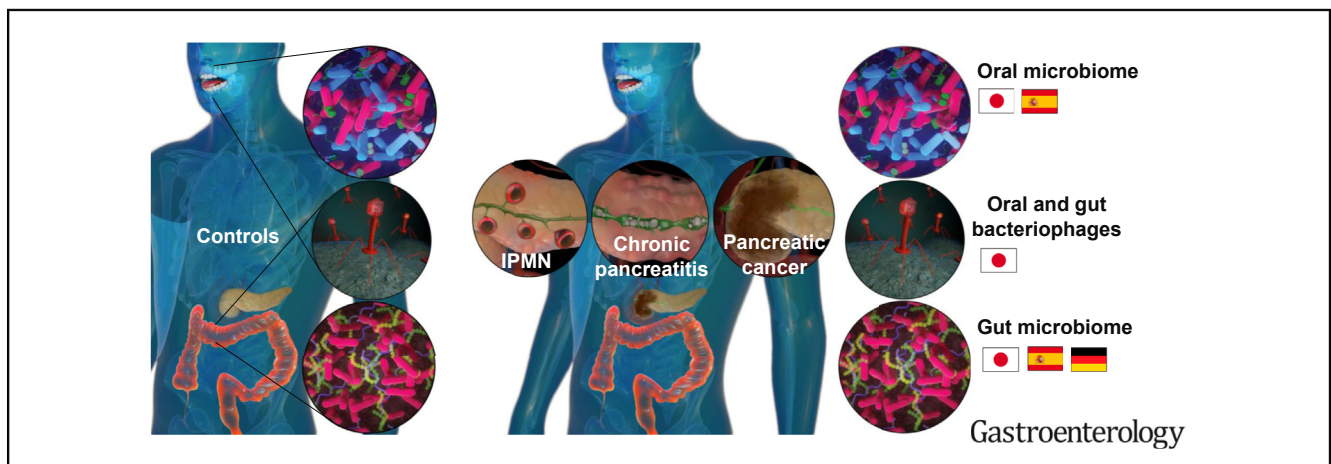# PANCREAS

Metagenomic Identification of Microbial Signatures Predicting Pancreatic Cancer From a Multinational Study

Naoyoshi Nagata,[1,2,*] **Suguru Nishijima,**[3,4,5,*] Yasushi Kojima,[2] Yuya Hisada,[2] Koh Imbe,[2] Tohru Miyoshi-Akiyama,[6] Wataru Suda,[7] Moto Kimura,[8] Ryo Aoki,[9] Katsunori Sekine,[10] Mitsuru Ohsugi,[11,12] Kuniko Miki,[1,2] Tsuyoshi Osawa,[13] Kohjiro Ueki,[14] Shinichi Oka,[15] Masashi Mizokami,[16] Ece Kartal,[5] Thomas S. B. Schmidt,[5] Esther Molina-Montes,[17] Lidia Estudillo,[17] Nuria Malats,[17] Jonel Trebicka,[18,19] Stephan Kersting,[20,21] Melanie Langheinrich,[20,21] Peer Bork,[5,22] Naomi Uemura,[1,10] Takao Itoi,[23] and Takashi Kawai[1]

[1]Department of Gastroenterological Endoscopy, Tokyo Medical University, Tokyo, Japan; [2]Department of Gastroenterology and Hepatology, National Center for Global Health and Medicine, Tokyo, Japan; [3]Computational Bio-Big Data Open Innovation Lab, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan; [4]Graduate School of Advanced Science and Engineering, Waseda University, Tokyo, Japan; [5]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany; [6]Pathogenic Microbe Laboratory, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan; [7]Laboratory for Microbiome Sciences, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; [8]Department of Clinical Research Strategic Planning Center for Clinical Sciences, National Center for Global Health and Medicine, Tokyo, Japan; [9]Institute of Health Sciences, Ezaki Glico Co., Ltd., Osaka, Japan; [10]Department of Gastroenterology and Hepatology, National Center for Global Health and Medicine, Kohnodai Hospital, Tokyo, Japan; [11]Department of Diabetes, Endocrinology, and Metabolism, Center Hospital, National Center for Global Health and Medicine, Tokyo, Japan; [12]Diabetes and Metabolism Information Center, Diabetes Research Center, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan; [13]Division of Nutriomics and Oncology, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan; [14]Diabetes Research Center, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan; [15]AIDS Clinical Center, National Center for Global Health and Medicine, Tokyo, Japan; [16]Genome Medical Sciences Project, Research Institute, National Center for Global Health and Medicine, Chiba, Japan; [17]Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Center (CNIO), Madrid, and CIBERONC, Spain; [18]Section for Translational Hepatology, Department of Internal Medicine I, Goehte University Frankfurt, Frankfurt, Germany; [19]European Foundation for the Study of Chronic Liver Failure, Barcelona, Spain; [20]Department of Surgery, University Hospital of Erlangen, Erlangen, Germany; [21]Department of Surgery, University Clinic Greifswald, Greifswald, Germany; [22]Department of Bioinformatics, Biocenter, University of Würzburg, Würzburg, Germany; and [23]Department of Gastroenterology and Hepatology, Tokyo Medical University, Tokyo, Japan

BACKGROUND & AIMS: To identify gut and oral metagenomic signatures that accurately predict pancreatic ductal carcinoma (PDAC) and to validate these signatures in independent cohorts. **METHODS:** We conducted a multinational study and performed shotgun metagenomic analysis of fecal and salivary samples collected from patients with treatment-naïve PDAC and non-PDAC controls in Japan, Spain, and Germany. Taxonomic and functional profiles of the microbiomes were characterized, and metagenomic classifiers to predict PDAC were constructed and validated in external datasets. **RESULTS:** Comparative metagenomics revealed dysbiosis of both the gut and oral microbiomes and identified 30 gut and 18 oral species significantly associated with PDAC in the Japanese cohort. These microbial signatures achieved high area under the curve values of 0.78 to 0.82. The prediction model trained on the Japanese gut microbiome also had high predictive ability in

Spanish and German cohorts, with respective area under the curve values of 0.74 and 0.83, validating its high confidence and versatility for PDAC prediction. Significant enrichments of *Streptococcus* and *Veillonella* spp and a depletion of *Faecalibacterium prausnitzii* were common gut signatures for PDAC in all the 3 cohorts. Prospective follow-up data revealed that patients with certain gut and oral microbial species were at higher risk of PDAC-related mortality. Finally, 58 bacteriophages that could infect microbial species consistently enriched in patients with PDAC across the 3 countries were identified. **CONCLUSIONS:** Metagenomics targeting the gut and oral microbiomes can provide a powerful source of biomarkers for identifying individuals with PDAC and their prognoses. The identification of shared gut microbial signatures for PDAC in Asian and European cohorts indicates the presence of robust and global gut microbial biomarkers.

Keywords: Pancreatic cancer; Microbiome; Shotgun metagenomics; Biomarker; Bacteriophage.

Pancreatic cancer remains one of the most lethal malignant neoplasms, with an overall 5-year survival rate of <5%.[1,2] Unfortunately, most patients present in the advanced stage of the disease; of the 15% to 20% of patients who undergo surgical resection, only 20% survive to 5 years.[3,4] Owing to population growth and aging, the number of cases and deaths from pancreatic cancer have doubled from 1990 to 2017.[1] Prior studies indicated smoking, alcohol, obesity, periodontal disease, and diabetes mellitus as risk factors for pancreatic cancer[4] and identified numerous potential markers for pancreatic cancer in blood and tumors. However, we are still unable to detect early pancreatic cancer.[5]

In the past decade, 16S ribosomal RNA (rRNA) gene-based microbiome studies have explored the association of pancreatic cancer with the human oral[6–9] and gut microbiomes.[10,11] Studies using mouse models have also revealed the involvement of the gut microbiome in pancreatic carcinogenesis.[12–16] These studies strongly suggest profound associations of the gut and oral microbiomes with pancreatic cancer. However, there still remains a need to identify high-resolution microbial and functional features by shotgun metagenomics and to evaluate the reproducibility of signatures across independent cohorts, which is crucial to establishing a robust and generalizable predictive model for pancreatic cancer.[17]

In this study, we analyzed and characterized the gut and oral microbiomes of Japanese (JP) patients with pancreatic ductal adenocarcinoma (PDAC) through a shotgun metagenomic sequencing approach. For case comparisons, we selected controls who were matched for possible confounding factors to identify true associations between PDAC and the microbiomes.[18] We then developed gut and oral metagenomic classifiers to predict PDAC, which were further validated in external datasets from Spanish (ES) and German (DE) cohorts. In addition, we explored microbial species that could serve as prognostic markers for patients with PDAC, and bacteriophages that potentially infect PDAC-associated species.

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

The human gut and oral microbiome are potentially associated with pancreatic ductal carcinoma. However, no previous studies have used shotgun metagenomics and validated results with external cohorts.

NEW FINDINGS

Metagenomic classifiers trained on the gut and oral microbiomes predicted pancreatic ductal carcinoma accurately, and were validated in independent cohorts. Several microbial species were associated with risk of pancreatic ductal carcinoma–related mortality. Bacteriophages infecting pancreatic ductal carcinoma–associated species were identified.

LIMITATIONS

In vitro and in vivo models would provide mechanistic insights behind the association between the microbiome and the development or progression of pancreatic ductal carcinoma.

IMPACT

Identification of shared gut microbial signatures across independent cohorts from different countries indicates the presence of a robust and global microbial signature for pancreatic ductal carcinoma.

PANCREAS

Material and Methods

Study Design and Sample Collection in the JP Cohort

We conducted a multinational study in Japan, Spain, and Germany. The ethics committees and the institutional review boards approved conducting this study in all participating institutions, and written informed consent was obtained from all participants (Supplementary Methods). In Japan, we recruited participants from the National Center for Global Health and Medicine (NCGM) Hospital and Tokyo Medical University Hospital, and the samples were analyzed at NCGM, Waseda University, the RIKEN Center for Integrative Medical Sciences, the National Institute of Advanced Industrial Science and Technology, and the European Molecular Biology Laboratory. Fecal and salivary samples were collected from 47 JP patients with treatment-naïve PDAC and 235 controls (1:5 case/control ratio, Table 1). The

Most current article

Table 1. Characteristics of Patients With Pancreatic Cancer and Controls in the 3 Cohorts

Japanese cohort	Pancreatic cancer	Matched controls	P
Total samples (feces, saliva)	47 (43, 47)	235 (235, 235)	NA
Age >70 (y)	27 (57.5)	139 (59.2)	.829
Sex, male	26 (55.3)	130 (55.3)	1.000
BMI > 25 (kg/m^2)	6 (12.8)	51 (21.7)	.231
Smoking index >400	12 (25.5)	72 (30.6)	.485
Alcohol	22 (46.8)	136 (57.9)	.163
Rice	5.28±0.95	5.45±0.96	.451
Bread	4.09±1.35	4.20±1.30	.605
Noodles	3.28±1.12	3.29±1.12	.977
Vegetables	4.70±1.49	4.80±1.38	.635
Fruit	4.02±1.50	3.99±1.48	.886
Seafood	3.77±1.16	3.79±0.88	.865
Meat	3.98±1.01	3.81±0.82	.473
Eggs	3.64±1.22	3.77±1.13	.641
Milk	3.79±1.67	3.39±1.82	.233
Coffee	4.13±1.88	4.36±1.88	.274
Periodontal disease	4 (8.5)	33 (14.0)	.476
Wearing dental crowns	30 (63.8)	177 (75.3)	.104
Implant or tooth bridge	17 (36.2)	101 (43.0)	.388
Denture wearing	17 (36.2)	85 (36.2)	1.000
Gingival bleeding	7 (14.9)	41 (17.5)	.832
Tooth root exposure	10 (21.3)	73 (31.1)	.179
Tooth wobbling	19 (40.4)	108 (46.0)	.487
Diabetes mellitus	14 (29.8)	47 (20.0)	.137
Dyslipidemia	14 (29.8)	86 (36.6)	.373
Hypertension	24 (51.1)	124 (52.8)	.831
Inflammatory bowel disease	0	0	NA
Irritable bowel syndrome	6 (12.8)	27 (11.5)	.804
Bristol Stool Scale, type 1 or 2	9 (19.2)	48 (20.4)	.842
Bristol Stool Scale, type 3, 4 or 5	42 (89.4)	221 (94.0)	.243
Bristol Stool Scale, type 6 or 7	10 (21.3)	36 (15.3)	.313
Antidiarrheal agent	0	0	NA
Laxatives	12 (25.5)	41 (17.5)	.195
Probiotics	9 (19.2)	44 (18.7)	1.000
PPIs	18 (38.3)	72 (30.6)	.304
NSAIDs	6 (12.8)	14 (6.0)	.117
Low-dose aspirin	5 (10.6)	23 (9.8)	.793

Table 1. Continued

Japanese cohort	Pancreatic cancer	Matched controls	P
Cancer stage, early/late	20 (42.6)/ 27 (57.5)	NA	NA
Death during follow-up	32 (68.1)	NA	NA
Median months of follow-up (IQR)	15.3 (6.2-27.8)	NA	NA

Spanish cohort (Kartal et al)[22]	Pancreatic cancer	Controls	P
Total samples (feces, saliva)	57 (57, 43)	50 (50, 45)	
Age mean	71.5	71.3	NA
Sex, male	36	31	.902
Obesity	16	13	.81
Smoking	28	26	.766
Alcohol	40	39	.358
Diabetes	17	5	.016
Cholesterol	25	13	.054
Periodontal disease	17	11	.358
Receding gums	23	13	.117
Antibiotics	32	27	.824
Probiotics	2	3	.663
PPIs	17	10	.243
Jaundice	32	2	<.001
Cancer stage, early/late	26/32	NA	NA

German cohort (Kartal et al)	Pancreatic cancer	Controls	P
Total samples (feces)	44	32	
Age mean	68.7	48	NA
Sex, male	19	15M	.749
Smoking	13	2	.018
Alcohol	18	5	.023
Diabetes	12	2	.033
Antibiotics	12	1	.006
PPIs	22	3	<.001
Cancer stage early/late	17/26	NA	NA

NOTE. Numbers in parentheses show percentages. Values presented with a plus/minus sign are means ± SD. Categorical data were compared using the χ^2 test or Fisher's exact test as appropriate. Continuous data were compared using the Mann–Whitney U test. For dietary habits, patients were specifically asked about typical eating patterns in the previous month, rated on a 7-point Likert scale (1, never or rarely; 2, 1–3 times/month; 3, 1–3 times/week; 4, 4–6 times/week; 5, once/day; 6, twice/day; and 7, 3 or more times/day).
BMI, body mass index; NA, not applicable; NSAIDs, nonsteroidal anti-inflammatory drugs; IQR, interquartile range.

controls were selected from noncancer JP subjects by matching for possible confounders and risk factors[4] such as age, sex, body mass index, lifestyle-related factors, dietary habits, dental/oral problems, comorbidities, bowel diseases, Bristol stool scale, and medications (Supplementary Methods). Because most of the subjects were elderly, we collected stimulated saliva samples at the hospital in the morning after overnight fasting.[19] For fecal samples, the participants defecated at home and placed the feces in a dedicated tube containing Cary–Blair medium, which is

PANCREAS
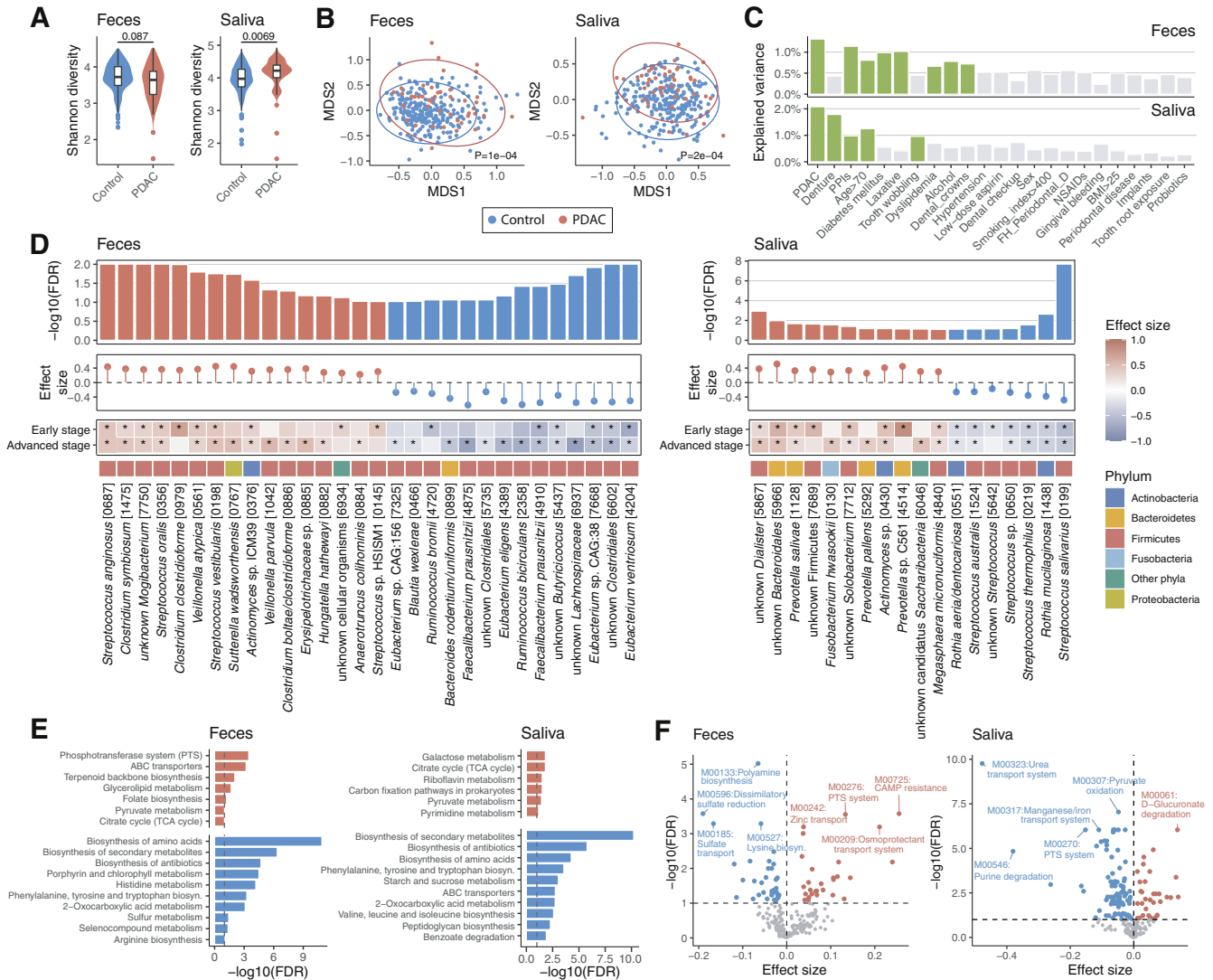
Figure 1. Comparison of the fecal and saliva microbiomes between controls and PDACs. (*A*) Violin plots showing the Shannon index of the gut and oral microbiomes at the species level in the controls and patients with PDAC. Numbers in the plots show *P* values obtained from the Wilcoxon rank-sum test. (*B*) MDS plots showing the similarity of the samples. *Red* and *blue circles* represent PDAC and controls, respectively. *P* values were obtained from permutational analysis of variance. (*C*) Bar plots showing the association of each metadata to the microbiome. *Green color* indicates a significant association (FDR < 0.05, permutational analysis of variance). (*D*) Bar plots showing the species significantly associated with PDAC (FDR < 0.1, generalized regression model, Supplementary Methods). *Red* and *blue colors* represent enriched and depleted species in PDAC, respectively. Line plots and heat maps show effect sizes (*β*-coefficients) of PDAC for the species obtained from the generalized regression models. Numbers in parentheses beside the species names indicate mOTU ID. (*E*) Associations between KEGG pathways and PDAC assessed by pathway-enrichment analysis. *Red* and *blue colors* represent significant enrichment and depletion of KOs in patients with PDAC, respectively (FDR < 0.1, Fisher's exact test, Supplementary Methods). Top 10 pathways with FDR < 0.1 are shown in the figure. (*F*) Volcano plots showing the associations between MOs and PDAC. *Red* and *blue colors* represent a significant enrichment and depletion of the MOs in patients with PDAC, respectively (FDR < 0.1, generalized regression model). MDS, multidimensional scaling.

relatively stable at 4°C and 25°C.[20] As bowel-cleansing agents for colonoscopy can have a profound effect on the gut microbiome,[21] we avoided collecting samples within 1 month after bowel cleansing. Both salivary and fecal samples were stored at −80°C until DNA extraction.

Sample Collection in the ES and DE Cohorts

In the ES cohort, participants were recruited from the Hospital Ramón y Cajal in Madrid and Hospital Vall d'Hebron in

Barcelona, Spain.[22] Fecal samples were collected from 57 cancer-treatment–naïve PDAC cases and 50 controls, whereas saliva samples were collected from 43 PDAC cases and 45 controls. In the DE cohort, fecal samples from 44 PDAC cases and 32 controls were collected from the Department of Surgery, University Hospital of Erlangen, and the Section for Translational Hepatology, Department of Internal Medicine I, Goethe University Clinic, Frankfurt.[22] Stool and saliva (mouthwash) samples were preserved in RNALater and stored at 4°C immediately for 12 hours, then transferred to −20°C for

another 24 hours, then stored at –80°C until DNA extraction. Saliva samples were collected in the morning of the recruitment day, before breakfast was taken and dental hygiene was done.

Shotgun Metagenomic Sequencing and Data Analysis in the JP Cohort

Bacterial DNA was isolated according to methods described in previous studies.[23,24] Sequencing libraries were generated using a NEBNext Ultra II DNA Library Prep Kit for Illumina, and 151-base pair paired-end reads were sequenced on a HiSeq X. Quality control of the metagenomic data was performed, and contamination with human and φX reads was excluded by mapping the metagenomic reads to the human (hg38) and φX genomes using bowtie2.[25] MEGAHIT[26] was used to assemble the quality-controlled reads, and Prodigal[27] was used to predict genes from the contigs (>1 kb). The genes were clustered using Linclust[28] with a ≥95% similarity threshold, and a nonredundant gene set was constructed. Functional annotation for the genes was performed using eggNOG-mapper.[29] Taxonomic profiles of the metagenomic samples were obtained with mOTUs2,[30] and functional profiles were obtained by mapping the metagenomic reads to the nonredundant genes using bowtie2.

Identification of Bacteriophages in the Metagenomic Data

Bacteriophage genomes were explored in the gut and oral metagenomic datasets of the JP cohort. Host prediction of the detected phages was performed by matching CRISPR spacers identified in the reference microbial genomes in the RefSeq.

Statistical Analysis and Model Construction

Differentially abundant microbial features between patients with PDAC and controls were detected by generalized linear regression analysis. The regression model was constructed on log10-transformed abundances of each feature while adjusting for age, sex, and body mass index. Multiple testing was adjusted for by the Benjamini-Hochberg method. A random forest classifier was used to predict PDAC based on the taxonomic and functional profiles of the microbiomes. The classifier was trained by 5-times repeated 10-fold cross-validation using the train function in the caret package. Relative abundances of each feature were log10-transformed and standardized by z-score transformation before the training.

We also conducted a prospective follow-up study of patients with PDAC to estimate the risk of mortality in the JP cohort. Survival analysis was performed using the Cox regression model with LASSO feature selection. The model was constructed with the glmnet function, and the top 10 most important microbial species predicting prognosis were selected.

Additional detailed methods are provided in the online Supplementary Methods.

Results

Species and Functional Characterization of the PDAC Microbiome

Fecal and salivary samples were collected from 47 JP patients with treatment-naïve PDAC and 235 controls (Table 1). The baseline characteristics were well balanced between PDAC cases and controls. PDAC included early (I [n = 4] and II [n = 13]) and advanced (III [n = 7] and IV [n = 20]) stages. Shotgun metagenomic analysis of these fecal and salivary samples (Supplementary Tables 1 and 2) identified 1151 and 517 species (Supplementary Figure 1), respectively, based on the marker-gene-based profiling method using mOTUs2,[30] and 6846 and 6457 Kyoto encyclopedia of genes and genomes (KEGG) orthologies (KOs) from the assembled contigs, respectively.

We found that the Shannon index was decreased (P = .087) in the gut microbiome, whereas it was increased in the oral microbiome (P = .0069) in patients with PDAC compared with controls (Figure 1A). Similar results were also observed for the inverse Simpson index and Pielou's evenness of the gut and oral microbiomes (Supplementary Figure 2A). Multidimensional scaling analysis revealed differences in β-diversity for both the gut and oral microbiomes between the patients with PDAC and controls (P = .0001, Figure 1B). Although various metadata, such as age, alcohol consumption, medication, and other diseases, were significantly associated with the microbiomes (false discovery rate [FDR] < 0.05), PDAC had the largest explanatory power for the variation in both the gut and oral microbiomes (Figure 1C). The deviation of the microbiomes of patients with PDAC from controls was slightly more pronounced for individuals in advanced stages than those in early stages (Supplementary Figure 3).

A comparison of species abundance between patients with PDAC and controls identified 30 and 18 species with significant changes in the gut and oral microbiomes, respectively (FDR < 0.1, Figure 1D, Supplementary Tables 3 and 4, Supplementary Figure 2B). Species increased in the gut microbiomes of patients with PDAC included some oral species, such as Streptococcus spp (Streptococcus oralis, Streptococcus vestibularis, and Streptococcus anginosus), Veillonella spp (Veillonella atypica and Veillonella parvula), and Actinomyces spp, whereas species depleted included several from the order Clostridiales, such as unknown Lachnospiraceae, Eubacterium ventriosum, and Faecalibacterium prausnitzii. Of these species with significant changes, 16 already showed differences at the early stage of the disease as compared with controls (FDR < 0.1, Figure 1D, Supplementary Table 3). We also found that the total abundance of typical oral species in the gut microbiome was significantly greater in patients with PDAC than in controls (P = .00018, Supplementary Figure 2C).

In the oral microbiomes of patients with PDAC, several unknown species in the phylum Firmicutes (unknown Firmicutes, Dialister, and Solobacterium spp) and Prevotella spp (Prevotella pallens and Prevotella sp C561) were significantly enriched, whereas some Streptococcus spp, such as Streptococcus salivarius, Streptococcus thermophilus, and Streptococcus australis were depleted (FDR < 0.1, Figure 1D, Supplementary Table 4). Of these significant species, 15 were also significantly different between patients with early-stage PDAC and controls (FDR < 0.1, Supplementary Table 4).

Functional analysis identified 996 and 1787 KOs with significant changes in the gut and oral microbiomes of

patients with PDAC compared with controls, respectively (FDR < 0.1, Supplementary Tables 5 and 6). Pathway-enrichment analysis showed that a large number of increased KOs were involved in phospho-transferase systems, ABC transporters, and terpenoid backbone biosynthesis, whereas KOs involved in amino acid and secondary metabolite biosynthesis and in porphyrin and chlorophyll metabolism were largely depleted in the gut (Figure 1E). At the KEGG module (MO) level, cationic antimicrobial peptide resistance, C5 isoprenoid biosynthesis, the mevalonate pathway, and the osmoprotectant transport system were significantly enriched in the gut microbiomes of patients
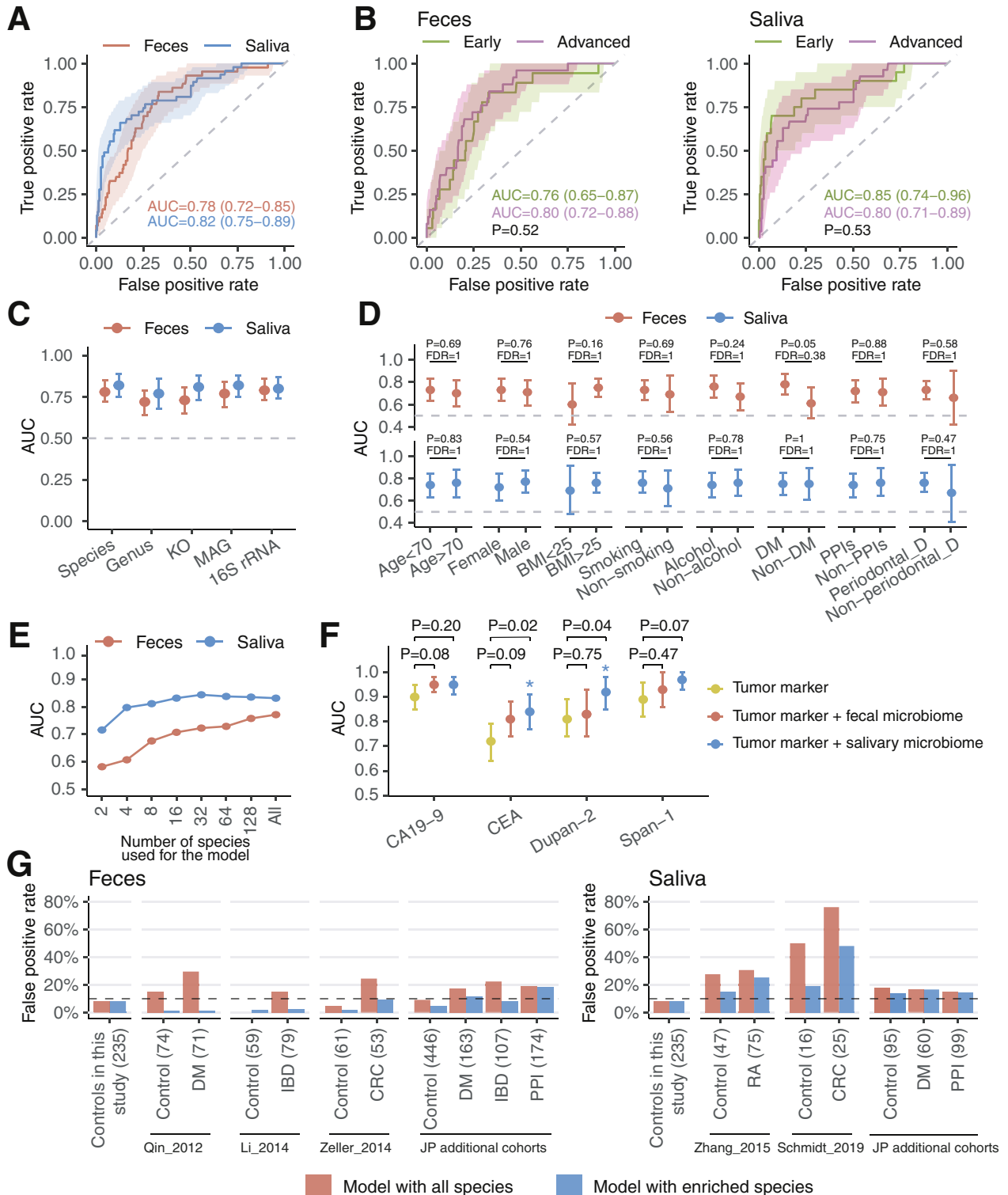
with PDAC, whereas polyamine biosynthesis, dissimilatory sulfate reduction, and the sulfate transport system were significantly depleted (Figure 1F, Supplementary Table 7).

The oral microbiomes of patients with PDAC were characterized by a significant enrichment of KOs involved in carbohydrate metabolism, such as galactose metabolism and the citrate cycle, and a depletion of those involved in metabolism of secondary metabolites and amino acids (Figure 1E, Supplementary Table 6). At the MO level, D-glucuronate degradation, thiamine biosynthesis, and the phosphonate transport system were significantly enriched in the oral microbiomes of patients with PDAC, whereas the urea transport system, the manganese/iron transport system, phosphotransferase system for trehalose, and pyruvate oxidation were significantly depleted (Figure 1F, Supplementary Table 8).

To further explore additional PDAC-associated microbial species, we constructed a total of 7546 metagenome-assembled genomes (MAGs) and dereplicated them into 720 species-level MAGs (average nucleotide identity < 95%) (Supplementary Figure 4, Supplementary Table 9, Supplementary Methods). Comparative analysis revealed 246 and 93 MAGs differentially abundant in the gut and oral samples between patients with PDAC and controls, respectively (FDR < 0.1, Supplementary Tables 10 and 11). Of the significant MAGs, 55 (22.3%) in the gut and 17 (18.3%) in the oral microbiomes had no species-level assignment in the Genome Taxonomy Database and were considered to be novel MAGs of uncultured species. Most of these MAGs in the gut were classified into *Streptococcus* (n = 20), *Granulicatella* (n = 9), or *Pauljensenia* (n = 6) at the genus level in the Genome Taxonomy Database, and those in the oral microbiome were classified into *Streptococcus* (n = 7) and *Bulleidia* (n = 5). These data suggest that as-yet uncultured and uncharacterized species in the gut and oral microbiomes are associated with PDAC.

Metagenomic Classifiers for PDAC in the JP Cohort

To construct prediction models for PDAC, we used a random forest classifier and trained the model on the species-level taxonomic profiles of the microbiomes (Supplementary Methods). The area under the curve (AUC) values were 0.78 and 0.82 for the gut and oral microbiomes, respectively (Figure 2A). The species that greatly contributed to the prediction (Supplementary Table 12) were consistent with those having significant changes in abundance between patients with PDAC and controls, such as unknown Lachnospiraceae, unknown *Mogibacterium* and *S vestibularis* for the fecal model, and *Prevotella salivae*, *S salivarius,* and unknown Bacteroidales for the saliva model (Figure 1D). These models accurately predicted patients with PDAC in both the early and advanced stages (Figure 2B). Also, models based on the genus, KO, and MAG profiles of the gut and oral microbiomes showed high predictive accuracy for the classification (Figure 2C). The accuracy of the gut species model was slightly higher for diabetic patients, with an AUC of 0.78, than for nondiabetic individuals, with an AUC of 0.60 (P = .05, FDR = 0.38), whereas other metadata, such as age, sex, smoking, alcohol, proton pump inhibitors (PPIs), and periodontal disease, did not substantially affect the performance of the gut and oral species models (Figure 2D). To further validate the accurate prediction of PDAC with the microbiome data, we performed 16S rRNA gene analysis of the same fecal and salivary samples, and found that the models yielded high AUCs, of 0.79 and 0.80, for the gut and oral microbiomes, respectively (Figure 2C, Supplementary Figure 5).

To evaluate the relationship between the number of species used in the model and prediction accuracy, we performed recursive feature elimination analysis (Supplementary Methods). The AUC decreased from 0.78 to 0.57 and from 0.82 to 0.71 in the gut and oral microbiomes, respectively, as the number of species used in the model decreased (Figure 2E). Nevertheless, AUCs of 0.71 and 0.83 were achieved for the gut and oral samples, respectively, with 16 species in the model (Figure 2E), indicating candidate microbial markers for screening PDAC.

As serological tumor markers have been used for PDAC screening in clinical practice,[31,32] we evaluated the predictive accuracy of known tumor markers and the combined use of the markers and microbiomes. The AUCs of tumor

Figure 2. Prediction of PDAC using the gut and oral microbiomes. (*A*) Prediction of PDAC using random forest classifiers trained on the species-level taxonomic profiles of the gut and oral microbiomes. Numbers in parentheses represent the 95% confidence interval of the AUC. (*B*) Comparison of the prediction accuracies of the models between early- and advanced-stage patients. (*C*) Prediction accuracies of the models trained on various profiles of the gut and oral microbiomes. Error bars show the 95% confidence interval of the AUC. (*D*) Comparison of the prediction accuracy of the models across various metadata. The models were trained on the species-level taxonomic profiles of the gut (*top*) and oral (*bottom*) microbiomes. (*E*) Association between the number of species used for the prediction models and AUCs. The prediction models were initially trained using all species (n = 218 and 266 for the gut and oral models, respectively); the number of species was reduced by recursive feature elimination; and model performance was evaluated at each number (n = 2, 4, 8, 16, 32, 64, 128, and all species). (*F*) Comparisons of prediction accuracy among tumor markers, microbiomes, and their combinations. Numbers above the plot show *P* values for the increase in the AUC for the combination use of the microbiomes and the tumor marker as compared with that of only tumor markers. *Red* and *blue colors* represent *P* values for the gut and oral microbiomes, respectively. *Asterisks* represent a significant increase in the AUC (*P* < .05). The number of patients with PDAC and controls for whom tumor markers were collected was 47 and 127 for CA 19-9, 47 and 155 for CEA, 40 and 29 for Dupan-2, and 39 and 29 for Span-1, respectively. (*G*) False positive rates by the prediction models in various datasets from previous studies. *Red* and *blue bars* represent the percentage of cases predicted as PDAC by the prediction model trained on all the species and that trained on enriched species, respectively. The *dotted line* represents the false positive rate of 10%. Numbers in parentheses show the number of samples in each dataset.

markers such as CA19-9, CEA, Dupan-2, and Span-1 ranged from 0.70 to 0.90 (Figure 2F), similar to those of the fecal and salivary microbiomes. The combination of tumor markers and microbiomes improved the AUCs to a range of 0.81 to 0.97 (Figure 2F).

Because dysbiotic signatures of the gut microbiomes are partially overlapped among different diseases (eg, depletion of beneficial microbes),[33] we investigated specificity of the models by applying them to publicly available disease microbiome datasets (as true negatives) in 5 previous studies (Supplementary Methods, Supplementary Table 13).[34–38] Also, we collected and sequenced additional fecal and salivary samples from individuals with diabetes mellitus (n = 163, n = 61, respectively) and inflammatory bowel disease (n = 107, n = 13, respectively), in those treated with PPIs (n = 174 and n = 105, respectively), and healthy controls (n = 394, n = 96, respectively) and used them for the analysis. When we used cutoff values that had a 10% false positive rate in our own dataset, the rates in the other disease datasets were 23.1% and 53.8%, on average, for the gut and oral models with all species, respectively (Figure 2G), suggesting a high proportion of the patients were mistakenly predicted as PDAC. To improve the specificity, we next focused on only species enriched in patients with PDAC, as enriched species could be more specific to each disease than depleted ones.[33] The AUCs of the new gut and oral models with only enriched species were found to be similar to those trained on all the species (0.78 and 0.84, respectively). The false positive rate of the new gut model was 4.5% in the other datasets, whereas that of the new oral model was still 35.7% on average (Figure 2G), suggesting that the gut model could be specific to PDAC, although the oral model had relatively low specificity to PDAC. In addition, we found that the dataset of PPI users still had a somewhat high false positive rate of 20% in the new gut model. This may be because of the partial overlap of the gut microbial signature between PDAC and PPI treatment, such as increases in *Streptococcus* spp and *Veillonella* spp (Supplementary Figure 6).

Analysis of the Gut Microbiomes of Patients With Intraductal Papillary Mucinous Neoplasm and Chronic Pancreatitis

Intraductal papillary mucinous neoplasm (IPMN) and chronic pancreatitis (CP), a premalignant condition and a strong risk factor of PDAC, respectively, are sometimes difficult to distinguish from PDAC with conventional biomarkers.[5] To characterize the microbiomes of these patients and compare the microbial signatures with that of PDAC, we collected additional fecal samples from patients with branch duct IPMN (n = 150) and CP (n = 65) and performed shotgun metagenomic analysis (Supplementary Methods). Multidimensional scaling analysis revealed that the gut microbial communities of patients with IPMN and CP had more similarities to those of PDAC than to the controls (Figure 3A). Comparison of relative abundances of each microbial species revealed that some PDAC-associated species, such as *S anginosus*, *Clostridium clostridioforme*, and *V atypica*, were also significantly enriched in IPMN or CP as compared with controls (Figure 3B). In addition, the effect sizes of the diseases on each microbial species showed significant positive correlations between PDAC and IPMN (rho = 0.54, P < .01) and between PDAC and CP (rho = 0.63, P < .01, Figure 3C). By contrast, the other PDAC-associated species, such as *Clostridium symbiosum*, *S oralis*, and *unknown Mogibacterium*, were still significantly enriched in PDAC as compared with IPMN and CP as disease controls. Next, we examined whether CP/IPMN and PDAC can be discerned using prediction models based on the gut microbiome. The AUCs obtained from random forest models were 0.70 and 0.68 for IPMN and CP, respectively, slightly lower than the values between PDAC and controls (Figure 3D). Almost the same values were obtained when we analyzed patients with early- and advanced-stage PDAC separately (0.67–0.72). Moreover, when we combined the gut microbiome and CA19-9, the AUC for IPMN slightly increased compared with that based on only CA19-9 (P = .06, Supplementary Figure 7). Collectively, these results suggested that the microbiomes of CP/IPMN are more similar to that of PDAC than to controls, but combining the microbiomes with tumor markers can potentially discriminate between PDAC and IPMN.

Validation of the Metagenomic Classifiers in the ES and DE Cohorts

To evaluate the robustness of the metagenomic classifiers trained on the JP dataset, we used external datasets from ES and DE cohorts (Table 1). These metagenomic datasets were analyzed with the same analytical pipeline as the JP dataset, and taxonomic and functional profiles were obtained (Supplementary Methods). Comparative analysis of whole community structures showed that both the gut and oral microbiomes differed significantly among the cohorts (Supplementary Figure 8), probably due to different lifestyles, geographic factors, or methodologies for sample processing.[39,40] Nevertheless, we found a high predictive accuracy when applying the JP-trained gut species model to the ES and DE datasets (AUC = 0.74 and 0.83, respectively, Figure 4A). Consistent with this result, the JP models based on the genus, KO, and MAG profiles also showed high AUCs, ranging from 0.64 to 0.79, in the ES and DE datasets (Figure 4A). Similarly, a high value was obtained when we applied the model trained on the ES and DE gut species to the JP datasets (AUC = 0.77 and 0.73, respectively, Figure 4A). When we focused on early- and advanced-stage patients separately, high AUC values were also obtained (Supplementary Figure 9A). In contrast, the oral models trained on the JP oral microbial profiles showed low AUCs, of 0.52–0.53, in the ES dataset (Figure 4A). Similarly, the ES oral models showed AUCs of 0.48–0.64 in the JP dataset (Figure 4A). The ES oral models showed AUCs of 0.54–0.60, even with internal cross-validation and no significant differences in α- and β-diversity between patients with PDAC and controls (Supplementary Figure 8C).

To further explore the consistent microbial signatures for PDAC across different cohorts, we compared PDAC-species associations among the 3 cohorts. The effect sizes of PDAC for each gut species showed significant positive
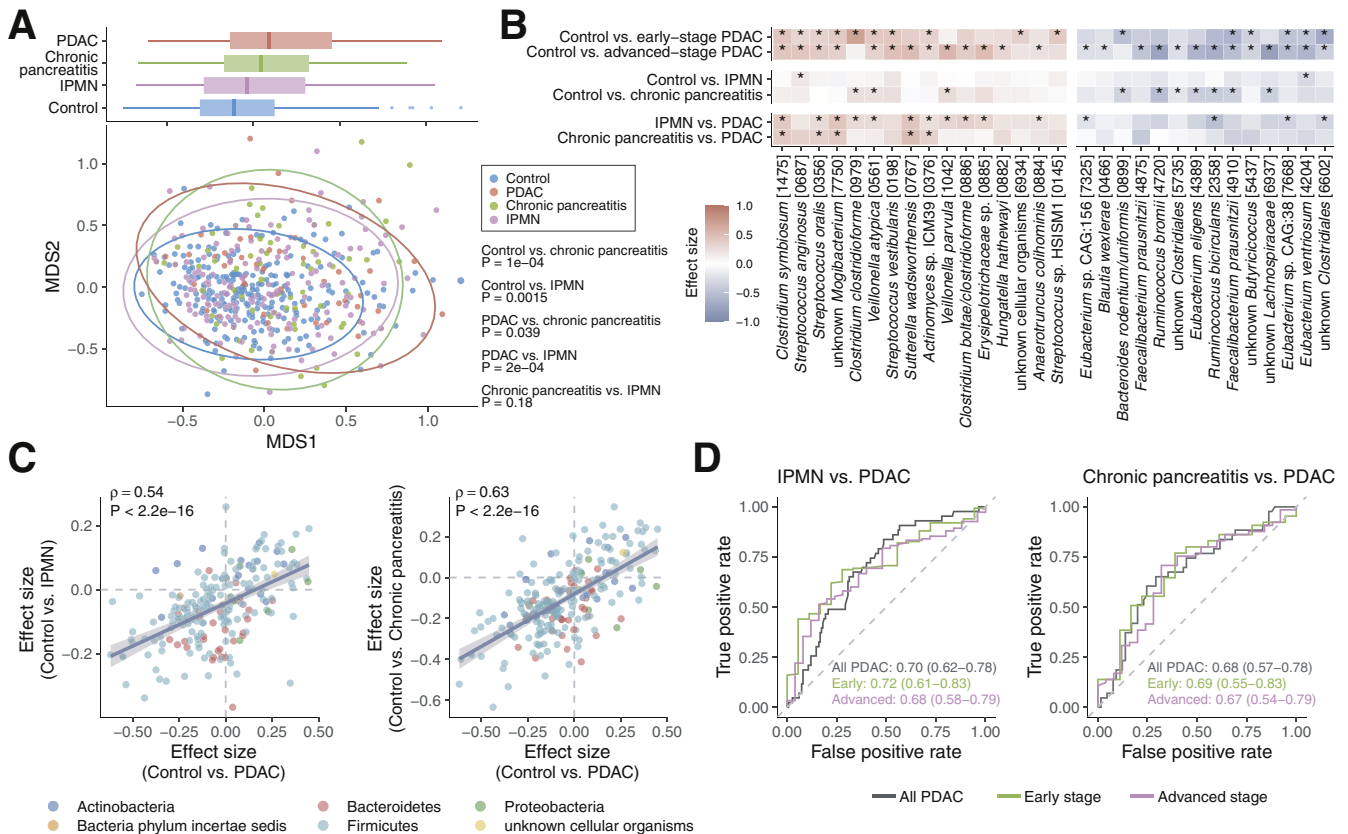
Figure 3. Comparison of fecal samples among IPMN, CP, controls, and PDAC. (*A*) MDS plots showing the similarity of the gut microbiomes among the controls (n = 235), IPMN (n = 150), CP (n = 65), and PDAC (n = 43). *Blue, purple, green*, and *red circles* represent controls, IPMN, CP, and PDAC, respectively. *P* values between the groups were obtained from permutational analysis of variance. (*B*) Heatmap summarizing results of comparisons between the groups. The results for 30 gut species that were differentially abundant between PDAC and controls are shown (see Figure 1*D*). *Asterisks* represent a significant association (FDR < 0.1). (*C*) Comparison of effect sizes of each disease on microbial species among PDAC, IPMN, and CP. Effect sizes were obtained from linear regression models in which the microbial abundances were compared between the control and disease patients. The color of the plot shows phylum-level taxonomy of the microbial species. The *blue line* represents the regression line, and the *gray shadow* shows the 95% confidence interval of the regression. (*D*) Prediction accuracies of the random forest classifiers trained on species-level taxonomic profiles of the gut microbiomes. Numbers in parentheses show the 95% confidence interval of the AUC. Numbers in parentheses beside the species names indicate mOTU ID. MDS, multidimensional scaling.

correlations between the JP and ES datasets and between the JP and DE datasets (*P* < .05), whereas there was no correlation for the salivary species between the JP and ES datasets (*P* > .05, Supplementary Figure 10*A*). We identified 32 gut species that were consistently enriched or depleted in patients with PDAC in at least 2 of the 3 cohorts (FDR < 0.1, Figure 4*B*, Supplementary Figure 9*B*, Supplementary Table 14). Among the 32 species, *V atypica, V parvula, S anginosus*, and *S oralis* were consistently enriched, whereas *F prausnitzii* was consistently decreased in the gut microbiome of patients with PDAC in all the 3 cohorts. In contrast, the JP and ES oral microbiome datasets showed no shared species with significant changes in abundance between patients with PDAC and controls (Supplementary Table 15).

At the functional level, we identified 670 KOs that were consistently associated with the gut microbiomes of patients with PDAC in at least 2 of the 3 cohorts (FDR < 0.1, Supplementary Table 16). Pathway-enrichment analysis of these KOs revealed that terpenoid backbone biosynthesis, ABC transporters, and pyruvate metabolism were enriched,

whereas biosynthesis of amino acids, biosynthesis of secondary metabolites, and histidine metabolism were depleted in patients with PDAC among the cohorts (FDR < 0.1, Supplementary Figure 10*B*). At the MO level, 42 MOs in the gut microbiome were shared between at least 2 of the 3 cohorts with significant changes in abundance (FDR < 0.1), which included consistent enrichments of C5 isoprenoid biosynthesis (the mevalonate pathway) and the osmoprotectant transport system and depletions of dissimilatory sulfate reduction (Figure 4*C*, Supplementary Table 17). There were no shared KOs or MOs that were significantly changed in abundance between the JP and ES oral microbiomes (Supplementary Tables 18 and 19).

## Gut and Oral Microbiomes Are Associated With the Risk of PDAC-related Death

A recent study suggests an association between the tumor microbiome and the prognosis of patients with PDAC.[41] To test whether the gut and oral microbial species were
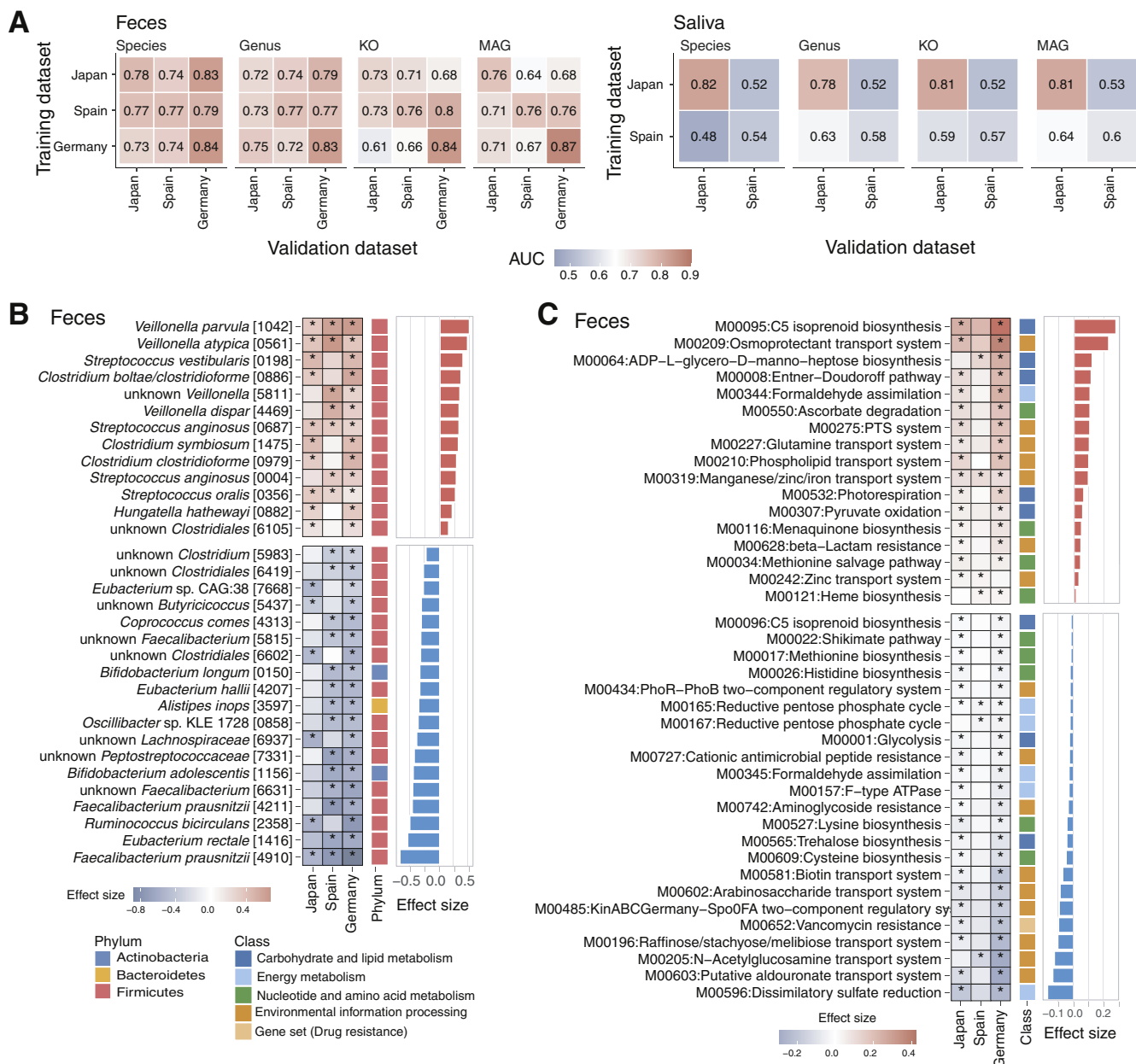
Figure 4. External validation of the prediction models in the JP cohort using the ES and DE cohorts. (*A*) Evaluation of the metagenomic classifiers constructed in each cohort. Y-axes show the dataset used for training, and X-axes represent the dataset used for validation. Values in the cells represent the AUC. The cells in which the training and prediction values are the same (on the diagonal line) indicate AUCs evaluated by internal cross-validation. (*B* and *C*) Species and MOs that were consistently associated with PDAC among the 3 cohorts. Heat map depicts the effect sizes of PDAC for each species and MO obtained from a generalized regression model in each cohort. *Asterisks* represent significant associations between PDAC and the feature (FDR < 0.1). Bar plots show the effect sizes of PDAC for each feature obtained from a generalized regression model using all datasets from the 3 cohorts (cases:controls = 144:317). Results for 32 species (*B*) and 42 MOs (*C*) that were consistently significant in at least 2 of the 3 cohorts are shown. Numbers in parentheses beside the species names indicate mOTU ID.

associated with PDAC-related mortality risk, we conducted a time-to-event analysis with prospective follow-up of the patients in the JP cohort. During a median follow-up of 15.3 months, 68.1% (32 of 47) of the patients with PDAC died (Table 1), and the cumulative death rates were 21% and 40% at 6 months and 1 year, respectively (Figure 5*A*). Using the Cox regression model with LASSO feature selection, we identified the top 10 microbial species strongly associated

with patient prognosis in both the gut and oral microbiomes, respectively (Figure 5*D*, Supplementary Table 20). C-indexes for these models were 0.62 and 0.45 for the gut and salivary microbiomes, respectively. Microbial species whose higher abundances were significantly associated with favorable prognosis included unknown *Alistipes*, *F praus-nitzii*, and Enterobacteriaceae species in the gut, and un-known *Capnocytophaga* and *Capnocytophaga* sp in the saliva
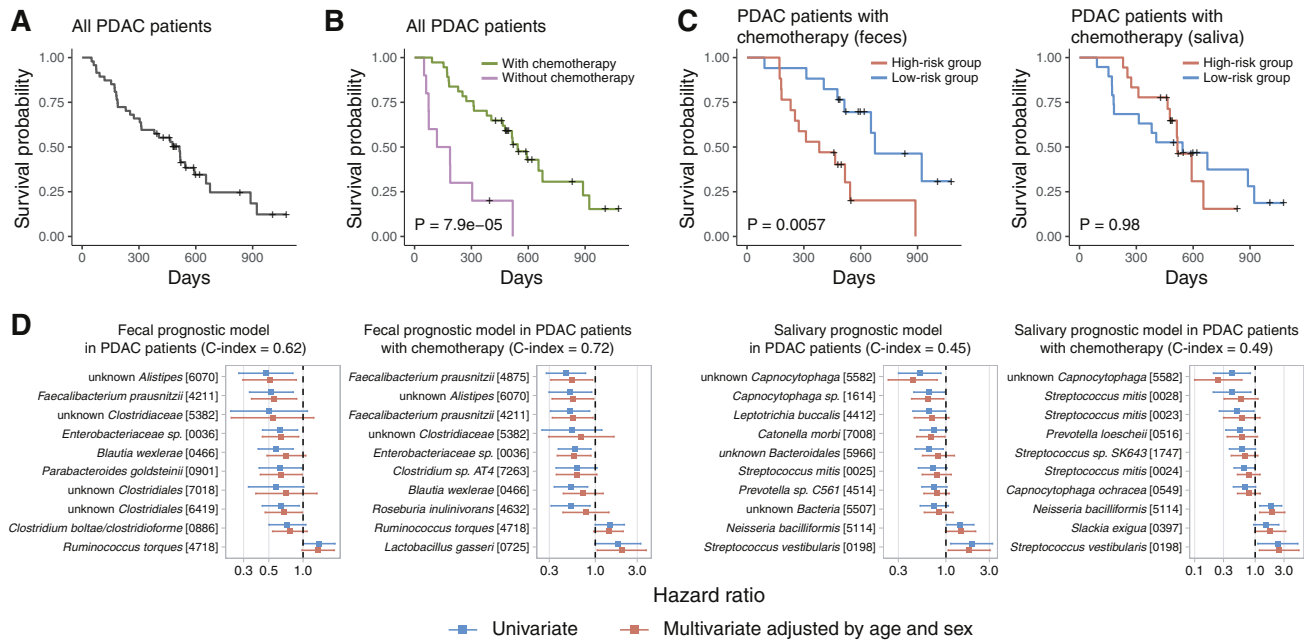
Figure 5. Survival analysis of patients with PDAC in the JP cohort. (*A* and *B*) Overall survival probability of patients with PDAC in the JP cohort (n = 47) (*A*), and that with (n = 37) and without chemotherapy (n = 10) (*B*). (*C*) Survival probability of patients with PDAC in the high- and low-risk group. The patients were divided into high- and low-risk groups based on the median value of the predicted score by the Cox regression model. The survival curves were compared between them by the log-rank test. (*D*) The hazard ratio of the top 10 strongly associated species selected by the Cox regression analysis with LASSO feature selection using all the patients (n = 47) (*B*) and patients with cancer therapy (n = 37) (*D*). Ten times 10-fold cross-validation was performed, and the top 10 species that were strongly associated with the survival probability of the patients were selected. *Blue* and *red colors* of the plot show the hazard ratio determined by univariate and multivariate Cox regression analyses adjusted by age and sex, respectively. The error bar of the plot represents a 95% confidence interval. Numbers in parentheses beside the species names indicate mOTU ID.

PANCREAS

($P < .05$, Supplementary Table 21). These significant associations were unchanged after adjustment with age and sex. In contrast, species associated with poor prognosis were *Ruminococcus torques* in the gut and *S vestibularis* and *Neisseria bacilliformis* in the saliva ($P = .05$, .05, and .02, respectively). In addition, we performed time-to-event analysis for progressive disease as an outcome for the patients with PDAC, and found that higher abundances of several species, such as *Phascolactobacterium* sp in the gut and *Streptococcus pneumoniae* in the saliva, were negatively correlated with the probability of developing progressive disease (Supplementary Figure 11*A*).

Because gut microbes may play a significant role in cancer therapy by modulating the host response to chemotherapeutic drugs,[42] we focused on survival in patients with PDAC who underwent cancer therapy (n = 37, Figure 5*B*). C-indexes from the constructed models for these patients were 0.72 and 0.49, for the gut and oral microbiomes, respectively. When we divided the patients into high- and low-risk groups according to scores obtained from the prediction models, we found a significant difference in survival probability between the 2 groups determined by the gut model ($P = .0057$, Figure 5*C*). Higher abundances of several microbial species in the gut, such as *F prausnitzii*, unknown *Alistipes,* and Enterobacteriaceae species, were associated with better survival; most of these species corresponded to those identified in the analysis including all patients with PDAC (Figure 5*D*, Supplementary Table 21). Although the prediction accuracy of the oral model was lower

than that of the gut model (Figure 5*C*), the model identified several species significantly associated with patient survival, such as unknown *Capnocytophaga* (Figure 5*D*).

Because various factors potentially affect the prognosis of patients with PDAC, we conducted a subgroup analysis: early- and advanced-stage patients, patients treated with adjuvant and nonadjuvant chemotherapy, and patients treated with and without FOLFIRINOX/nab-paclitaxel. Of the 10 microbial species detected in the analysis using all patients with PDAC, 7 gut and oral species were also associated with survival in at least one group ($P < .1$, univariate Cox regression analysis, Supplementary Figure 11*A* and *B*). In particular, *F prausnitzii* and *Blautia wexlerae* in the gut and *N bacilliformis* in the saliva, detected in the 5 and 4 groups, respectively, may be robust markers.

Bacteriophages Infecting Species Associated With PDAC and PDAC-related Mortality

If the PDAC-associated species identified in this study are responsible for the development or progression of PDAC, phage therapy may be a promising approach.[43] We therefore explored bacteriophages in the metagenomic data and found 58 phages that potentially infect microbial species consistently enriched in patients with PDAC among the 3 cohorts (*S oralis, Streptococcus parasanguinis, V atypica, V parvula*) (Supplementary Methods, Supplementary Table 22). None of these phage genomes showed sequence

similarity to those in the RefSeq (average nucleotide identity <95%), suggesting they are novel phages. Cluster analysis based on the proteome profiles of the phages identified 2 major clusters of phages that were predicted to infect *S oralis* and *S parasanguinis* (Figure 6*A*, with black squares). The *Streptococcus* phages in the first cluster (14 genomes included) had a genome size of 55 to 61 kb and were predicted to be lytic phages because they did not encode any integrases and showed no sequence similarity to the host genomes (Figure 6*B*). The *Streptococcus* phages in the second cluster (30 genomes included) had a genome size of 35 to 46 kb and were predicted to be lysogenic phages based on the presence of integrases and sequence similarity to some genomic regions of the host (Figure 6*C*). The phages predicted to infect *V atypica* and *V parvula* showed 2 small clusters as well as individual clusters distinct from each other (Figure 6*D*), in which 1 jumbo lytic phage with a genome size of 201 kb was included (Figure 6*E*). In addition to the phages, we also identified 15 new phages possibly infecting *S vestibularis*, which was associated with poor prognosis for the patients with PDAC (Supplementary Table 23). Collectively, these results demonstrated that there are diverse phages in the gut and oral environments that could infect the PDAC-associated species.

Discussion

In this study, we characterized the gut and oral microbiomes of patients with PDAC by shotgun metagenomic sequencing analysis and demonstrated reproducible gut microbial and functional signatures across 3 different cohorts. Our analysis first revealed dysbiosis of the gut and oral microbiomes of the JP patients with PDAC (Figures 1 and 2) and then validated the results in the external ES and DE cohorts (Figure 3). Particularly, the metagenomic classifiers trained on the JP gut microbiome showed high AUCs, of 0.74 to 0.83, in the ES and DE cohorts, even though the microbiomes differed significantly among the cohorts (Supplementary Figure 6*D*). The large overlaps of the PDAC-associated gut signatures among the cohorts indicate the robustness of the association, which is mostly independent of the lifestyle, geography, and genetic backgrounds of the populations and technical batch effects, as seen in recent studies of colorectal-cancer–associated gut microbes.[17,44] Although our study does not provide evidence of a causal relationship between the changes in the microbiome and PDAC, the results suggest the feasibility of constructing a global, specific, and reproducible predictive model to screen for PDAC based on noninvasive gut microbiome profiling.

Comparison of the JP gut microbiome between patients with PDAC and matched controls revealed a decrease in α-diversity and a difference in β-diversity between them (Figure 1), conditions that were also observed in the ES and DE cohorts (Supplementary Figure 6*A* and *B*), consistent with previous 16S rRNA gene-based studies.[10,11] Among the gut species significantly associated with PDAC in the JP cohort, *Veillonella* spp (*V parvula*, and *V atypica*) and *Streptococcus* spp (*S anginosus* and *S oralis*) were consistently enriched in the guts of patients with PDAC in both the ES and DE cohorts. *Veillonella* and *Streptococcus* spp can potentially interact metabolically and frequently co-occur in gut ecosystems,[45,46] and they have been shown to induce
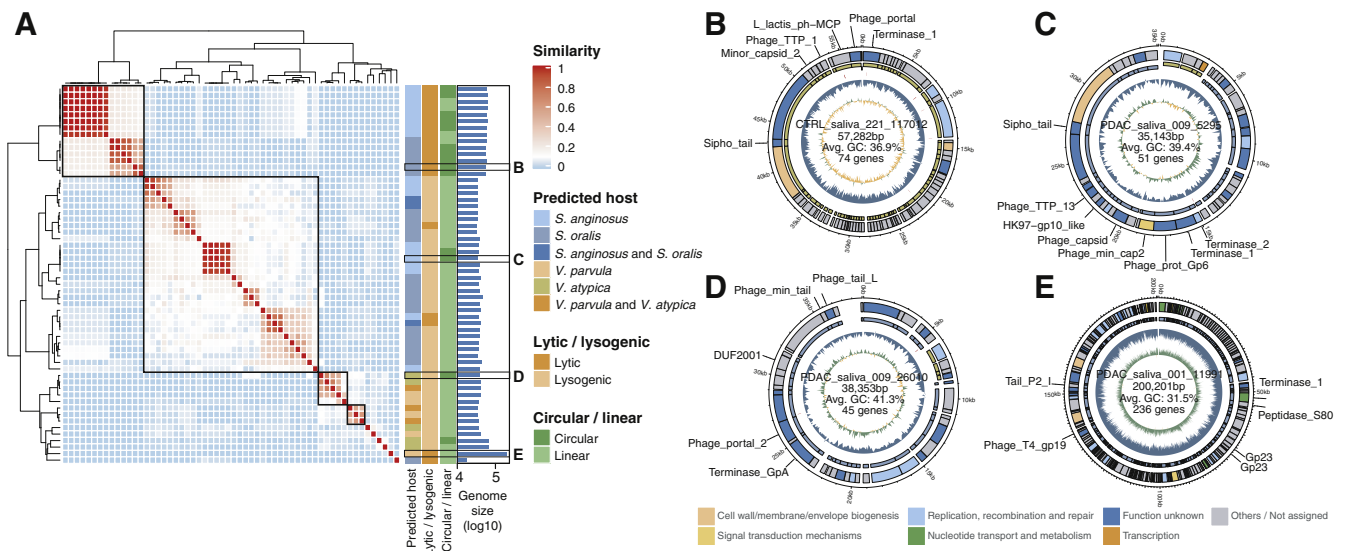
Figure 6. Identification of novel bacteriophages that infect species associated with PDAC and mortality. (*A*) Clustering of the 58 phages that could infect the 4 species (*S anginosus*, *S oralis*, *V parvula*, and *V atypica*) consistently enriched in the PDAC patients across the 3 countries. *Red* and *blue colors* represent high and low similarities of proteome profiles between 2 phages. The color chart on the right shows predicted host, phage lifestyle, contig configuration, and genome size of the phage. The dendrogram on the heatmap was constructed using the ward.D2 method based on 1 - Pearson's correlation as a distance. *Black squares* in the heatmap show phage clusters. Four representative phages in each cluster are shown in panels *B*–*E*. (*B*–*E*) Genome structures of 4 representative phages. From inner to outside, GC skew, GC content, the strand of encoded genes, and functional annotation of the genes are represented. Phage hallmark genes annotated with the Pfam database are shown outside of the circle.

interleukin (IL)-6, IL-8, IL-10, and tumor necrosis factor-α reactions in dendritic cells.[47] This induction is consistent with the increase in these cytokines in patients with PDAC.[48] We also observed the consistent depletion of short-chain fatty acid producers (eg, *F prausnitzii*, *Eubacterium rectale*, and *Ruminoccus bicirculans*) in the guts of patients with PDAC among the cohorts (Figure 1D), which partially overlapped with the signatures of the gut microbiome in various diseases.[34,35] Short-chain fatty acids regulate intestinal immune functions through cell surface G protein-coupled receptors, and their depletion results in inflammation,[49] which might promote the development of PDAC.[50]

Functional analysis of the gut microbiome across the 3 cohorts revealed the significant enrichment of C5 isoprenoid biosynthesis (the mevalonate pathway) in the patients with PDAC of the JP and DE cohorts. This pathway generates isoprenoids, such as farnesyl pyrophosphate and geranylgeranyl pyrophosphate, which are indispensable for activity of GTPases through protein prenylation.[51] One of the key GTPases for initiation and maintenance of PDAC is the human Ras gene family, which is often mutated and upregulated in the tumor tissue of the pancreas.[52] The upregulated metabolic changes in the gut microbiome and cancer tissue may suggest the possibility of bacterial carcinogenesis in the pancreas. Also, we found significant enrichment of microbial genes for biosynthesis of ADP-L-glycero-D-manno-heptose, which generates a precursor of lipopolysaccharide, in the guts of patients with PDAC in the ES and DE cohorts. Lipopolysaccharide induces inflammation through the nuclear factor-kappa-B pathway in the host, which could support the development of pancreatic cancer.[53]

We used public and additional metagenomic datasets to show the specificity of the metagenomic classifier with the enriched species (Figure 2G). However, we identified some overlaps between PDAC- and PPI-associated species, resulting in a somewhat high false positive rate for PPI users (~20%). Given epidemiological evidence of positive associations of PPI use and the risk of PDAC,[54] the overlap of the microbial signatures (eg, increase in *Streptococcus* spp and *Veillonella* spp) implies that the gut microbiome altered by PPIs is responsible for the development or progression of PDAC.

Consistent with 16S rRNA-based studies,[6,7,9,55] we also observed dysbiosis of the oral microbiomes of patients with PDAC in the JP cohort (Figure 1A and B). The significant depletion of *S salivarius*, which has been reported to have anti-inflammatory effects,[56] was the most prominent signature in the PDAC oral microbiome (Figure 1D). The lowered abundance of *Streptococcus* spp in the oral microbiomes of patients with PDAC was concordant with the results of several studies,[6,55] although increases in *Porphyromonas gingivalis*[57] and decreases in *Neisseria* spp[6,7] reported previously were not replicated in the JP cohort. Moreover, no oral species or genes were significantly different between patients with PDAC and controls in the ES cohort, resulting in no overlap between the results in the JP and ES cohorts. Various technical and biological factors, such as storage conditions, DNA extraction, oral hygiene, and diurnal oscillations, could affect the structure of the oral microbiome[58,59] and confound PDAC-oral microbiome associations. Compared with methodologies for fecal samples that have been intensively evaluated and improved,[40] those for saliva samples have been less investigated and could be immature. Standardization of salivary sample collection, storage, and processing will be necessary.

Consistent with past reports,[4,60] fewer than 40% of patients with PDAC survived for 1 year or more after diagnosis in the JP cohort (Figure 4A), emphasizing the need for prognostic biomarkers for the disease. Survival analysis revealed associations of favorable prognosis with higher abundances of short-chain fatty acid producers in the gut, such as *F prausnitzii* and *Clostridiales* species that have been considered to be health-promoting species.[61] In contrast, species relevant to poor prognosis included those previously linked with diseases; for example, *R torques* with inflammatory bowel disease,[62] *Haemophilus parainfluenzae* with respiratory disease, and *N bacilliformis* with endocarditis.[63,64] These signatures had no overlaps with those identified in the tumor microbiome identified in a previous study,[41] suggesting that the gut and oral microbiomes could provide unique signatures to predict prognosis of patients with PDAC.

Furthermore, we successfully identified 58 new phages that could infect the 4 microbial species consistently enriched in the gut of patients with PDAC (*S anginosus*, *S oralis*, *V parvula*, and *V atypica*) (Figure 6). In contrast to antibiotic treatment that has disruptive effects on the entire structure of the microbiome, select for antibiotic resistance genes, or produce multiple side effects, phage therapy may be a promising approach to modulate the microbiome by eliminating only certain species.[43] The genomic information of the phages could be the basis for establishing future phage therapy to treat or prevent PDAC.

One of the limitations of this study is that we could not perform experimental models confirming a causal relationship between the microbiome and the development and progression of PDAC. The PDAC-associated microbial species identified in this study could be targets for further investigating the role of microbial species in PDAC. Another limitation is that the sample size for prognostic analysis of patients with PDAC was small, so a larger sample size is needed for validating the robustness of the prognostic marker. Moreover, the antibiotics users in the European cohorts could not be excluded, which might result in the underestimation of consistent microbial signatures among the cohorts.

In conclusion, our results have shown highly consistent and reproducible associations between the gut microbiome and PDAC across cohorts, which strongly support the presence of global gut microbial signatures for pancreatic cancer. These signatures could be a basis for establishing a robust and accurate screening tool for PDAC in clinical practice and for understanding the roles of the microbiome in the etiology of PDAC.

Supplementary Material

PANCREAS

PANCREAS

References

1. GBD 2017 Pancreatic Cancer Collaborators. The global, regional, and national burden of pancreatic cancer and its attributable risk factors in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol Hepatol 2019;4:934–947.

2. Rawla P, Sunkara T, Gaduputi V. Epidemiology of pancreatic cancer: global trends, etiology and risk factors. World J Oncol 2019;10:10–27.

3. Li D, Xie K, Wolff R, et al. Pancreatic cancer. Lancet 2004;363:1049–1057.

4. McGuigan A, Kelly P, Turkington RC, et al. Pancreatic cancer: a review of clinical diagnosis, epidemiology, treatment and outcomes. World J Gastroenterol 2018; 24:4846–4861.

5. Singhi AD, Koay EJ, Chari ST, et al. Early detection of pancreatic cancer: opportunities and challenges. Gastroenterology 2019;156:2024–2040.

6. Farrell JJ, Zhang L, Zhou H, et al. Variations of oral microbiota are associated with pancreatic diseases including pancreatic cancer. Gut 2012;61:582–588.

7. Torres PJ, Fletcher EM, Gibbons SM, et al. Characterization of the salivary microbiome in patients with pancreatic cancer. PeerJ 2015;3:e1373.

8. Vogtmann E, Han Y, Caporaso JG, et al. Oral microbial community composition is associated with pancreatic cancer: a case-control study in Iran. Cancer Med 2020; 9:797–806.

9. Fan X, Alekseyenko AV, Wu J, et al. Human oral microbiome and prospective risk for pancreatic cancer: a population-based nested case-control study. Gut 2018; 67:120–127.

10. Ren Z, Jiang J, Xie H, et al. Gut microbial profile analysis by MiSeq sequencing of pancreatic carcinoma patients in China. Oncotarget 2017;8:95176–95191.

11. Half E, Keren N, Reshef L, et al. Fecal microbiome signatures of pancreatic cancer patients. Sci Rep 2019;9:16801.

12. Akshintala VS, Talukdar R, Singh VK, et al. The gut microbiome in pancreatic disease. Clin Gastroenterol Hepatol 2019;17:290–295.

13. Thomas RM, Gharaibeh RZ, Gauthier J, et al. Intestinal microbiota enhances pancreatic carcinogenesis in preclinical models. Carcinogenesis 2018;39:1068–1078.

14. Pushalkar S, Hundeyin M, Daley D, et al. The Pancreatic cancer microbiome promotes oncogenesis by induction of innate and adaptive immune suppression. Cancer Discov 2018;8:403–416.

15. Sethi V, Kurtom S, Tarique M, et al. Gut microbiota promotes tumor growth in mice by modulating immune response. Gastroenterology 2018;155:33–37.e6.

16. Thomas RM, Jobin C. Microbiota in pancreatic health and disease: the next frontier in microbiome research. Nat Rev Gastroenterol Hepatol 2020; 17:53–64.

17. Thomas AM, Manghi P, Asnicar F, et al. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. Nat Med 2019;25:667–678.

18. Vujkovic-Cvijin I, Sklar J, Jiang L, et al. Host variables confound gut microbiota studies of human disease. Nature 2020;587:448–454.

19. Tsuda A, Suda W, Morita H, et al. Influence of proton-pump inhibitors on the luminal microbiota in the gastrointestinal tract. Clin Transl Gastroenterol 2015;6:e89.

20. Nagata N, Tohya M, Takeuchi F, et al. Effects of storage temperature, storage time, and Cary-Blair transport medium on the stability of the gut microbiota. Drug Discov Ther 2019;13:256–260.

21. Nagata N, Tohya M, Fukuda S, et al. Effects of bowel preparation on the human gut microbiome and metabolome. Sci Rep 2019;9:4042.

22. Kartal E, Schmidt TSB, Molina-Montes E, et al. A faecal microbiota signature with high specificity for pancreatic cancer. Gut 2022 Mar 8:gutjnl-2021-324755. https://doi.org/10.1136/gutjnl-2021-324755.

23. Kim S-W, Suda W, Kim S, et al. Robustness of gut microbiota of healthy adults in response to probiotic intervention revealed by high-throughput pyrosequencing. DNA Res 2013;20:241–253.

24. Nishijima S, Suda W, Oshima K, et al. The gut microbiome of healthy Japanese and its microbial and functional uniqueness. DNA Res 2016;23:125–133.

25. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–359.

26. Li D, Liu C-M, Luo R, et al. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 2015;31:1674–1676.

27. Hyatt D, Chen G-L, Locascio PF, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119.

28. Steinegger M, Söding J. Clustering huge protein sequence sets in linear time. Nat Commun 2018;9:2542.

29. Huerta-Cepas J, Forslund K, Coelho LP, et al. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol Biol Evol 2017; 34:2115–2122.

30. Milanese A, Mende DR, Paoli L, et al. Microbial abundance, activity and population genomic profiling with mOTUs2. Nat Commun 2019;10:1014.

31. Kawa S, Tokoo M, Oguchi H, et al. Epitope analysis of SPan-1 and DUPAN-2 using synthesized glyco-conjugates sialyllact-N-fucopentaose II and sialyllact-N-tetraose. Pancreas 1994;9:692–697.

32. Kaur S, Baine MJ, Jain M, et al. Early diagnosis of pancreatic cancer: challenges and new developments. Biomark Med 2012;6:597–612.

33. Wirbel J, Zych K, Essex M, et al. Microbiome meta-analysis and cross-disease comparison enabled by the SIAMCAT machine learning toolbox. Genome Biol 2021; 22:93.

34. Zeller G, Tap J, Voigt AY, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol 2014;10:766.

35. Qin J, Li Y, Cai Z, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 2012; 490:55–60.

36. Schmidt TSB, Hayward MR, Coelho LP, et al. Extensive transmission of microbes along the gastrointestinal tract. Elife 2019;8:e42693.

37. Zhang X, Zhang D, Jia H, et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nat Med 2015;21:895–905.

38. Li J, Jia H, Cai X, et al. An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol 2014;32:834–841.

39. Yatsunenko T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. Nature 2012;486:222–227.

40. Costea PI, Zeller G, Sunagawa S, et al. Towards standards for human fecal sample processing in metagenomic studies. Nat Biotechnol 2017;35:1069–1076.

41. Riquelme E, Zhang Y, Zhang L, et al. Tumor microbiome diversity and composition influence pancreatic cancer outcomes. Cell 2019;178:795–806.e12.

42. Alexander JL, Wilson ID, Teare J, et al. Gut microbiota modulation of chemotherapy efficacy and toxicity. Nat Rev Gastroenterol Hepatol 2017;14:356–365.

43. Gordillo Altamirano FL, Barr JJ. Phage therapy in the postantibiotic era. Clin Microbiol Rev 2019; 32:e00066–e00118.

44. Wirbel J, Pyl PT, Kartal E, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nat Med 2019;25:679–689.

45. Zoetendal EG, Raes J, van den Bogert B, et al. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. ISME J 2012; 6:1415–1426.

46. Egland PG, Palmer RJ Jr, Kolenbrander PE. Interspecies communication in Streptococcus gordonii-Veillonella atypica biofilms: signaling in flow conditions requires juxtaposition. Proc Natl Acad Sci U S A 2004; 101:16917–16922.

47. van den Bogert B, Meijerink M, Zoetendal EG, et al. Immunomodulatory properties of Streptococcus and Veillonella isolates from the human small intestine microbiota. PLoS One 2014;9:e114277.

48. Błogowski W, Deskur A, Budkowska M, et al. Selected cytokines in patients with pancreatic cancer: a preliminary report. PLoS One 2014;9:e97613.

49. Parada Venegas D, De la Fuente MK, Landskron G, et al. Short chain fatty acids (SCFAs)-mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. Front Immunol 2019;10:277.

50. Padoan A, Plebani M, Basso D. Inflammation and pancreatic cancer: focus on metabolism, cytokines, and immunity. Int J Mol Sci 2019;20:676.

51. Oni TE, Biffi G, Baker LA, et al. SOAT1 promotes mevalonate pathway dependency in pancreatic cancer. J Exp Med 2020;217:e20192389.

52. Magliano MP di, Logsdon CD. Roles for KRAS in pancreatic tumor development and progression. Gastroenterology 2013;144:1220–1229.

53. Hoesel B, Schmid JA. The complexity of NF-κB signaling in inflammation and cancer. Mol Cancer 2013;12:86.

54. Hong H-E, Kim A-S, Kim M-R, et al. Does the use of proton pump inhibitors increase the risk of pancreatic cancer? A systematic review and meta-analysis of epidemiologic studies. Cancers (Basel) 2020;12:2220.

55. Olson SH, Satagopan J, Xu Y, et al. The oral microbiota in patients with pancreatic cancer, patients with IPMNs, and controls: a pilot study. Cancer Causes Control 2017; 28:959–969.

56. Kaci G, Goudercourt D, Dennin V, et al. Anti-inflammatory properties of Streptococcus salivarius, a commensal bacterium of the oral cavity and digestive tract. Appl Environ Microbiol 2014;80:928–934.

57. Michaud DS, Izard J, Wilhelm-Benartzi CS, et al. Plasma antibodies to oral bacteria and risk of pancreatic cancer in a large European prospective cohort study. Gut 2013; 62:1764–1770.

58. Armstrong AJS, Parmar V, Blaser MJ. Assessing saliva microbiome collection and processing methods. NPJ Biofilms Microbiomes 2021;7:81.

59. Belstrøm D. The salivary microbiota in health and disease. J Oral Microbiol 2020;12:1723975.

60. Luo J, Xiao L, Wu C, et al. The incidence and survival rate of population-based pancreatic cancer patients: Shanghai Cancer Registry 2004-2009. PLoS One 2013;8: e76052.

61. Lopez-Siles M, Duncan SH, Garcia-Gil LJ, et al. Faecalibacterium prausnitzii: from microbiology to diagnostics and prognostics. ISME J 2017;11:841–852.

62. Joossens M, Huys G, Cnockaert M, et al. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. Gut 2011; 60:631–637.

63. Frankard J, Rodriguez-Villalobos H, Struelens MJ, et al. Haemophilus parainfluenzae: an underdiagnosed pathogen of biliary tract infections? Eur J Clin Microbiol Infect Dis 2004;23:46–48.

64. Masliah-Planchon J, Breton G, Jarlier V, et al. Endocarditis due to Neisseria bacilliformis in a patient with a bicuspid aortic valve. J Clin Microbiol 2009; 47:1973–1975.

Correspondence
Address correspondence to Naoyoshi Nagata, MD, PhD, Department of Gastroenterological Endoscopy, Tokyo Medical University, 6-7-1 Nishishinjuku, Shinjuku-ku, Tokyo 160-0023, Japan. e-mail: nnagata_ncgm@yahoo.co.jp, or Suguru Nishijima, PhD, Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. e-mail: nishijima.suguru@gmail.com.

PANCREAS

CRediT Authorship Contributions

Naoyoshi Nagata, MD, PhD (Conceptualization: Equal; Data curation: Equal; Funding acquisition: Lead; Investigation: Equal; Methodology: Equal; Writing – original draft: Equal; Writing – review & editing: Equal).

Suguru Nishijima, PhD (Conceptualization: Equal; Data curation: Equal; Formal analysis: Equal; Investigation: Equal; Methodology: Equal; Visualization: Equal; Writing – original draft: Equal; Writing – review & editing: Equal).

Yasushi Kojima, MD, PhD (Writing – original draft: Equal).

Yuya Hisada, MD (Data curation: Equal; Writing – original draft: Equal).

Koh Imbe, MD, PhD (Data curation: Equal; Writing – original draft: Equal).

Tohru Miyoshi-Akiyama, PhD (Writing – original draft: Equal).

Wataru Suda, PhD (Methodology: Equal; Writing – original draft: Equal).

Moto Kimura, PhD (Writing – original draft: Equal).

Ryo Aoki, PhD (Writing – original draft: Equal).

Katsunori Sekine, MD (Writing – original draft: Equal).

Mitsuru Ohsugi, MD, PhD (Writing – original draft: Equal).

Kuniko Miki, PhD (Writing – original draft: Equal).

Tsuyoshi Osawa, PhD (Writing – original draft: Equal).

Kohjiro Ueki, MD, PhD (Writing – original draft: Equal).

Shinichi Oka, MD, PhD (Writing – original draft: Equal).

Masashi Mizokami, MD, PhD (Writing – original draft: Equal).

Ece Kartal, PhD (Writing – original draft: Equal).

Thomas S. B. Schmidt, PhD (Methodology: Equal; Writing – original draft: Equal).

Esther Molina-Montes, PhD (Writing – original draft: Equal).

Lidia Estudillo, PhD (Writing – original draft: Equal).

Nuria Malats, MD, PhD (Writing – original draft: Equal).

Jonel Trebicka, PhD (Writing – original draft: Equal).

Stephan Kersting, PhD (Writing – original draft: Equal).

Melanie Langheinrich, PhD (Writing – original draft: Equal).

Peer Bork, PhD (Writing – original draft: Equal).

Naomi Uemura, MD, PhD (Writing – original draft: Equal).

Takao Itoi, MD, PhD (Writing – original draft: Equal).

Takashi Kawai, MD, PhD (Writing – original draft: Equal)

PANCREAS