients;

however, no statistically significant differences in phylogenies were detected for tumor and normal tissue sites except for the genus *Johnsonella*(31). In contrast, colorectal tumor tissues harbor distinct microbial communities compared to nearby healthy tissue (32). In this study, our analysis showed significant differences in the microbiota between tumor and normal tissues. The six major phyla of bacteria identified in tumor tissues are similar to those in normal or diseased esophageal mucosa, perhaps due to proximity (29). However, they still show obvious clustering characteristics, which also implies that specific flora components exist in tumors and normal tissues, and identifying the difference between the two specific types of bacteria may be the key to understanding tumorigenesis.

Genus-level analysis showed that the abundances of *Fusobacterium* and *Streptococcus* increased significantly, while the abundances of *Butyrivibrio* and *Lactobacillus* decreased, in tumor tissues compared with those in normal tissues. *Butyrivibrio* species are considered nonpathogenic bacteria that can utilize cellulose, starch and other polysaccharides for chemical organic nutrition, and the main product of glucose fermentation by these species is butyric acid (33). Butyric acid is an important short-chain fatty acid (SCFA) that can promote the growth of beneficial bacteria, such as *Lactobacillus*, in the body. *Lactobacillus* species are generally considered to be probiotics and can inhibit the growth of harmful bacteria(34). The relative abundances of these two genera of bacteria decreased significantly, resulting in changes in the diversity of the microbiota in tissues and possibly increasing the number of pathogenic or harmful bacteria. Correspondingly, we found a significant increase in the abundances of *Fusobacterium* and *Streptococcus*, which are positively correlated with tumorigenesis and inflammation. Several studies have found that *Streptococcusgallolyticus* is related to colorectal cancer and can release toxins, regulate the tumor microenvironment or stimulate the immune response of immune cells (35,36). Moreover, *Streptococcus* abundance retained its association with unfavorable survival, suggesting that this may be an independent prognostic indicator for ESCC (37). *Fusobacterium* is a genus that encompasses several species known to be opportunistic pathogens in humans; they are obligate anaerobes with known sites of infection in the oral cavity as well as in the gastrointestinal tract (38,39). Relative *F. nucleatum* DNA levels in ESCC tissues compared with those in corresponding nontumor tissues were examined by qPCR, and FISH was used to analyze the distribution of *F. nucleatum* in ESCC tissue. Consistent with a previous study, the relative abundance of *F. nucleatum* was significantly higher in ESCC tissues than in adjacent nontumor tissues, which suggests that it is involved in the development of malignant tumors(15,40). One possibility is that these bacteria may play a role in the etiology of ESCC; another possibility is that these species are enriched as the tumor has formed a niche that favors these bacteria.

Furthermore, we explored the relationship between specific microbiota composition and the pathological development of ESCC. In regard to pT stage, previous studies have shown that the abundance of only *Streptococcus* in pT3-4 was significantly higher than that in pT1-2 of ESCC, while the other genera showed no significant change(37). However, in our results, *Fusobacteria* and *Bacteroidetes* were found in pT3-4 at the phylum level, and the relative abundance of *F. nucleatum*, an important representative strain of *Fusobacterium*, in pT3-4 was also significantly higher than that in pT1-2. As a strictly anaerobic gram-negative bacterium, *F. nucleatum* is mainly distributed in human and animal intestines and in the oral

mucosa. *F. nucleatum* is the most abundant species in the oral cavity and has come to the forefront of scientific interest because of an increasing number of associations with extraoral diseases(7). In many studies, it has been shown that *F. nucleatum* is closely related to the occurrence of tumors, especially colon cancer (6,15,16,41). In addition, it is positively related to the poor prognosis of esophageal cancer (15). Recently, research has found that a higher *F. nucleatum* burden correlates with a poor response to neoadjuvant chemotherapy in ESCC (22). According to previous studies on *F. nucleatum*, combined with the different relative abundances of *F. nucleatum* in different pT and clinical stages of ESCC in this study, *F. nucleatum* might be an important factor in the tumor progression of ESCC (16,42,43). This also provides a new way for us to understand the causes of the occurrence and development of ECSS.

Gene mutation is an important factor that affects signal pathway regulation and participates in tumor pathogenesis. Accumulated gene mutations accelerate genome instability, which eventually leads to uncontrollable growth of the tumor. However, it is not clear whether *F. nucleatum* affects the progression of ESCC by inducing gene mutations. It is hypothesized that bacterial species such as *Escherichia coli*, *F. nucleatum*, and enterotoxigenic *Bacteroides fragilis* have a role in colorectal carcinogenesis. In addition, bacteria produce toxins that inhibit the immune response and cause DNA damage(44). However, *F. nucleatum* encodes no known toxins and very few canonical 'virulence factors'(7). However, epidemiological associations suggest that *F. nucleatum* can promote genome instability and mutation. A previous study showed that *TP53*, *KRAS*, and *BRAF* mutations were additionally associated with *F. nucleatum*-positive colorectal cancers (CRCs) (45). We analyzed 13 samples of *F. nucleatum*-positive ESCC by WES and found that the function of the mutant gene is mainly concentrated in the pathway of positive regulation of apoptosis and the epidermal growth factor-like protein domain. Cancer cells have developed mechanisms by which apoptosis is evaded through the mutation of essential genes involved in the regulation of the process. Epidermal growth factor (EGF) and its receptor play an important role in signaling pathways and in regulating cell proliferation, migration, differentiation and apoptosis (46). We found a common mutant gene, *TP53*, in 13 *F. nucleatum*-positive ESCC tumor tissues, and we found genes that were present only in *F. nucleatum*-positive tumor tissues, such as *COL22A1*, *TRBV10-1*, *CSMD3*, *SCN7A* and *PSG11*. This finding is different from previous reports showing that *KRAS* and *BRAF* frequently occur in CRCs(45). We infer that this may be caused by the small sample size and different cancer types. Nevertheless, our results support the relationship between *F. nucleatum* and tumor genetics. We advocate the need for large-scale research and further systems analysis to obtain conclusive evidence.

Tumor mutation burden (TMB), a quantitative measure of the total number of coding mutations in the tumor genome, is emerging as a potential biomarker. Both exogenous DNA damage caused by DNA-damaging agents and endogenous DNA damage caused by increased production of reactive oxygen species could produce gene mutations. Specific mutations in tumor cells may lead to peptide epitopes or tumor-specific antigens for T-cells, and these antigens could serve as immunotherapeutic targets. The number of somatic mutations varies widely both between and within cancer types. Previous clinical trials in multiple tumors have also shown that there is a positive correlation between TMB and the efficacy of PD-1/PD-L1 inhibitors and a trend toward longer PFS(47-49). Li *et al* found that a high TMB was

significantly associated with a worse prognosis and could promote tumor metastasis and development (50). Our results showed that the presence of *F. nucleatum* in tumors and the TMB could be combined to predict postoperative metastasis in esophageal cancer. In this study, we found that *F. nucleatum* was positively correlated with the degree of malignancy of ESCC. Therefore, *F. nucleatum* positivity in tumor tissue and high TMB may together serve as a biomarker for predicting human ESCC progression and metastasis.

Conclusion

In conclusion, we found that the abundance of *F. nucleatum* in ESCC tissue is closely related to the development of ESCC tumor tissues in the pT and clinical stages, suggesting that it may play an important role in the progression of ESCC. *F. nucleatum* might participate in the mechanisms by which epidermal growth factor (EGF) is induced to interfere with the cell apoptosis-mediated regulation of related genes in this process. Importantly, the abundance of *F. nucleatum* and the TMB might be used in combination as a marker to predict the potential of metastasis in ESCC.

Abbreviations

ESCC

esophageal squamous cell carcinoma

16S rRNA

16S ribosomal RNA

OTUs

Operational taxonomic units

QIIME

quantitative insights into microbial ecology

N

normal group

T

tumor group

G

tumor grades

pT

tumor invasive depth

GO

Gene Ontology

WES

whole exome sequencing

TMB

tumor mutation burden

Declarations

Acknowledgement

The authors would like to acknowledge the experimental platform provided by the microbiology laboratory of Henan Provincial People's Hospital and the helpful suggestions from colleagues.

Author contributions

These authors contributed equally to this work; Concept and design: Chao Shi, Zhen Li and Yongjun Guo; Acquisition of samples: Chao Shi, Huan Zhao; Experimental verification: Zhen Li, Weimin Kong and Dongdong Jian; Analysis or interpretation of data: Chao Shi, Zhen Li, Jiawen Zheng and Xiaolei Cheng; Drafting of the manuscript: Chao Shi, Zhen Li; Administrative, technical, or material support: Taibing Fan, Jie Ma, Hao Tang, Yongjun Guo. Supervision: Jie Ma, Hao Tang, Yongjun Guo.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 81802444 and No. 81901412) and Henan Medical Science and Technology Project (LHGJ20190811).

Availability of data and materials

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The Ethics Committee of Henan Cancer Hospital affiliated with Zhengzhou University approved this study (2017407). All surgical tissue specimens were collected after obtaining informed and written consents from the patients.

Conflicts of interest

None of the authors has any potential conflicts to disclose.

Consent for publication

Written informed consent for publication was obtained from all participants.

References

1. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, *et al.* Cancer statistics in China, 2015. *CA Cancer J Clin* **2016**;66(2):115-32 doi 10.3322/caac.21338.
2. Pennathur A, Gibson MK, Jobe BA, Luketich JD. Oesophageal carcinoma. *Lancet* **2013**;381(9864):400-12 doi 10.1016/S0140-6736(12)60643-6.

3. Young VB. The role of the microbiome in human health and disease: an introduction for clinicians. *BMJ* **2017**;356:j831 doi 10.1136/bmj.j831.
4. Dinan TG, Cryan JF. Gut-brain axis in 2016: Brain-gut-microbiota axis - mood, metabolism and behaviour. *Nature reviews Gastroenterology & hepatology* **2017**;14(2):69-70 doi 10.1038/nrgastro.2016.200.
5. Marshall BJ, Windsor HM. The relation of *Helicobacter pylori* to gastric adenocarcinoma and lymphoma: pathophysiology, epidemiology, screening, clinical presentation, treatment, and prevention. *Med Clin North Am* **2005**;89(2):313-44, viii doi 10.1016/j.mcna.2004.09.001.
6. Mima K, Nishihara R, Qian ZR, Cao Y, Sukawa Y, Nowak JA, *et al.* *Fusobacterium nucleatum* in colorectal carcinoma tissue and patient prognosis. *Gut* **2016**;65(12):1973-80 doi 10.1136/gutjnl-2015-310101.
7. Brennan CA, Garrett WS. *Fusobacterium nucleatum* - symbiont, opportunist and oncobacterium. *Nat Rev Microbiol* **2019**;17(3):156-66 doi 10.1038/s41579-018-0129-6.
8. Atanasova KR, Yilmaz O. Looking in the *Porphyromonas gingivalis* cabinet of curiosities: the microbium, the host and cancer association. *Mol Oral Microbiol* **2014**;29(2):55-66 doi 10.1111/omi.12047.
9. Groeger S, Domann E, Gonzales JR, Chakraborty T, Meyle J. B7-H1 and B7-DC receptors of oral squamous carcinoma cells are upregulated by *Porphyromonas gingivalis*. *Immunobiology* **2011**;216(12):1302-10 doi 10.1016/j.imbio.2011.05.005.
10. Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, *et al.* The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science* **2020**;368(6494):973-80 doi 10.1126/science.aay9189.
11. Schwabe RF, Jobin C. The microbiome and cancer. *Nat Rev Cancer* **2013**;13(11):800-12 doi 10.1038/nrc3610.
12. Zitvogel L, Ayyoub M, Routy B, Kroemer G. Microbiome and Anticancer Immunosurveillance. *Cell* **2016**;165(2):276-87 doi 10.1016/j.cell.2016.03.001.
13. Dzutsev A, Goldszmid RS, Viaud S, Zitvogel L, Trinchieri G. The role of the microbiota in inflammation, carcinogenesis, and cancer therapy. *Eur J Immunol* **2015**;45(1):17-31 doi 10.1002/eji.201444972.
14. Hunt RH, Yaghoobi M. The Esophageal and Gastric Microbiome in Health and Disease. *Gastroenterol Clin North Am* **2017**;46(1):121-41 doi 10.1016/j.gtc.2016.09.009.
15. Yamamura K, Baba Y, Nakagawa S, Mima K, Miyake K, Nakamura K, *et al.* Human Microbiome *Fusobacterium Nucleatum* in Esophageal Cancer Tissue Is Associated with Prognosis. *Clinical cancer research : an official journal of the American Association for Cancer Research* **2016**;22(22):5574-81 doi 10.1158/1078-0432.CCR-16-1786.
16. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, *et al.* *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Res* **2012**;22(2):299-306 doi 10.1101/gr.126516.111.

17. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **2010**;7(5):335-6 doi 10.1038/nmeth.f.303.
18. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **2006**;72(7):5069-72 doi 10.1128/AEM.03006-05.
19. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**;30(15):2114-20 doi 10.1093/bioinformatics/btu170.
20. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **2016**;13(7):581-3 doi 10.1038/nmeth.3869.
21. Yu T, Guo F, Yu Y, Sun T, Ma D, Han J, *et al.* *Fusobacterium nucleatum* Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell* **2017**;170(3):548-63 e16 doi 10.1016/j.cell.2017.07.008.
22. Yamamura K, Izumi D, Kandimalla R, Sonohara F, Baba Y, Yoshida N, *et al.* Intratumoral *Fusobacterium Nucleatum* Levels Predict Therapeutic Response to Neoadjuvant Chemotherapy in Esophageal Squamous Cell Carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* **2019**;25(20):6170-9 doi 10.1158/1078-0432.CCR-19-0318.
23. Loy A, Arnold R, Tischler P, Rattei T, Wagner M, Horn M. probeCheck—a central resource for evaluating oligonucleotide probe coverage and specificity. *Environmental microbiology* **2008**;10(10):2894-8 doi 10.1111/j.1462-2920.2008.01706.x.
24. Swidsinski A, Dörffel Y, Loening-Baucke V, Theissig F, Rückert JC, Ismail M, *et al.* Acute appendicitis is characterised by local invasion with *Fusobacterium nucleatum/necrophorum*. *Gut* **2011**;60(1):34-40 doi 10.1136/gut.2009.191320.
25. Li Y, Ge Q, Cao J, Zhou Y, Du Y, Shen B, *et al.* Association of *Fusobacterium nucleatum* infection with colorectal cancer in Chinese patients. *World journal of gastroenterology* **2016**;22(11):3227-33 doi 10.3748/wjg.v22.i11.3227.
26. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, *et al.* The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* **2007**;8(9):R183 doi 10.1186/gb-2007-8-9-r183.
27. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **2014**;15(12):550 doi 10.1186/s13059-014-0550-8.
28. Liu Y, Fatheree NY, Mangalat N, Rhoads JM. Human-derived probiotic *Lactobacillus reuteri* strains differentially reduce intestinal inflammation. *Am J Physiol Gastrointest Liver Physiol* **2010**;299(5):G1087-96 doi 10.1152/ajpgi.00124.2010.
29. Di Pilato V, Freschi G, Ringressi MN, Pallecchi L, Rossolini GM, Bechi P. The esophageal microbiota in health and disease. *Ann N Y Acad Sci* **2016**;1381(1):21-33 doi 10.1111/nyas.13127.

30. Gall A, Fero J, McCoy C, Claywell BC, Sanchez CA, Blount PL, *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(6):e0129055 doi 10.1371/journal.pone.0129055.
31. Pushalkar S, Ji X, Li Y, Estilo C, Yegnanarayana R, Singh B, *et al.* Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma. *BMC Microbiol* **2012**;12:144 doi 10.1186/1471-2180-12-144.
32. Burns MB, Lynch J, Starr TK, Knights D, Blekhman R. Virulence genes are a signature of the microbiome in the colorectal tumor microenvironment. *Genome Med* **2015**;7(1):55 doi 10.1186/s13073-015-0177-8.
33. Sechovcova H, Kulhava L, Fliegerova K, Trundova M, Morais D, Mrazek J, *et al.* Comparison of enzymatic activities and proteomic profiles of *Butyrivibrio fibrisolvens* grown on different carbon sources. *Proteome Sci* **2019**;17:2 doi 10.1186/s12953-019-0150-3.
34. Duar RM, Lin XB, Zheng J, Martino ME, Grenier T, Perez-Munoz ME, *et al.* Lifestyles in transition: evolution and natural history of the genus *Lactobacillus*. *FEMS Microbiol Rev* **2017**;41(Supp_1):S27-S48 doi 10.1093/femsre/fux030.
35. Butt J, Werner S, Willhauck-Fleckenstein M, Michel A, Waterboer T, Zornig I, *et al.* Serology of *Streptococcus gallolyticus* subspecies *gallolyticus* and its association with colorectal cancer and precursors. *Int J Cancer* **2017**;141(5):897-904 doi 10.1002/ijc.30765.
36. Kumar R, Herold JL, Schady D, Davis J, Kopetz S, Martinez-Moczygemba M, *et al.* *Streptococcus gallolyticus* subsp. *gallolyticus* promotes colorectal tumor development. *PLoS Pathog* **2017**;13(7):e1006440 doi 10.1371/journal.ppat.1006440.
37. Liu Y LZ, Lin Y, Chen Y, Peng XE, He F, Liu S, Yan S, Huang L, Lu W, Xiang Z, Hu Z. *Streptococcus* and *Prevotella* are associated with the prognosis of oesophageal squamous cell carcinoma. *J Med Microbiol* **2018**;67(8):1058-68.
38. Allen-Vercoe E, Strauss J, Chadee K. *Fusobacterium nucleatum*: an emerging gut pathogen? *Gut Microbes* **2011**;2(5):294-8 doi 10.4161/gmic.2.5.18603.
39. Kolenbrander PE. Oral microbial communities: biofilms, interactions, and genetic systems. *Annu Rev Microbiol* **2000**;54:413-37 doi 10.1146/annurev.micro.54.1.413.
40. Shao D, Vogtmann E, Liu A, Qin J, Chen W, Abnet CC, *et al.* Microbial characterization of esophageal squamous cell carcinoma and gastric cardia adenocarcinoma from a high-risk region of China. *Cancer* **2019**;125(22):3993-4002 doi 10.1002/cncr.32403.
41. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, *et al.* Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Res* **2012**;22(2):292-8 doi 10.1101/gr.126573.111.
42. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/beta-catenin signaling via its FadA adhesin. *Cell Host Microbe* **2013**;14(2):195-206 doi 10.1016/j.chom.2013.07.012.

43. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, *et al.* *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* **2013**;14(2):207-15 doi 10.1016/j.chom.2013.07.007.
44. Nougayrede JP, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, Gottschalk G, *et al.* *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* **2006**;313(5788):848-51 doi 10.1126/science.1127059.
45. Ito M, Kanno S, Nosho K, Sukawa Y, Mitsuhashi K, Kurihara H, *et al.* Association of *Fusobacterium nucleatum* with clinical and molecular features in colorectal serrated pathway. *Int J Cancer* **2015**;137(6):1258-68 doi 10.1002/ijc.29488.
46. Appert-Collin A, Hubert P, Cremel G, Bennasroune A. Role of ErbB Receptors in Cancer Cell Migration and Invasion. *Front Pharmacol* **2015**;6:283 doi 10.3389/fphar.2015.00283.
47. Hellmann M, Nathanson T, Rizvi H, Creelan B, Sanchez-Vega F, Ahuja A, *et al.* Genomic Features of Response to Combination Immunotherapy in Patients with Advanced Non-Small-Cell Lung Cancer. *Cancer cell* **2018**;33(5):843-52.e4 doi 10.1016/j.ccell.2018.03.018.
48. Ready N, Hellmann MD, Awad MM, Otterson GA, Gutierrez M, Gainor JF, *et al.* First-Line Nivolumab Plus Ipilimumab in Advanced Non-Small-Cell Lung Cancer (CheckMate 568): Outcomes by Programmed Death Ligand 1 and Tumor Mutational Burden as Biomarkers. *J Clin Oncol* **2019**;37(12):992-1000 doi 10.1200/jco.18.01042.
49. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, *et al.* Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet* **2019**;51(2):202-6 doi 10.1038/s41588-018-0312-8.
50. Li L, Chen X, Hao L, Chen Q, Liu H, Zhou Q. Exploration of immune-related genes in high and low tumor mutation burden groups of chromophobe renal cell carcinoma. *Bioscience reports* **2020**;40(7) doi 10.1042/bsr20201491.

Tables

Table 1. Patient clinical characteristics and detection level of *F. nucleatum*

Clinical Characteristic	Total Number of Patients	Total Detect Number	Samples for detect <i>F. nucleatum</i>		
			[N=98]		<i>P</i> value
	Positive	Negative			
111	98	68(69.4%)	30(29.6%)		
Age (mean, range)					
62.9±7.1					
<60	31(27.9%)	29	22(75.9%)	7(24.1%)	0.1412
≥60	80(72.7%)	69	46(66.7%)	23(33.3%)	
Gender					
Male	64(57.7%)	56	39(69.6%)	17(30.4%)	0.9989
Female	47(42.3%)	42	29(69%)	13(31%)	
Smoking					
Yes	44(39.6%)	39	26(66.7%)	13(33.3%)	0.9428
No	67(60.4%)	59	42(71.2%)	17(28.8%)	
Drinking					
Yes	46(41.4%)	39	25(64.1%)	14(35.9%)	0.5561
No	65(58.6%)	59	43(72.9%)	16(28.8%)	
Histopathological grading					
G1-2	79(71.2%)	72	50(69.4%)	22(30.6%)	0.5305
G3	32(28.8%)	26	18(69.2%)	8(30.8%)	
clinical T stage (pT)					
pT1-2	44(39.6%)	35	24(68.6%)	11(31.4%)	0.039
pT3-4	67(60.4%)	63	44(69.8%)	19(30.2%)	
lymph node					
pN0-1	97(87.4%)	89	61(68.5%)	28(31.5%)	0.1343
pN2-3	14(12.6%)	9	7(77.8%)	2(22.2%)	
Clinical tumor stage					
I	8(7.2%)	4	3(75%)	1(25%)	0.0033

Figure 2

Microbial comparison of paired tumor and paired-normal tissues from patients with ESCC for α (A) and β diversity (B). The α diversity of 21 paired ESCC tumor and nontumor tissues was assessed with the Chao1 index, Shannon index, observed species and PD_whole_tree. The β diversity, based on the weighted UniFrac distance, was compared with PCoA plots for 21 paired tumor and nontumor tissues from patients with ESCC. PCoA, principal coordinate analysis. (C) The relative abundance of bacteria at the phylum level in both the normal and tumor groups of ESCC. (D) Differential abundance analysis

revealed that Bacteroidetes (P=0.03149), Fusobacteria (P=0.01796) and Spirochaetae (P=0.004155) were the top 3 significantly enriched taxa in the tumor group compared with the matched normal group samples. **P-value<0.01.

Figure 3

Microbiota associated with clinical tumor stage in ESCC. (A) Twenty-one pairs of samples according to their respective clinical stages and clustered trees based on the unweighted UniFrac distance matrix and

UPGMA method; the clustering results were integrated with the relative abundance of each sample species. The α diversity at the clinical pT stage (B) and clinical tumor stage (C) was assessed with the Chao1 index, Shannon index, observed species and PD_whole_tree.

Figure 4

Enrichment of *Fusobacterium nucleatum* in ESCC detected by FISH. Graph A: ESCC specimens stained with DAPI. Cell nuclei were stained in blue; B: ESCC specimens stained with both DAPI and a universal bacterial probe (EUB338). Bacterial conserved regions were stained in green; C: ESCC specimen stained with both DAPI and a *Fusobacterium*-specific probe (FUSO). The *F. nucleatum*-specific regions were stained in red; D: ESCC specimens triply stained with DAPI, FUSO and EUB338. Cell nuclei (in blue), bacterial conserved regions (in green) and *F. nucleatum*-specific regions (in red) were clearly visible (all $\times 400$).

Figure 5

Distribution of *F. nucleatum* in ESCC. (A) The relative abundance of *F. nucleatum* in tumor tissues was significantly higher than that in paired normal tissue samples in 21 patients with ESCC ($P=0.0262$). (B) The relative *F. nucleatum* levels in the pT3-4 stage were significantly higher than those in the pT1-2 stage ($P=0.039$), and the enrichment was higher in the III stage than in the I-II stage ($P=0.0039$).

Figure 6

(A) Gene Ontology (GO) enrichment results; (B) protein domain enrichment analysis results; (C) mutation frequencies and signatures, and significantly mutated genes in 20 ESCCs. The number of somatic

mutations of each examined case (top) and the significantly mutated genes (SMGs) colored by the types of mutations and their mutational frequency (bottom). Columns, examined cases; rows, genes; N, F. nucleatum-negative group; P, F. nucleatum-positive group. (D) Combination of the mutational burden and the F. nucleatum content to predict metastasis in ESCC. The red horizontal line indicates the cutoff score for mutational burden. Clinical and pathological features that include the presence or absence of distant metastases in the surgical specimen are annotated for each patient. A sample was considered F. nucleatum positive if the qPCR Ct value was <37 and the melt curve could be generated; otherwise, it was considered negative.

Supplementary Files

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

- [SupplementaryMaterial.docx](#)