Journal of Translational Medicine

RESEARCH

Open Access

Metagenomic characterization of oral microbiome signatures to predict upper gastrointestinal and pancreaticobiliary cancers: a case–control study

Sujin Oh[1], Jaihwan Kim[2], Cheol Min Shin[2], Hyo-Jung Lee[3], Hye Seung Lee[4,5,6*] and Kyoung Un Park[1,7*]

Abstract

Background This study investigated the oral microbiome signatures associated with upper gastrointestinal (GI) and pancreaticobiliary cancers.

Methods Saliva samples from cancer patients and age- and sex-matched healthy controls were analyzed using 16S rRNA-targeted sequencing, followed by comprehensive bioinformatics analysis.

Results Significant dissimilarities in microbial composition were observed between cancer patients and controls across esophageal cancer (EC), gastric cancer (GC), biliary tract cancer (BC), and pancreatic cancer (PC) groups ($R^2 = 0.067, = 0.075, = 0.068$, and $= 0.044$; $p = 0.001, = 0.001, = 0.002$, and $= 0.004$, respectively). Additionally, the oral microbiome composition significantly differed by the four cancer sites ($p = 0.001$ for EC vs. GC, EC vs. BC, EC vs. PC, GC vs. BC, and GC vs. PC; $p = 0.013$ for BC vs. PC). We built oral metagenomic classifiers to predict cancer and selected specific microbial taxa with diagnostic properties. For EC, the classifier differentiated cancer patients and controls with good accuracy (area under the curve [AUC] = 0.791) and included three genera: *Akkermansia*, *Escherichia-Shigella*, and *Subdoligranulum*. For GC, the classifier exhibited high discriminative power (AUC = 0.961); it included five genera (*Escherichia-Shigella*, *Gemella*, *Holdemanella*, *Actinomyces*, and *Stomatobaculum*) and three species (*Eubacterium sp. oral clone* EI074, *Ruminococcus sp. Marseille*-P328, and *Leptotrichia wadei* F0279). However, microbial taxa with diagnostic features for BC and PC were not identified.

Conclusions These findings suggested that the oral microbiome composition may serve as an indicator of tumorigenesis in upper GI and pancreaticobiliary cancers. The development of oral metagenomic classifiers for EC and GC demonstrates the potential value of microbial biomarkers in cancer screening.

Keywords Oral microbiome, Esophageal cancer, Gastric cancer, Biliary tract cancer, Pancreatic cancer, Microbial signatures

*Correspondence:
Hye Seung Lee
hye2@snu.ac.kr
Kyoung Un Park
m91w95pf@snu.ac.kr
Full list of author information is available at the end of the article

Introduction

Gastrointestinal (GI) tract cancers are present in over one-quarter of all cancer cases and cause over one-third of cancer-related deaths worldwide [1]. The five most common GI cancers include esophageal, gastric, colorectal, liver, and pancreatic cancers (PCs), together constituting more than 90% of all GI cancers; more than 60% of the cases and deaths occur in Asia [2, 3]. Although biliary tract cancer (BC) is relatively rare compared with these common GI cancers, its incidence and mortality rate are steadily increasing. The Republic of Korea has the highest BC-related mortality rate and the second highest incidence of BC [4, 5]. Previous studies have explored several risk factors and potential markers for GI cancers. However, a more precise and multifaceted understanding of GI cancers is required for improved disease prevention and treatment strategies.

The GI tract, extending from the oral cavity to the distal colon, is inhabited by an estimated 10^{14} microbial cells that outnumber human cells by at least tenfold [6]. The coding capacity of the microbes significantly exceeds that of the human genome. Consequently, human genetic features are now assumed to arise from the combination of microbial and human genomes [7]. The term "microbiome" refers to the cumulative genetic composition of microorganisms within various body habitats, which plays a crucial role in physiological processes ranging from host metabolism to immune reactions [8]. The fecal microbiome has been most widely investigated; the oral microbiome, the second largest and most diverse microbial ecosystem (approximately 770 species), has also been associated with human health [9, 10]. Notably, nitrate-reducing bacteria, including *Veillonella dispar* and *Actinomyces odontolyticus*, convert nitrate into nitrite, which is further metabolized into nitric oxide (NO). This pathway plays a crucial role in blood pressure regulation, antimicrobial defense, and oral health improvement by reducing inflammation and preventing caries [11, 12].

There has been increasing interest in the role of the human microbiome in human health, particularly because several studies have revealed associations between microbial alterations and systemic diseases [13–19]. Cancer has also drawn attention because *Helicobacter pylori* was classified as a class-I carcinogen by the World Health Organization based on its ability to promote stomach cancer after chronic infection [20, 21]. An emerging concept in cancer biology is that the microbiome constitutes an influential environmental factor modulating the carcinogenic process; tumorigenesis is presumably influenced by compositional instability in microbial communities. For example, researchers have demonstrated intestinal microbiome involvement in several cancer types, including colorectal, gastric, and liver cancers [14, 15]. Additionally, periodontopathogens, such as *P. gingivalis*, *T. forsythia*, and *T. denticola*, reportedly increase the risk of certain cancer types, including pancreatic, colon, and lung cancers, indicating an association between oral dysbiosis and tumorigenesis in organs distant from the oral cavity [17–19].

Despite mounting evidence regarding the roles of microbiomes in cancer development, the association between oral dysbiosis and GI cancer development remains poorly understood. The present study investigated the oral microbiome composition in patients with esophageal cancer (EC), gastric cancer (GC), BC, and PC. We hypothesized that upper GI and pancreaticobiliary cancer patients would exhibit distinctive oral microbiome characteristics compared with healthy individuals. Comparative analyses were also used to assess oral microbiome profiles among cancer patients with different primary tumor sites. Subsequently, oral metagenomic classifiers to predict upper GI and pancreaticobiliary cancers were constructed; we identified microbial taxa that could potentially serve as diagnostic markers for cancer patients.

Materials and methods

Study design

This study utilized a case–control design (Fig. 1) involving 42 EC, 19 GC, 14 BC, and 21 PC patients diagnosed at the Seoul National University Bundang Hospital (SNUBH) between December 2019 and February 2022. In accordance with best practices for observational studies, we have followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [22]. A completed STROBE checklist has been included in Supplementary Table 1 to ensure transparency and comprehensiveness in reporting. Samples were collected before patients underwent any surgical or therapeutic treatment. Clinicopathological data, including demographics, TNM cancer stage (American Joint Committee on Cancer, 8th edition), histologic subtype, and tumor differentiation, were collected. Pathological TNM stage was determined for patients who underwent surgical resection, whereas clinical staging was used for those who did not undergo surgical resection. Healthy controls were recruited from among visitors to the dental clinic of SNUBH, selected based on the absence of systemic disease. All potential participants were informed about the study's purpose and procedures. A questionnaire survey was conducted to gather information regarding disease history, including hypertension, diabetes, chronic infections, liver and kidney diseases, cardiovascular disease, autoimmune diseases, and malignant tumors, as well as the history of antibiotic use within the preceding 6 months. Electronic medical records, routine laboratory

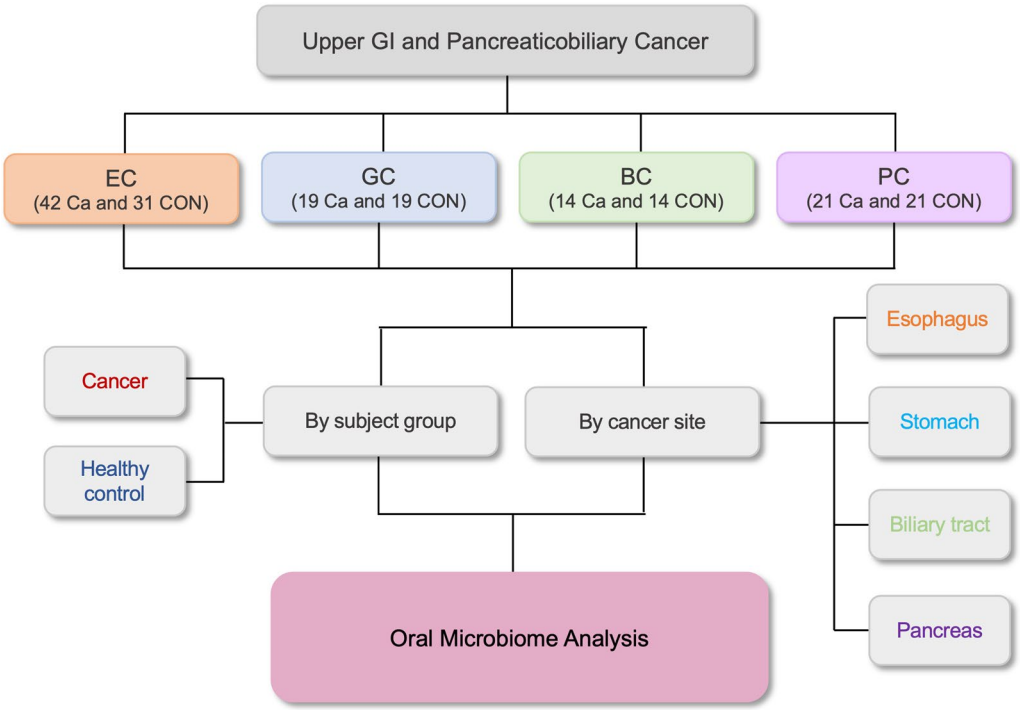

Fig. 1 Schematic of the study design. The study included 42 esophageal cancer (EC), 19 gastric cancer (GC), 14 biliary tract cancer (BC), and 21 pancreatic cancer (PC) patients, along with age- and sex-matched healthy controls [n = 31 (EC), = 19 (GC), = 14 (BC), and = 21 (PC)]. Comparative oral microbiome analyses were performed between each cancer and its matched control group, as well as among the cancer groups

test results, and dental charts were reviewed to verify participants' medical histories.3 Age- and sex-matched controls were selected for each cancer group [n = 31 (EC), = 19 (GC), = 14 (BC), and = 21 (PC)]. In addition to the aforementioned clinical variables, information about potential microbiome-related confounding factors was collected, including alcohol consumption, tobacco smoking, and periodontitis severity, assessed by dental experts and classified according to the Centers for Disease Control and Prevention-American Academy of Periodontology definition. However, dental inspections could not be performed on GC patients; periodontitis severity data were missing for this group in downstream analyses. The clinical characteristics of the patients and healthy controls are presented in Tables 1 and 2. Detailed information regarding periodontal status and oral hygiene is provided in Supplementary Table 2.

Sample collection and preparation

Participants were instructed to refrain from oral hygiene activities for at least 2 h before sample collection. Saliva was collected from the oral cavity for 20 min without stimulation prior to any dental procedures that could alter the oral microbiome. All specimens were immediately transported to the laboratory and stored at − 70 °C until DNA extraction. Microbial DNA was isolated from each specimen using the QIAamp DNA Microbiome Kit (QIAGEN, Venlo, The Netherlands), in accordance with the manufacturer's standard protocol.

16S rRNA-targeted sequencing

The PHiCS Institute (Seoul, Korea) performed 16S rRNA-targeted sequencing. DNA quality was assessed using Qubit dsDNA HS Assay Kits (Thermo Fisher Scientific, Waltham, MA, USA). Polymerase chain reaction (PCR) targeting the V3 and V4 hypervariable regions of 16S rRNA genes was conducted using KAPA HiFi HotStart ReadyMix PCR kits (Roche, Basel, Switzerland), in accordance with the manufacturer's instructions. The primer sequences used for PCR amplification were 519F: 5′-CCTACGGGNGGCWGCAG-3′ and 806R: 5′-GACTACHVGGGTATCTAATCC-3′. Libraries were constructed using Nextera XT DNA library preparation kits (Illumina, San Diego, CA, USA) and pooled to achieve a final loading concentration of 8 pM. Subsequently, paired-end (2 × 300 bp) sequencing was performed using the MiSeq platform (Illumina).

Microbiome data analysis

Taxonomic classification

Reads were processed using a Divisive Amplicon Denoising Algorithm (DADA)−2-based pipeline within the

Table 1 Clinicopathological characteristics of esophageal cancer (EC) and gastric cancer (GC) patients, and their age- and sex-matched healthy controls

Clinicopathological feature			EC				GC			
			Ca	CON	Total	P-value	Ca	CON	Total	P-value
Age (years)	Median (range)		68 (49–86)	68 (48–82)	67 (48–86)	0.232	67 (48–80)	68 (49–80)	67.5 (48–80)	0.965
Sex	N (%)	M	36 (85.7)	24 (77.4)	60 (82.2)	0.544	10 (52.6)	10 (52.6)	20 (52.6)	1.000
		F	6 (14.3)	7 (22.6)	13 (17.8)		9 (47.4)	9 (47.4)	18 (47.4)	
Weight (kg)	Median (range)		63.2 (39.8–81.7)	75.2 (52.45–91.1)*	68.0 (39.8–91.1)	0.006	54.0 (36.9–71.8)	68.2 (55.2–86.0)*	57.5 (36.9–86.0)	0.008
BMI (kg/m2)	Median (range)		23.3 (14.6–27.8)	25.5 (18.8–29.3)*	24.1 (14.6–29.3)	0.009	20.9 (13.9–26.6)	24.9 (21.0–29.1)*	22.1 (13.9–29.1)	0.003
Alcohol consumption	N (%)	Yes	21 (63.6)	15 (53.6)	36 (59.0)	0.593	10 (55.6)	6 (46.2)	16 (51.6)	0.879
		No	12 (36.4)	13 (46.4)	25 (41.0)		8 (44.4)	7 (53.8)	15 (48.4)	
		NA	9	3	12		0	0	0	
Tobacco smoking	N (%)	Yes	7 (21.2)	3 (10.0)	10 (15.9)	0.308	3 (16.7)	0 (0.0)	3 (8.3)	0.229
		No	26 (78.8)	27 (90.0)	53 (84.1)		15 (83.3)	18 (100.0)	33 (91.7)	
		NA	9	1	10		1	1	2	
Severity of periodontitis†	N (%)	Mild	0 (0.0)	1 (3.7)	1 (1.5)	0.008	0 (NA)	2 (11.8)	2 (11.8)	NA
		Moderate	4 (10.3)	10 (37.0)	14 (21.2)		0 (NA)	9 (52.9)	9 (52.9)	
		Severe	35 (89.7)	16 (59.3)	51 (77.3)		0 (NA)	6 (35.3)	6 (35.3)	
		NA	3	4	7		19‡	2	21	
Histologic subtype	N (%)	Adenoca	0 (0.0)				19 (100.0)			
		Sqcc	41 (100.0)				0 (0.0)			
		NA§	1				0			
Tumor differenciation	N (%)	WD	4 (13.3)				1 (5.6)			
		MD	3 (10.0)				4 (22.2)			
		PD	23 (76.7)				13 (72.2)			
		NA	12				1			
Lymph node metastasis	N (%)	Negative	27 (65.9)				11 (61.1)			
		Positive	14 (34.1)				7 (38.9)			
		NA	1				1			
Distant metastasis	N (%)	M0	41 (97.6)				16 (88.9)			
		M1	1 (2.4)				2 (11.1)			
		NA	0				1			

Ca cancer patients, *CON* controls, *M* male, *F* female, *NA* not available, *Adenoca* adenocarcinoma, *Sqcc* squamous cell carcinoma, *WD* well-differentiated, *MD* moderately differentiated, *PD* poorly differentiated

*Weight and BMI data were not available for the subset of healthy controls; 11 and 8 were missing for CON of EC and GC patients, respectively

† Severity of periodontitis was classified according to the Centers for Disease Control and Prevention-American Academy of Periodontology (CDC/AAP) definition

‡ Dental inspections of GC patients could not be performed. Therefore, periodontitis severity in GC patients was unknown

§ Tumors that could not be classified as Adenoca or Sqcc were regarded as NA. Otherwise, NA indicates missing or undetermined data

Table 2 Clinicopathological characteristics of biliary tract cancer (BC) and pancreatic cancer (PC) patients, and their age- and sex-matched healthy controls

Clinicopathological feature			BC				PC			
			Ca	CON	Total	P-value	Ca	CON	Total	P-value
Age (years)	Median (range)		69 (53–82)	69.5 (51–82)	69 (51–82)	0.963	70 (51–83)	68 (51–82)	69.5 (51–83)	0.772
Sex	N (%)	M	5 (35.7)	5 (35.7)	10 (35.7)	1.000	11 (52.4)	11 (52.4)	22 (52.4)	1.000
		F	9 (64.3)	9 (64.3)	18 (64.3)		10 (47.6)	10 (47.6)	20 (47.6)	
Weight (kg)	Median (range)		59.8 (45.5–80.5)	65.6 (52.5–86.0)*	60.7 (45.5–86.0)	0.244	55.4 (40.8–87.3)	66.0 (42.2–77.2)*	57.6 (40.9–87.3)	0.076
BMI (kg/m^2)	Median (range)		22.4 (17.9–29.0)	24.9 (18.8–29.1)*	23.0 (17.9–29.1)	0.125	21.9 (17.8–27.1)	24.9 (19.7–28.7)*	22.5 (17.8–28.7)	0.006
Alcohol consumption	N (%)	Yes	4 (30.8)	7 (50.0)	11 (40.7)	0.533	6 (33.3)	8 (44.4)	14 (38.6)	0.732
		No	9 (69.2)	7 (50.0)	16 (59.3)		12 (66.7)	10 (55.6)	22 (61.1)	
		NA	0	0	0		0	0	0	
Tobacco smoking	N (%)	Yes	2 (15.4)	2 (14.3)	4 (14.8)	1.000	1 (5.6)	1 (4.8)	2 (5.1)	1.000
		No	11 (84.6)	12 (85.7)	23 (85.2)		17 (94.4)	20 (95.2)	37 (94.9)	
		NA	1	0	1		3	0	3	
Severity of periodontitis[†]	N (%)	Mild	0 (0.0)	0 (0.0)	0 (0.0)	1.000	0 (0.0)	0 (0.0)	0 (0.0)	1.000
		Moderate	6 (46.2)	6 (54.5)	12 (50.0)		9 (45.0)	7 (50.0)	16 (47.1)	
		Severe	7 (53.8)	5 (45.5)	12 (50.0)		11 (55.5)	7 (50.0)	18 (52.9)	
		NA	1	3	4		1	7	8	
Histologic subtype	N (%)	Adenoca	14 (100.0)				19 (90.5)			
		Sqcc	0 (0.0)				0 (0.0)			
		NA[‡]	0				2 (9.5)			
Tumor differenciation	N (%)	WD	3 (30.0)				0 (0.0)			
		MD	5 (50.0)				8 (80.0)			
		PD	2 (20.0)				2 (20.0)			
		NA	4				11			
Lymph node metastasis	N (%)	Negative	6 (42.9)				8 (38.1)			
		Positive	8 (57.1)				13 (61.9)			
		NA	0				0			
Distant metastasis	N (%)	M0	14 (100.0)				17 (81.0)			
		M1	0 (0.0)				4 (19.0)			
		NA	0				0			

Ca cancer patients, *CON* controls, *M* male, *F* female, *NA* not available, *Adenoca* adenocarcinoma, *Sqcc* squamous cell carcinoma, *WD* well-differentiated, *MD* moderately differentiated, *PD* poorly differentiated

*Weight and BMI data were not available for the subset of healthy controls; 3 and 8 were missing for CON of BC and PC patients, respectively

[†] Severity of periodontitis was classified according to the Centers for Disease Control and Prevention-American Academy of Periodontology (CDC/AAP) definition

[‡] Tumors that could not be classified as Adenoca or Sqcc were regarded as NA. Otherwise, NA indicates missing or undetermined data

QIIME2 platform [23, 24]. An amplicon sequencing variant (ASV) table was produced through quality-based filtering and trimming, read deduplication, and ASV inference, followed by paired-end merging and chimera removal. For taxonomic analysis of microbial composition, sequences were taxonomically classified against the 99% SILVA rRNA taxonomy using a pre-trained scikit-learn naive Bayes machine-learning classifier within the QIIME q2-feature-classifier plugin [25, 26].

Diversity analysis

The following analyses were conducted using R software (version 4.1.2; R Development Core Team, Vienna, Austria). QIIME artifacts were imported into the R environment using the *qiime2R* package [27]. To address artifactual biases, feature tables were normalized by rarefaction. Alpha and beta diversity indices were calculated using the *phyloseq* package [28]. Shannon entropy was used for alpha diversity analysis. Beta diversity indices, including Bray–Curtis and unweighted UniFrac distance matrices, were computed to estimate dissimilarities in microbial composition between samples. Principal coordinates analysis was used to visualize broad trends of sample dissimilarities, whereas permutation multivariate analysis of variance (PERMANOVA) and distance-based redundancy analysis (db-RDA) were conducted to quantify the explanatory power of clinical variables with respect to microbial community variance using the *vegan* package [29].

Differential abundance (DA) testing

To identify differentially abundant genera and species in cancer patients compared with their matched controls, we conducted Wilcoxon rank-sum tests on centered log-ratio (CLR)-transformed taxonomic abundances. Before analysis, microbial taxa with prevalence < 0.05 across all samples were removed. DA tests on taxonomic abundance were also performed among cancer patients to identify any microbial taxa that varied among the patients with four different tumor sites. All pairwise comparisons were conducted between pairs of cancer patient groups. Multiple testing correction was performed using the Benjamini-Hochberg (BH) method, and taxa were considered differentially abundant when q < 0.05. Effect size was calculated as the mean of the decile differences of CLR abundance between groups.

Multivariable modeling

To obtain multivariable models for the classification of cancer patients and healthy individuals, multivariable logistic regression models with L1 (LASSO) regularization were constructed on CLR-transformed microbial abundances using the *glmnet* R package [30]. Abundance tables were initially filtered by selecting the univariably associated microbial taxa identified through DA analysis. Data were randomly split into test and training sets in 10 repeated tenfold cross-validation. For each split, the regression model was trained on the training sets, and the trained model was then used to predict the left-out test set. Lambda parameters were selected to minimize mean binomial deviance, ensuring that more than three non-zero coefficients were retained for the final models. In addition to the binomial microbial classifiers for discriminating between EC, GC, BC, and PC and their control groups, a multinomial logistic regression model was fitted to determine the optimal variable combination for discriminating among the four cancer groups. This was achieved through filtration, normalization, and LASSO variable selection procedures, as described above.

Statistical analysis

CLR transformation of raw feature counts was performed to enable the application of conventional statistical techniques to microbial abundance data. Associations between categorical variables in contingency tables were assessed using either the Chi-square test or Fisher's exact test. For continuous variables, nonparametric differences were analyzed using either the Wilcoxon rank-sum test or the Kruskal–Wallis H test. All statistical tests were two-sided, and p-values < 0.05 were considered statistically significant. Correction for multiple testing was carried out using the BH method, and statistical significance was determined based on a q-value threshold of < 0.05.

Multivariable models were evaluated using the *pROC* and *multiROC* R packages for binomial and multinomial logistic regression models, respectively [31]. Sensitivities, specificities, and area under the curve (AUC) of the receiver operating characteristic curves for the classifications were calculated. Evaluation metrics for multiclass classification were computed by reducing the multiclass predictions to multiple sets of binary predictions in a one vs. rest manner, followed by macro- and micro-averaging across all groups. The 95% confidence interval (CI) of the AUC was computed using 2000 bootstrap replicates. Moreover, the predicted probability of cancer derived from the abundance of each taxon within the models was estimated by maintaining other variables at their mean values.

Results

Associations of oral microbiome with upper GI and pancreaticobiliary cancers

We investigated the microbial compositions in the oral cavities of patients with EC, GC, BC, and PC, along with their matched healthy controls. The predominant phylum in both cancer and control groups was

Firmicutes, followed by *Actinobacteria* and *Proteobacteria* (Fig. 2A), representing 48.0%, 20.6%, and 13.9% of healthy controls and 52.1%, 23.0%, and 11.0% of cancer patients, respectively. At the genus level, *Streptococcus* and *Rothia* were the most abundant taxa across all control and cancer groups, representing 24.1% and 16.9% of all healthy individuals and 27.8% and 19.8% of all patients, respectively.

Although dominant taxa remained consistent between control and cancer groups, diversity analyses revealed microbial shifts in the oral cavity associated with upper GI and pancreaticobiliary cancers. A comparison of Shannon index values showed that EC patients exhibited greater alpha diversity compared with healthy individuals ($p=0.002$, Wilcoxon). Similarly, GC patients showed increased microbial richness in the oral cavity compared with controls ($p=0.011$, Wilcoxon), whereas no significant differences were observed between patients and healthy controls in the BC and PC groups ($p=0.1$ and $=0.58$, respectively, Wilcoxon; Fig. 2B).

To explore dissimilarities in microbial composition between cancer and corresponding control groups, beta diversity analysis was conducted using the Bray–Curtis distance. Although baseline characteristics were balanced between cancer cases and controls (Tables 1 and 2), we investigated potential confounders, including age, sex, alcohol consumption, smoking habits, and periodontitis severity, to adjust for covariates as appropriate (Supplementary Table 3). Microbial composition differed according to periodontitis severity in EC patients and their control group ($p=0.003$, $R^2=0.052$, PERMANOVA), whereas tobacco smoking contributed to 4.7% of microbial variance in PC patients and their matched controls. Principal coordinates analysis and PERMANOVA of the Bray–Curtis distance revealed significant dissimilarities in microbial composition between cancer patients and controls; disease status explained 6.7%, 7.5%, 6.8%, and 4.4% of the total oral microbiome variance in the four cancer types, respectively ($p=0.001$, $=0.001$, $=0.002$, and $=0.003$, respectively; Fig. 2B). Additionally, despite adjustment for covariates, oral microbiome differences persisted between cancer patients and healthy individuals [EC: $p=0.001$ (adjusted for periodontitis severity, PERMANOVA), PC: $p=0.004$ (adjusted for tobacco smoking, PERMANOVA)].

We next examined community-level microbial diversity in cancer cases to determine potential differences in oral microbiome composition among patients with four distinct cancer sites. Regarding alpha diversity, significant variations were observed in Shannon entropy of the oral microbiome among each of the cancer groups, except between EC and GC and between BC and PC ($p=0.007$ for EC vs. BC, $p=0.02$ for EC vs. PC, $p=0.009$ for GC vs. BC, and $p=0.044$ for GC vs. PC, Wilcoxon; Fig. 3B). EC and GC patients exhibited higher microbial richness levels in their oral cavities compared with BC and PC patients. Additionally, primary sites of upper GI and pancreaticobiliary cancers were correlated with dissimilarities in oral microbiome composition ($R^2=0.130$, $p=0.001$, PERMANOVA); pairwise differences remained statistically significant for all cancer groups ($p=0.001$ for EC vs. GC, EC vs. BC, EC vs. PC, GC vs. BC, and GC vs. PC; $p=0.013$ for BC vs. PC, PERMANOVA; Fig. 3C and D). Considering the potential effect of patient heterogeneity on our findings, we explored associations between microbial variance and other clinical features. We found that tobacco smoking habit (yes/no) and presence of distant metastasis (M0/M1) also contributed to salivary microbial variance ($R^2=0.020$ and $=0.018$, respectively, $p=0.019$, PERMANOVA; Fig. 3C, Supplementary Table 4). Despite adjustment for smoking and M stage as covariates to quantify the marginal effects of primary site on beta diversity in cancer patients, tumor sites continued to explain 12.8% of total variance in the oral microbiome (Supplementary Table 4).

Disease-associated oral microbial taxa in upper GI and pancreaticobiliary cancer patients

Building on the associations of community-level diversity of the oral microbiome with upper GI and pancreaticobiliary cancers, we conducted DA analyses to identify cancer-associated genera and species ($q<0.05$, BH-adjusted, Wilcoxon). Our findings revealed significant differences in the abundances of eight and 14

(See figure on next page.)

Fig. 2 Comparison of community-level oral microbiome characteristics between upper gastrointestinal (GI) and pancreaticobiliary cancer patients and healthy controls. **A** Genus and species compositions of the oral microbiome in each group [top left: esophageal cancer (EC), top right: gastric cancer (GC), bottom left: biliary tract cancer (BC), bottom right: pancreatic cancer (PC); CON, controls; Ca, cancer patients]. Phyla and genera representing less than 1% and less than 2.5% of the microbial community are shown as < 1% and < 2.5%, respectively. **B** Alpha (Shannon entropy) and beta (Bray–Curtis distance) diversity, and their associations with disease status [top left: EC, top right: GC, bottom left: BC, bottom right: PC (*$p<0.05$; **$p<0.01$; ns, not significant, Wilcoxon)]. R^2 value indicates the proportion of variance explained by disease status in each group, as confirmed by PERMANOVA

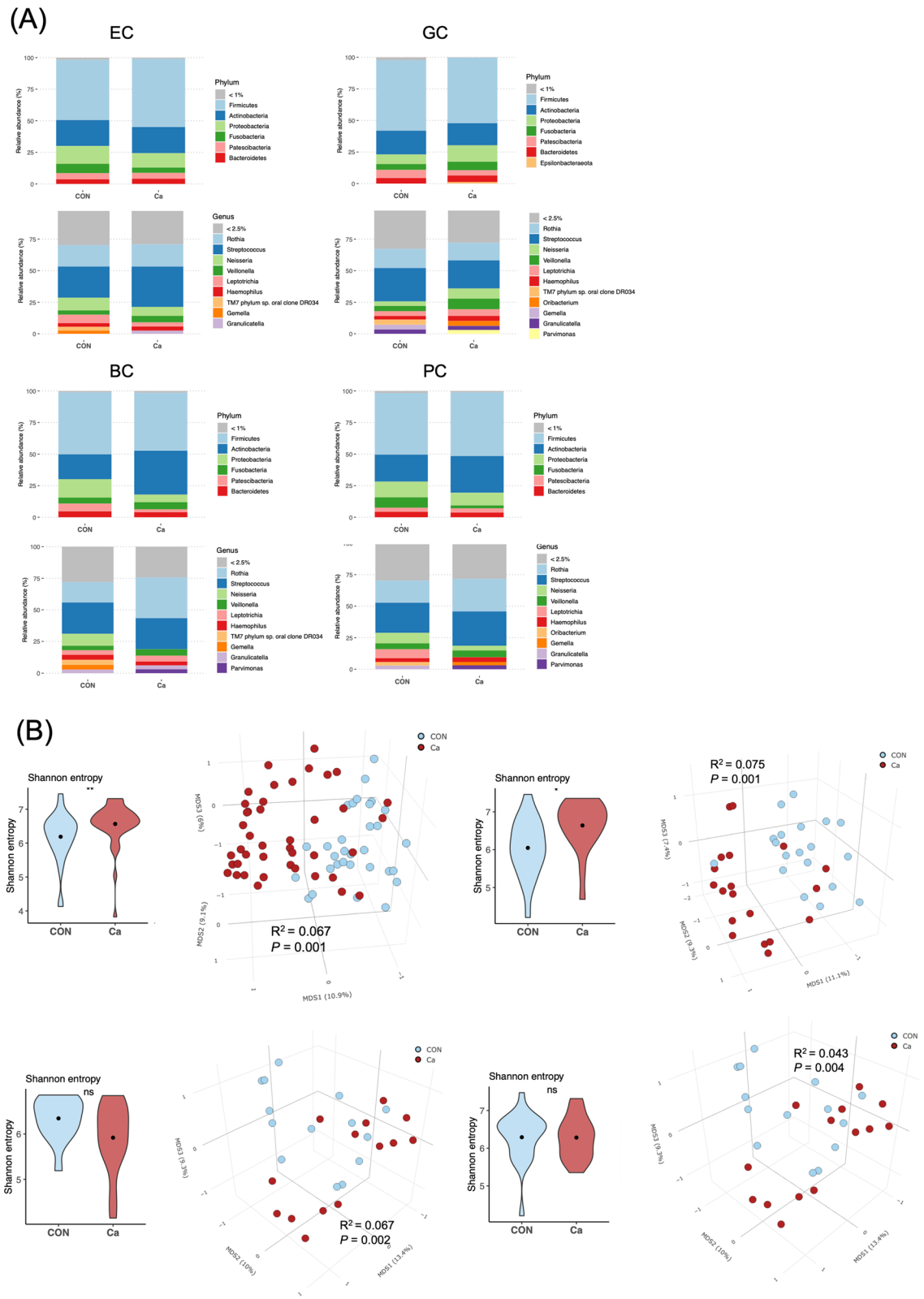

Fig. 2 (See legend on previous page.)

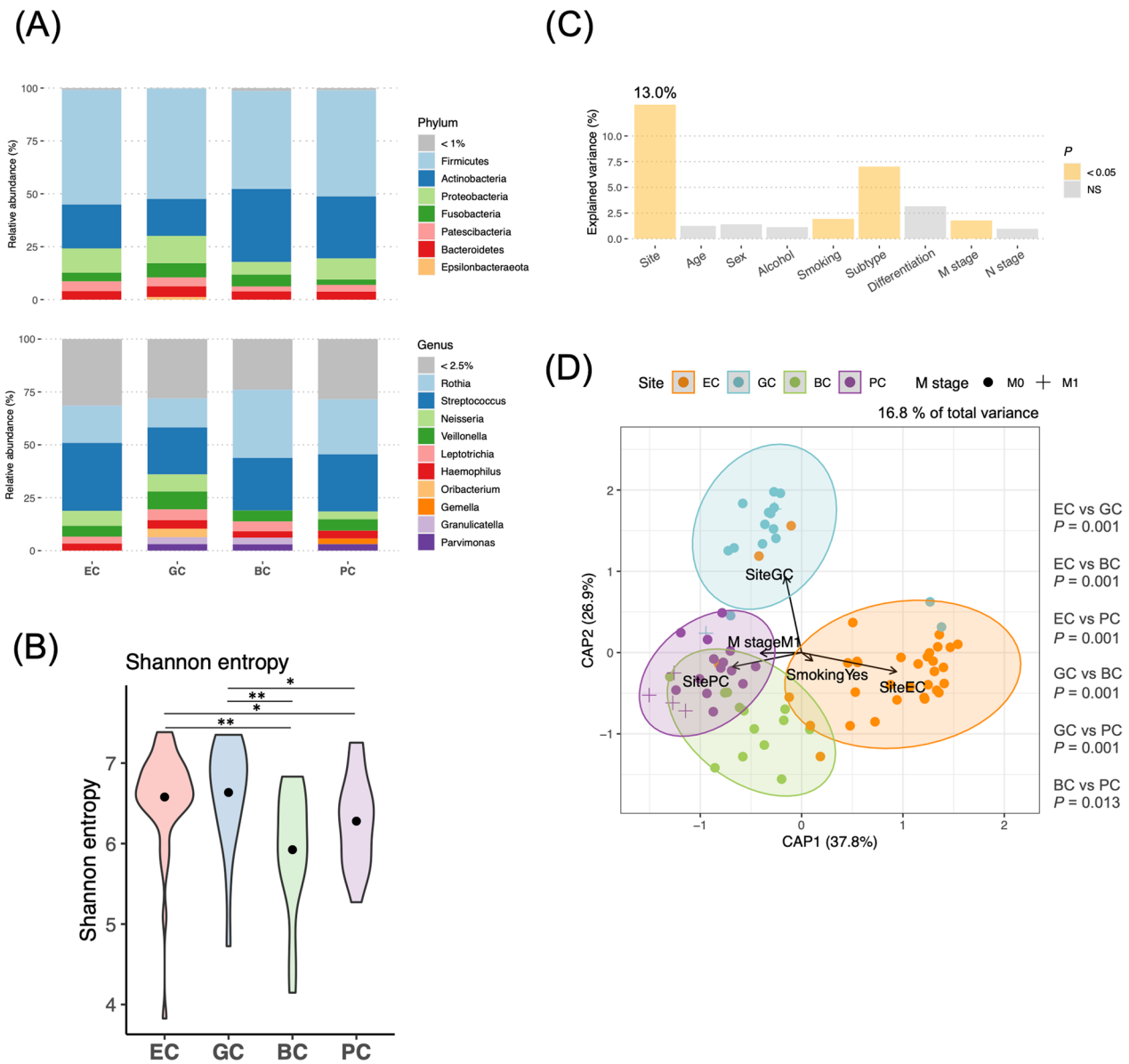

Fig. 3 Comparison of community-level oral microbiome characteristics among esophageal cancer (EC), gastric cancer (GC), biliary tract cancer (BC), and pancreatic cancer (PC) patients. **A** Genus and species compositions of the oral microbiome in EC, GC, BC, and PC patients. Phyla and genera representing less than 1% and less than 2.5% of the microbial community are shown as < 1% and < 2.5%, respectively. **B** Differences in alpha diversity (Shannon entropy) among cancer groups (*$p < 0.05$, **$p < 0.01$, Wilcoxon). **C** Associations between beta diversity (Bray–Curtis distance) and clinicopathological variables in cancer patients. Bars represent the percentage of explained variance relative to the total microbial variance, and the orange color indicates statistical significance ($p < 0.05$, PERMANOVA). **D** Bray–Curtis distance-based redundancy analysis (db-RDA) of cancer patients. Variance constrained by the clinical variables included in db-RDA explained 16.8% of the total microbial variance. Variables are illustrated using arrows, where arrow length indicates the contribution of each variable to microbial variance. The colors of dots and ellipses represent tumor sites

genera in EC and GC patients, respectively, compared with their respective controls. However, one genus was differentially abundant in BC patients, and none was identified in PC patients. At the species level, we detected two and six species with significant DA in EC and GC patients, respectively, compared with healthy controls; none showed significant DA in BC and PC groups (Fig. 4A and B).

Seven genera (*Akkermansia*, *Parabacteroides*, *Blautia*, *Collinsella*, *Escherichia-Shigella*, *Subdoligranulum*, and *Fusicatenibacter*) and two species (*Bacteroides plebeius* and *Parabacteroides merdae*) were consistently

decreased in both BC and GC patients compared with healthy individuals. Meanwhile, we also observed microbial taxa with disease associations specific to certain cancers. For example, *Ruminococcaceae UCG-013* showed a significant decrease in EC patients compared with controls. Additionally, five genera (*Gemella*, *Paraprevotella*, *Holdemanella*, *Actinomyces*, and *Stomatobaculum*) and four species (*Eubacterium sp. oral clone* EI074, *Ruminococcus sp. Marseille*-P328, *Bacteroides dorei*, and *Leptotrichia wadei* F0279) were significantly decreased, whereas two genera (*Actinomyces* and *Stomatobaculum*) were enriched in GC patients (Fig. 4C).

We observed several microbial taxa in the EC and GC groups whose direction of change in cancer patients compared with healthy subjects was opposite to that in the BC and PC groups (Fig. 4C). *Akkermansia*, *Parabacteroides*, *Blautia*, *Collinsella*, *Escherichia-Shigella*, *Subdoligranulum*, and *Fusicatenibacter* were decreased in EC and GC patients, whereas they showed increasing trends in BC and PC groups, compared with healthy individuals. Conversely, *Actinomyces*, *Neisseria*, and *Stomatobaculum* were significantly or tended to be enriched in EC and GC patients, but were diminished in BC and PC patients, compared with controls.

Furthermore, we conducted comparative analyses to identify oral microbial taxa that varied among cancer patients with different primary tumor sites. Intriguingly, no taxa showed DA between EC and GC patients or between BC and PC patients. However, 15 and 23 genera were differentially abundant in EC and GC patients, respectively, compared with BC patients. We also detected three and 10 genera that were significantly reduced in EC and GC patients, respectively, compared with PC patients. *Subdoligranulum* and *Bifidobacterium* were among the oral genera with the greatest differences between cancer groups, followed by *Rothia*, *Bacteroides*, and *Blautia*. Similarly, the top five species exhibiting DA were *Bacteroides dorei*, *Parabacteroides merdae*, *Bacteroides plebeius*, *Streptococcus intermedius*, and *Rothia dentocariosa* (Fig. 4D). Detailed DA testing results for the cancer groups are provided in Supplementary Table 5.

Oral metagenomic classification models of upper GI and pancreaticobiliary cancers

To consolidate univariably associated microbial taxa into a comprehensive model and select predictive microbial taxa while eliminating less informative ones, we constructed oral microbiome-based classifiers using LASSO logistic regression models. Because DA analysis did not identify any significant taxa in PC patients and only one taxon in BC patients compared with controls, multivariable modeling was conducted for EC and GC groups. The classifier for EC effectively discriminated between cancer patients and healthy individuals with good accuracy [AUC = 0.791 (95% CI: 0.679–0.892)]. It included three genera, *Akkermansia*, *Escherichia-Shigella*, and *Subdoligranulum*, for which increased abundance reduced the probability of cancer (Fig. 5A). Similarly, the binomial model for differentiating GC patients from healthy controls exhibited high discriminative power [AUC = 0.961 (95% CI: 0.889–1)]. Variables with non-zero coefficients in this model comprised five genera (*Escherichia-Shigella*, *Gemella*, *Holdemanella*, *Actinomyces*, and *Stomatobaculum*) and three species (*Eubacterium sp. oral clone* EI074, *Ruminococcus sp. Marseille*-P328, and *Leptotrichia wadei* F0279). An increase in *Stomatobaculum*, *Actinomyces*, or *Leptotrichia wadei* F0279 increased the risk of GC, whereas increases in the remaining taxa decreased the risk (Fig. 5B).

Moreover, a multinomial logistic regression model was obtained to identify the optimal combination of microbial taxa for discriminating among cancer patients with four primary sites. Of the 32 genera and 14 species with univariable associations, the model selected 18 genera and eight species, achieving high accuracy for classifying cancer groups [macro-average AUC = 0.950 (95% CI: 0.912–0.972), micro-average AUC = 0.948 (95% CI: 0.917–0.979)]. The coefficients of the variables are presented in Fig. 6B.

(See figure on next page.)

Fig. 4 Differentially abundant microbial taxa in the oral microbiome of upper gastrointestinal (GI) and pancreaticobiliary cancer patients. **A**–**C** Genera and species exhibiting significant differences between esophageal cancer (EC), gastric cancer (GC), biliary tract cancer (BC), and pancreatic cancer (PC) patients and their control groups. Venn diagrams show significant genera (**A**) and species (**B**) with disease associations in EC, GC, and BC patients relative to healthy individuals, some of which were common between EC and GC patients. Heatmap illustrates the results of DA tests, highlighting the directions of changes in microbial abundance in cancer patients compared with controls. The effect size was calculated as the mean of the decile differences of centered log-ratio (CLR) abundance between two groups; an effect size > 0 indicated that a taxon was increased in patients compared with healthy individuals. **D** Genera and species exhibiting significant differences among cancer groups. The five most differentially abundant taxa are shown in the plots. Correction for multiple testing was performed using the Benjamini–Hochberg method, and statistical significance was determined as $q < 0.05$ (*$q < 0.05$, **$q < 0.01$, ***$q < 0.001$, Wilcoxon)

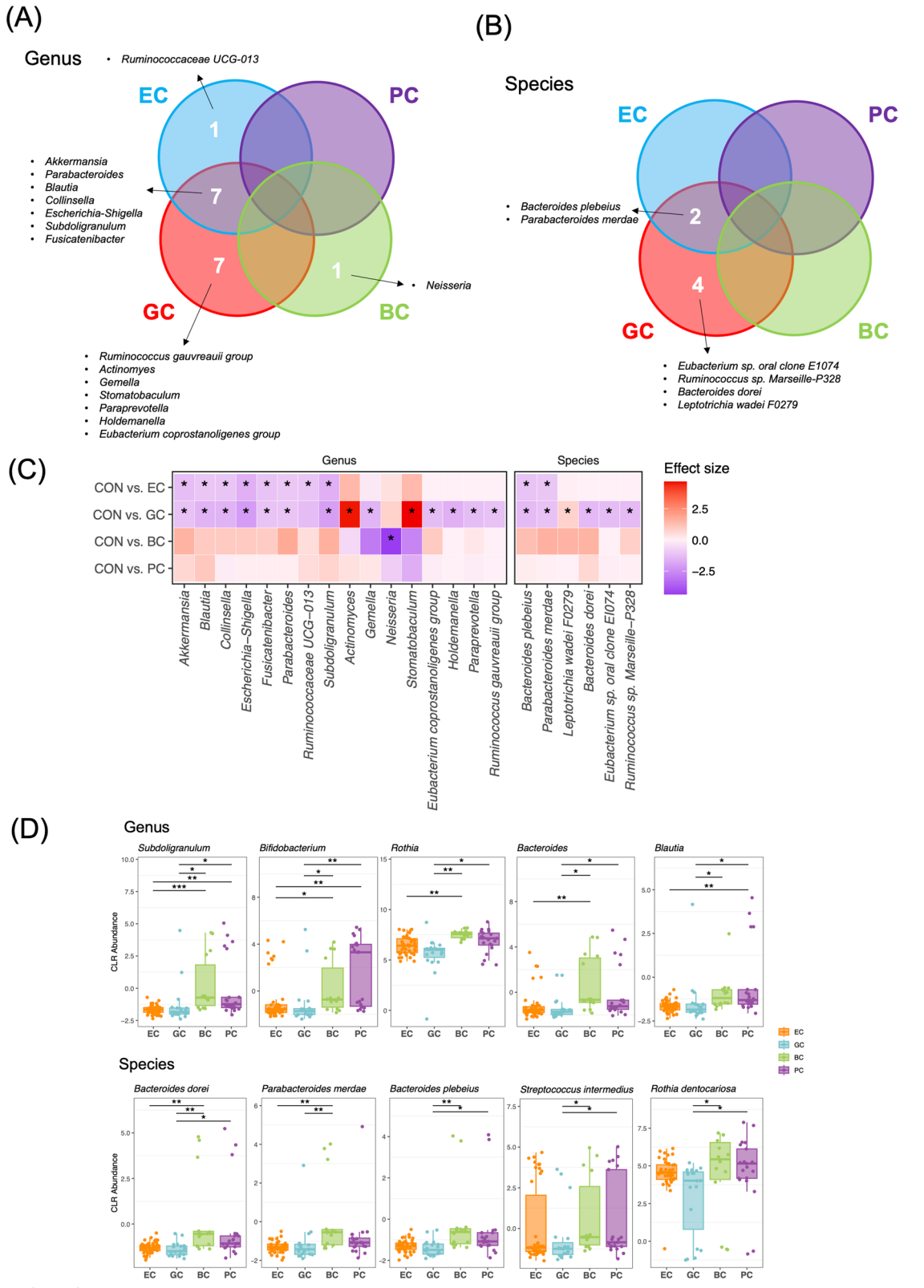

Fig. 4 (See legend on previous page.)

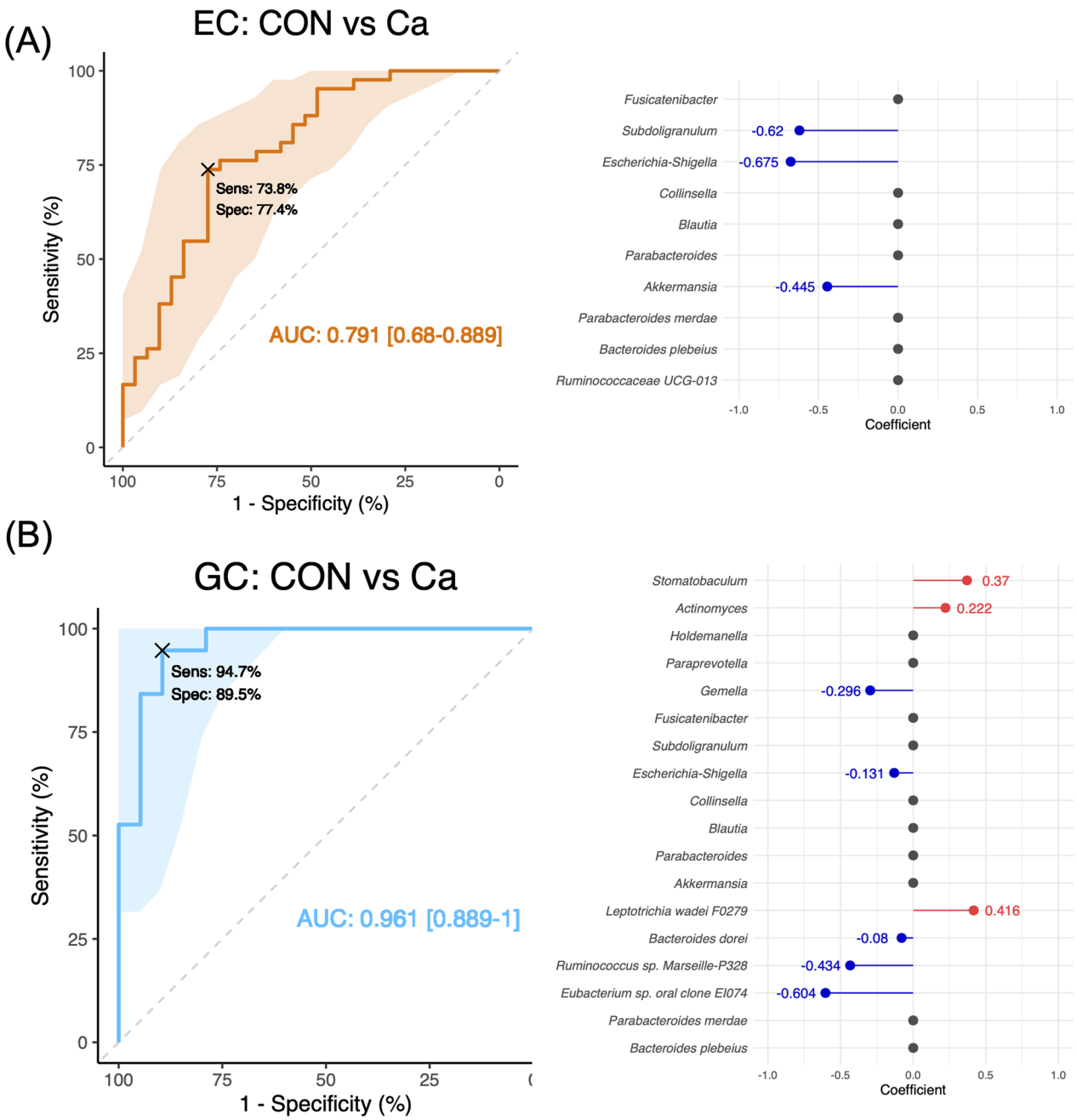

Fig. 5 Performances of oral metagenomic models in the classification of cancer patients and healthy individuals, obtained through LASSO logistic regression analysis with cross-validation. **A**, **B** Left panel shows classifier predictive accuracies as receiver operating characteristic (ROC) curves, with 95% confidence intervals (CIs) shaded. Values in parentheses represent CIs for area under the curve (AUC) values. Black 'X' markers indicate the optimal cutoff points, where the models achieved 73.8% sensitivity and 77.4% specificity (**A**), and 94.7% sensitivity and 89.5% specificity (**B**). Right panel shows the coefficients of selected variables in the model. Red, blue, and gray colors indicate positive, negative, and zero coefficients, respectively. **A**: esophageal cancer (EC); **B**: gastric cancer (GC); CON, controls; Ca, cancer patients; Sens, sensitivity; Spec, specificity

Discussion

To our knowledge, this study represents the first comprehensive analysis of the oral microbiome in upper GI and pancreaticobiliary cancer patients. We investigated the oral microbiome signatures of EC, GC, BC, and PC by 16S rRNA-targeted sequencing of saliva samples from cancer patients and their age- and sex-matched healthy controls. Our study demonstrated associations between oral dysbiosis and upper GI and pancreaticobiliary cancers (Figs. 2 and 3). The oral microbiome can

be influenced by various demographic and biological factors, leading to inter-individual variability in microbial composition; this variability may explain the lack of consistent findings across studies [32, 33]. Therefore, we examined the extent to which confounding factors, such as smoking habits and periodontitis severity, explained microbial variance to identify genuine associations between oral microbial variance and cancer. Comparisons between cancer patients and healthy individuals revealed significant decreases in alpha diversity in EC and GC patients, as well as beta diversity differences across all cancer types, compared with their respective controls, both with and without adjustment for covariates (Supplementary Tables 3 and 4). Furthermore, significant differences in the oral microbiome were observed among cancer patients with different tumor sites; microbial richness was higher in EC and GC patients than in BC and PC patients, and dissimilarities in composition were present among the groups.

EC is the seventh most common cancer type and the sixth most common cause of cancer mortality worldwide; more than half of the cases occur in Asia [3]. Because EC is often asymptomatic during the early stages, many cases are diagnosed at an advanced stage, contributing to high mortality rates [34]. Considering the need for early detection markers, previous studies have explored the association between the salivary microbiome and EC risk, yielding potential microbiological markers for EC [35, 36]. Our study successfully identified characteristic microbial taxa in the oral microbiome of EC patients (Figs. 4 and 5). Specifically, *Akkermansia*, *Escherichia-Shigella*, and *Subdoligranulum* exhibited predictive features; a multivariate model achieved an AUC of 79% in classifying EC patients and healthy individuals. A decrease in abundance among these genera corresponded to an increase in the probability of EC. Notably, *Akkermansia*, a probiotic isolated from the human intestine, has been detected in saliva and inversely correlated with inflammatory diseases [37–40]. Its presence, particularly that of *Akkermansia muciniphila*, appears to downregulate the production of inflammatory cytokines associated with periodontal and systemic inflammation (such as interleukin-10 and interleukin-12) [39] and stimulate anticancer immune responses targeting the PD-1/PD-L1 interaction [41]. The lower abundance of *Subdoligranulum*, a butyrate-producing bacterium, has been associated with various diseases, including type 2 diabetes and inflammatory bowel disease (IBD). Its co-occurrence with *Akkermansia* species has also been demonstrated in several studies [42–44].

Although *H. pylori* is a key pathogen in gastric carcinogenesis, only 3% of *H. pylori*-positive individuals develop GC, and *H. pylori* colonization is dramatically reduced in cancerous lesions. These findings imply the involvement of other microbes in GC progression [21, 45, 46]. Our results suggested that the oral microbiome plays a role in GC pathogenesis, consistent with previous reports [47–50]. We successfully constructed oral metagenomic classifiers with high accuracy (AUC: 96%) for predicting GC. *Escherichia-Shigella*, *Gemella*, *Holdemanella*, *Actinomyces*, *Stomatobaculum*, *Eubacterium sp. oral clone* EI074, *Ruminococcus sp. Marseille*-P328, and *Leptotrichia wadei* F0279 demonstrated diagnostic properties. In particular, *Gemella* and *Holdemanella*, the reduction of which appeared to increase GC risk, have been highlighted for their health implications in previous studies. *Gemella*, an indigenous anaerobic bacterium in the oral cavity, has been described as a component of the symbiotic microbial flora because its abundance was higher in healthy controls than in periodontitis patients [51, 52]. *Holdemanella biformis* has exhibited antitumor activity by reducing tumor cell proliferation through mechanisms involving histone acetylation and nuclear factor of activated T cells-3 [53].

Pancreaticobiliary cancer, although less common than EC and GC, lacks diagnostic markers other than cancer antigen 19–9 for early detection; thus, most patients are diagnosed with metastatic or locally advanced cancer, precluding curative surgery [3, 54]. Despite its relatively low cancer morbidity, PC has been widely studied for its association with oral dysbiosis [19, 55–58]. Salivary microbiome composition significantly differs between PC patients and healthy controls [56, 57]. Some species with diagnostic properties for PC—for example, decreases in *Neisseria elongate*, *Neisseria mucosa*, and *Streptococcus mitis*, and increases in *Granulicatella adiacens* and

(See figure on next page.)

Fig. 6 Performances of oral metagenomic model in the classification of upper gastrointestinal (GI) and pancreaticobiliary cancer patients, obtained through LASSO multinomial logistic regression analysis with cross-validation. **A** Model predictive accuracies were evaluated by reducing multiclass [esophageal cancer (EC) vs. gastric cancer (GC) vs. biliary tract cancer (BC) vs. pancreatic cancer (PC)] predictions to multiple sets of binary predictions in the one vs. rest manner, depicted as receiver operating characteristic (ROC) curves. Values in parentheses represent 95% confidence intervals (CIs) for area under the curve (AUC) values. The optimal cutoff points for each model yielded the following sensitivity and specificity values: EC—87.0% and 88.1%; GC – 94.8% and 89.5%; BC—98.8% and 92.9%; PC—90.7% and 95.2%. For the macro- and micro-average models, the sensitivity and specificity were 89.3% and 89.2%, and 89.9% and 92.7%, respectively. **B** Lollipop plots show the coefficients of selected variables in the model. Red, blue, and gray colors represent positive, negative, and zero coefficients, respectively

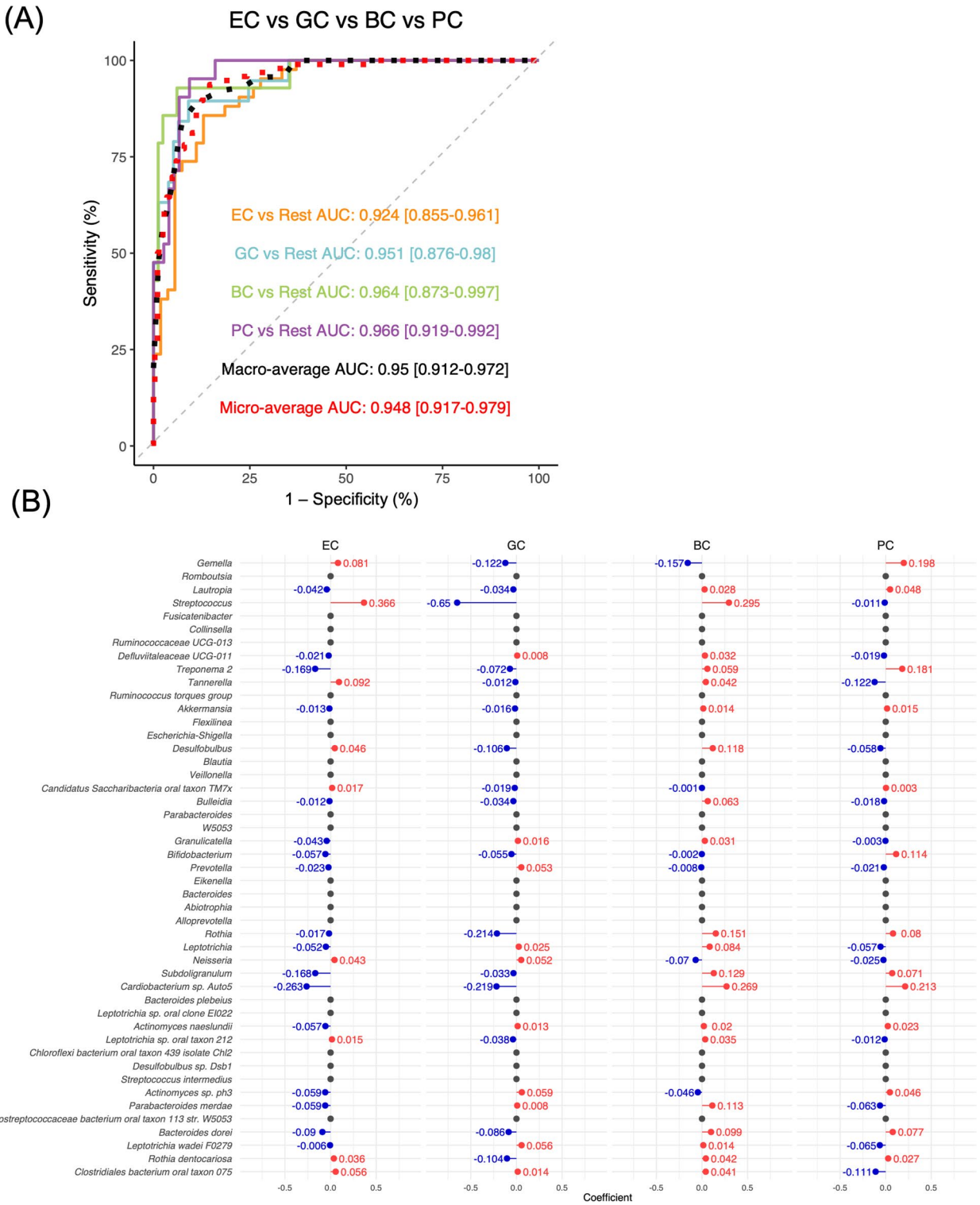

Fig. 6 (See legend on previous page.)

Fusobacterium periodonticum—have been reported, but there has been minimal consistency across studies [55, 56, 58]. Conversely, there are limited data available regarding the microbiome in BC; most studies have analyzed bile or biliary tract tissue based on the hypothesis that introducing gut bacteria into the biliary system alters

bile composition and behavior towards tumorigenesis [59–61]. Although we observed community-level dissimilarities in the oral microbiome between pancreaticobiliary cancer patients and the control groups, we could not detect any specific microbial taxa with diagnostic features.

Intriguingly, we observed an inverse correlation between the magnitude of oral dysbiosis and the distance from the oral cavity to the organ where the tumor occurred. Specifically, a cancer site closer to the mouth was associated with more pronounced cancer-related changes in the oral microbiome. This relationship was evident in our data, where PC showed the lowest contributions to oral microbial variance, as indicated by the smallest R^2 value (4.3%) compared with other cancer groups (Fig. 2). Although predictive models with high accuracy were obtained for patients with upper GI cancer, no differentially abundant taxa were detected for PC patients compared with healthy controls, and only one was identified for BC patients. Furthermore, taxonomic changes in cancer patients were correlated with the distance between the oral cavity and the primary tumor site. Oral taxa with significant decreases in upper GI cancer patients relative to healthy individuals showed a decreasing trend among pancreaticobiliary cancer patients, and vice versa (Fig. 4). The db-RDA plot in Fig. 5 revealed overlap between BC and PC patients, whereas other groups formed distinct clusters.

Researchers have proposed several plausible mechanisms by which oral dysbiosis may induce systemic diseases, particularly cancer. Oral infection or dysbiosis can influence the production of pathogen-associated molecular patterns, such as lipopolysaccharides, leading to initiation of systemic inflammatory and immune responses [62]. This chronic inflammation contributes to cancer progression by increasing the production of anti-apoptotic proteins, growth factors, and cytokines that support cancer growth and dissemination [63, 64]. Another mechanism involves the direct translocation of microbes between the oral cavity and distant organs [65, 66]. Because the oral cavity serves as the entry to the GI tract, oral bacteria can be swallowed and easily reach other organs, and vice versa. For example, resident oral species, such as *Fusobacterium nucleatum*, have been found in the gut microbiome of patients with IBD and colorectal cancer [67]. Considering the inter-organ networking behavior of the oral microbiome, the degrees and patterns of oral dysbiosis may be correlated with the distance between the oral cavity and the organ where tumor occurs.

This study had some limitations that must be addressed. First, because our study population was small and from a single institution, the generalizability of our findings might be limited. Further studies involving multi-center collections from larger populations with diverse geographic and genetic backgrounds are necessary to validate the robustness of the classification models. Furthermore, while we assessed the presence and severity of periodontitis, we acknowledge that additional periodontal factors—such as the use of antimicrobial mouthwash in the six months prior to the study, the presence of gingivitis, and other periodontal diseases—could influence the oral microbiota. Although we included self-reported data on participants' toothbrushing habits, more detailed oral hygiene behaviors and their potential impact on oral dysbiosis should be further investigated. Finally, as this is an observational study, causal relationships between variables cannot be firmly established. To better understand the dynamic interactions between oral health and cancers, future studies should employ longitudinal designs and incorporate additional in vitro and in vivo models, including animal studies, to explore the causal links between oral dysbiosis and upper GI and pancreaticobiliary cancers and further validate the diagnostic potential of the screened differential bacteria.

Conclusions

In conclusion, we identified distinctive oral microbiome features in upper GI and pancreaticobiliary cancer patients compared with healthy individuals, and we demonstrated that oral dysbiosis patterns were correlated with primary cancer sites. Although the mechanisms by which oral microbial taxa identified in this study contribute to pathogenesis are not yet fully understood, the oral metagenomic models for EC and GC suggest the feasibility of constructing diagnostic microbial markers for cancer screening.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-024-05989-9.

Additional file 1

Acknowledgements
Not applicable.

Author contributions
SO, JK, CS, HL, HSL, and KP conceived and planned the study. SO prepared samples, conducted next-generation sequencing, analyzed the data, performed statistical analyses, and wrote the paper. SO, JK, CS, HL, HSL, and KP interpreted the results. SO, HL, HSL, and KP contributed to the final version of the manuscript. All the authors approved the submitted version.

Funding
This work was supported by Grant no. 16-2021-0007 from the SNUBH Research Fund.

Availability of data and materials
Sequence data supporting the findings of this study have been deposited in the Sequence Read Archive (SRA) with the BioProject accession numbers PRJNA1109323 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1109323) and PRJNA1027093 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1027093).

Declarations

Ethics approval and consent to participate
This study was approved by the Institutional Review Board of SNUBH (IRB number: B-1810-499-301) and conducted in accordance with the Declaration of Helsinki. All participants provided written informed consent before participation.

Consent for publication
Written informed consent for publication of their clinical details was obtained from the participants.

Competing interests
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author details
[1]Department of Laboratory Medicine, Seoul National University College of Medicine, 103, Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea. [2]Department of Internal Medicine, Seoul National University Bundang Hospital, 82, Gumi-ro 173beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do 13620, Republic of Korea. [3]Department of Periodontology, Section of Dentistry, Seoul National University Bundang Hospital, 82, Gumi-ro 173beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do 13620, Republic of Korea. [4]Department of Pathology, Seoul National University College of Medicine, 103, Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea. [5]Department of Pathology, Seoul National University Hospital, Seoul National University College of Medicine, 103, Daehak-Ro, Jongno-Gu, Seoul 03080, Republic of Korea. [6]Cancer Research Institute, Seoul National University College of Medicine, 101, Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea. [7]Department of Laboratory Medicine, Seoul National University Bundang Hospital, 82, Gumi-ro 173Beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do 13620, Republic of Korea.

Received: 9 August 2024 Accepted: 13 December 2024
Published online: 06 January 2025

References

1. Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, et al. Cancer statistics for the year 2020: An overview. Int J Cancer. 2021;149(4):778–89.
2. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020;396(10258):1204–22.
3. Arnold M, Abnet CC, Neale RE, Vignat J, Giovannucci EL, McGlynn KA, et al. Global burden of 5 major types of gastrointestinal cancer. Gastroenterology. 2020;159(1):335–49.
4. Patel T. Increasing incidence and mortality of primary intrahepatic cholangiocarcinoma in the United States. Hepatology. 2001;33(6):1353–7.
5. Baria K, De Toni EN, Yu B, Jiang Z, Kabadi SM, Malvezzi M. Worldwide incidence and mortality of biliary tract cancer. Gastro Hep Adv. 2022;1(4):618–26.
6. Zoetendal EG, Rajilić-Stojanović M, de Vos WM. High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. Gut. 2008;57(11):1605–15.
7. Manson JM, Rauch M, Gilmore MS. The commensal microbiology of the gastrointestinal tract. GI microbiota and regulation of the immune system. 2008:15–28.
8. Sonnenburg JL, Angenent LT, Gordon JI. Getting a grip on things: how do communities of bacterial symbionts become established in our intestine? Nat Immunol. 2004;5(6):569–73.
9. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):207–14.
10. Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. Database (Oxford). 2010;2010:baq013.
11. Hyde ER, Andrade F, Vaksman Z, Parthasarathy K, Jiang H, Parthasarathy DK, et al. Metagenomic analysis of nitrate-reducing bacteria in the oral cavity: implications for nitric oxide homeostasis. PLoS ONE. 2014;9(3):e88645.
12. Rosier BT, Moya-Gonzalvez EM, Corell-Escuin P, Mira A. Isolation and characterization of nitrate-reducing bacteria as potential probiotics for oral and systemic health. Front Microbiol. 2020;11: 555465.
13. Ni J, Wu GD, Albenberg L, Tomov VT. Gut microbiota and IBD: causation or correlation? Nat Rev Gastroenterol Hepatol. 2017;14(10):573–84.
14. Meng C, Bai C, Brown TD, Hood LE, Tian Q. Human gut microbiota and gastrointestinal cancer. Genom Proteom Bioinform. 2018;16(1):33–49.
15. Yu LX, Schwabe RF. The gut microbiome and liver cancer: mechanisms and clinical translation. Nat Rev Gastroenterol Hepatol. 2017;14(9):527–39.
16. Persson GR, Persson RE. Cardiovascular disease and periodontitis: an update on the associations and risk. J Clin Periodontol. 2008;35(8 Suppl):362–79.
17. Dizdar O, Hayran M, Guven DC, Yılmaz TB, Taheri S, Akman AC, et al. Increased cancer risk in patients with periodontitis. Curr Med Res Opin. 2017;33(12):2195–200.
18. Irfan M, Delgado RZR, Frias-Lopez J. The oral microbiome and cancer. Front Immunol. 2020;11: 591088.
19. Fan X, Alekseyenko AV, Wu J, Peters BA, Jacobs EJ, Gapstur SM, et al. Human oral microbiome and prospective risk for pancreatic cancer: a population-based nested case-control study. Gut. 2018;67(1):120–7.
20. Uemura N, Okamoto S, Yamamoto S, Matsumura N, Yamaguchi S, Yamakido M, et al. Helicobacter pylori infection and the development of gastric cancer. N Engl J Med. 2001;345(11):784–9.
21. Polk DB, Peek RM Jr. Helicobacter pylori: gastric cancer and beyond. Nat Rev Cancer. 2010;10(6):403–14.
22. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. The lancet. 2007;370(9596):1453–7.
23. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581–3.
24. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852–7.
25. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2012;41(D1):D590–6.
26. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome. 2018;6:1–17.
27. Bisanz JE. qiime2R: Importing QIIME2 artifacts and associated data into R sessions. Version 099. 2018;13.
28. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE. 2013;8(4): e61217.
29. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci. 2003;14(6):927–30.
30. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1.
31. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:1–8.
32. Tsilimigras MC, Fodor AA. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. Ann Epidemiol. 2016;26(5):330–5.
33. Gerber GK. The dynamic microbiome. FEBS Lett. 2014;588(22):4131–9.
34. Arnold M, Rutherford MJ, Bardot A, Ferlay J, Andersson TM, Myklebust T, et al. Progress in cancer survival, mortality, and incidence in seven

high-income countries 1995–2014 (ICBP SURVMARK-2): a population-based study. Lancet Oncol. 2019;20(11):1493–505.
35. Peters BA, Wu J, Pei Z, Yang L, Purdue MP, Freedman ND, et al. Oral microbiome composition reflects prospective risk for esophageal cancers. Cancer Res. 2017;77(23):6777–87.
36. Chen X, Winckler B, Lu M, Cheng H, Yuan Z, Yang Y, et al. Oral microbiota and risk for esophageal squamous cell carcinoma in a high-risk area of China. PLoS ONE. 2015;10(12): e0143603.
37. Dao MC, Everard A, Aron-Wisnewsky J, Sokolovska N, Prifti E, Verger EO, et al. Akkermansia muciniphila and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. Gut. 2016;65(3):426–36.
38. Huck O, Mulhall H, Rubin G, Kizelnik Z, Iyer R, Perpich JD, et al. Akkermansia muciniphila reduces Porphyromonas gingivalis-induced inflammation and periodontal bone destruction. J Clin Periodontol. 2020;47(2):202–12.
39. Mulhall H, DiChiara JM, Deragon M, Iyer R, Huck O, Amar S. Akkermansia muciniphila and its pili-like protein Amuc_1100 modulate macrophage polarization in experimental periodontitis. Infect Immun. 2020;89(1).
40. Song B, Xian W, Sun Y, Gou L, Guo Q, Zhou X, et al. Akkermansia muciniphila inhibited the periodontitis caused by Fusobacterium nucleatum. NPJ Biofilms Microbiomes. 2023;9(1):49.
41. Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillère R, et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science. 2018;359(6371):91–7.
42. Holmstrøm K, Collins MD, Møller T, Falsen E, Lawson PA. Subdoligranulum variabile gen. nov., sp. Nov. from human feces. Anaerobe. 2004;10(3):197–203.
43. Van Immerseel F, Ducatelle R, De Vos M, Boon N, Van De Wiele T, Verbeke K, et al. Butyric acid-producing anaerobic bacteria as a novel probiotic treatment approach for inflammatory bowel disease. J Med Microbiol. 2010;59(Pt 2):141–3.
44. Everard A, Lazarevic V, Derrien M, Girard M, Muccioli GG, Neyrinck AM, et al. Responses of gut microbiota and glucose and lipid metabolism to prebiotics in genetic obese and diet-induced leptin-resistant mice. Diabetes. 2011;60(11):2775–86.
45. Conteduca V, Sansonno D, Lauletta G, Russi S, Ingravallo G, Dammacco FH. pylori infection and gastric cancer: state of the art (review). Int J Oncol. 2013;42(1):5–18.
46. Hooi JKY, Lai WY, Ng WK, Suen MMY, Underwood FE, Tanyingoh D, et al. Global prevalence of helicobacter pylori infection: systematic review and meta-analysis. Gastroenterology. 2017;153(2):420–9.
47. Sun JH, Li XL, Yin J, Li YH, Hou BX, Zhang Z. A screening method for gastric cancer by oral microbiome detection. Oncol Rep. 2018;39(5):2217–24.
48. Wu J, Xu S, Xiang C, Cao Q, Li Q, Huang J, et al. Tongue coating microbiota community and risk effect on gastric cancer. J Cancer. 2018;9(21):4039–48.
49. Huang K, Gao X, Wu L, Yan B, Wang Z, Zhang X, et al. Salivary microbiota for gastric cancer prediction: an exploratory study. Front Cell Infect Microbiol. 2021;11: 640309.
50. Yang Y, Long J, Wang C, Blot WJ, Pei Z, Shu X, et al. Prospective study of oral microbiome and gastric cancer risk among Asian, African American and European American populations. Int J Cancer. 2022;150(6):916–27.
51. Chen WP, Chang SH, Tang CY, Liou ML, Tsai SJ, Lin YL. Composition analysis and feature selection of the oral microbiota associated with periodontal disease. Biomed Res Int. 2018;2018:3130607.
52. Miyoshi T, Oge S, Nakata S, Ueno Y, Ukita H, Kousaka R, et al. Gemella haemolysans inhibits the growth of the periodontal pathogen Porphyromonas gingivalis. Sci Rep. 2021;11(1):11742.
53. Zagato E, Pozzi C, Bertocchi A, Schioppa T, Saccheri F, Guglietta S, et al. Endogenous murine microbiota member Faecalibaculum rodentium and its human homologue protect from intestinal tumour growth. Nat Microbiol. 2020;5(3):511–24.
54. Lepage C, Capocaccia R, Hackl M, Lemmens V, Molina E, Pierannunzio D, et al. Survival in patients with primary liver cancer, gallbladder and extrahepatic biliary tract cancer and pancreatic cancer in Europe 1999–2007: results of EUROCARE-5. Eur J Cancer. 2015;51(15):2169–78.
55. Farrell JJ, Zhang L, Zhou H, Chia D, Elashoff D, Akin D, et al. Variations of oral microbiota are associated with pancreatic diseases including pancreatic cancer. Gut. 2012;61(4):582–8.
56. Torres PJ, Fletcher EM, Gibbons SM, Bouvet M, Doran KS, Kelley ST. Characterization of the salivary microbiome in patients with pancreatic cancer. PeerJ. 2015;3: e1373.
57. Ertz-Archambault N, Keim P, Von Hoff D. Microbiome and pancreatic cancer: a comprehensive topic review of literature. World J Gastroenterol. 2017;23(10):1899–908.
58. Sun H, Zhao X, Zhou Y, Wang J, Ma R, Ren X, et al. Characterization of oral microbiome and exploration of potential biomarkers in patients with pancreatic cancer. Biomed Res Int. 2020;2020:4712498.
59. Zhou D, Wang JD, Weng MZ, Zhang Y, Wang XF, Gong W, et al. Infections of Helicobacter spp. in the biliary system are associated with biliary tract cancer: a meta-analysis. Eur J Gastroenterol Hepatol. 2013;25(4):447–54.
60. Saab M, Mestivier D, Sohrabi M, Rodriguez C, Khonsari MR, Faraji A, et al. Characterization of biliary microbiota dysbiosis in extrahepatic cholangiocarcinoma. PLoS ONE. 2021;16(3): e0247798.
61. Wheatley RC, Kilgour E, Jacobs T, Lamarca A, Hubner RA, Valle JW, et al. Potential influence of the microbiome environment in patients with biliary tract cancer and implications for therapy. Br J Cancer. 2022;126(5):693–705.
62. Thaiss CA, Zmora N, Levy M, Elinav E. The microbiome and innate immunity. Nature. 2016;535(7610):65–74.
63. Zhang Y, Niu Q, Fan W, Huang F, He H. Oral microbiota and gastrointestinal cancer. Onco Targets Ther. 2019;12:4721–8.
64. Asili P, Mirahmad M, Rezaei P, Mahdavi M, Larijani B, Tavangar SM. The association of oral microbiome dysbiosis with gastrointestinal cancers and its diagnostic efficacy. J Gastrointest Cancer. 2023;54(4):1082–101.
65. Schmidt TS, Hayward MR, Coelho LP, Li SS, Costea PI, Voigt AY, et al. Extensive transmission of microbes along the gastrointestinal tract. Elife. 2019;8.
66. Park SY, Hwang BO, Lim M, Ok SH, Lee SK, Chun KS, et al. Oral-gut microbiome axis in gastrointestinal disease and cancer. Cancers (Basel). 2021;13(9):2124.
67. Hashemi Goradel N, Heidarzadeh S, Jahangiri S, Farhood B, Mortezaee K, Khanlarkhani N, et al. Fusobacterium nucleatum and colorectal cancer: a mechanistic overview. J Cell Physiol. 2019;234(3):2337–44.

Publisher's Note