**Supplemental Material**

DNA extraction for tissue samples

All the tissue samples were homogenized using a Fast Prep instrument with lysing matrix Y (MP Biomedicals, CA, USA) after addition of 180µl of ATL buffer (QIAGEN, CA, USA). Bead-beating was performed for one minute at 6.0 m sec-1. The supernatant was then recovered and 200µl of lysozyme (20mg/ml) buffer was added. The samples were then incubated at 37ºC for 30 min. Subsequently, 20-µl of proteinase K (20 mg/ml) was added and samples were further incubated at 56ºC for 30 min. Afterwards, 200-µl lysis buffer AL (QIAGEN) was added and samples were incubated at 56ºC for 30 min. Subsequently, 900-µl of 4M guanidine thiocyanate buffer was added, samples were mixed by inversion. Subsequently, 700-µl cold ethanol was added and the DNA was purified using the DNeasy Blood and Tissue kit (69506, QIAGEN) as per the manufacturer’s tissue protocol (DNeasy Blood & Tissue Handbook, version 07/2006). DNA concentration was fluorometrically measured using the QuantiT PicoGreen dsDNA High-Sensitivity Assay (Q-33120, Life Technologies) with a BioTek fluorescence plate reader instrument (Ex.λ/Em.λ of 485/548 nm) using the BioTek Gen5 software package.

Bacterial amplification

Bacterial amplification of all the DNA extracted from samples was validated using primers targeting the V3-V4 hypervariable region of the 16S ribosomal RNA gene with identical sequences designed for Illumina sequencing (Primers: 341F 5’-CCTACGGGAGGCAGCAG-3’ and 806R 5’GGACTACHVGGGTWTCTAAT-3’) without the addition of Illumina adapters and sample barcode sequencing.

Bacterial mock community

Genomic DNA from ten bacterial strains were used to build a bacterial mock community that was sequenced as a positive control in every MiSeq run performed throughout this research. The ten bacterial strains that comprised this bacterial mock community were as follows: *Cryptobacterium curtum* Oral Taxon 579, Bacteroidales [G-2] sp. Oral Taxon 274, *Capnocytophaga sp.* Oral Taxon 338, *Streptococcus anginosus* Oral Taxon 543, *Peptoniphilus sp.* Oral Taxon 386, *Selenomonas noxia* Oral Taxon 130, *Fusobacterium nucleatum ss polymorphum* Oral Taxon 202, *Aggregatibacter aphrophilus* Oral Taxon 545, and *Pyramidobacter piscolens* Oral Taxon 357. Results for relative abundance of these control samples were consistent, and as expected, across each MiSeq run.

PCR

PCR mixtures of 50-µl contained 10-µl of diluted DNA template, 20-µl of HotMasterMix, 1-µl of each primer (10µM). The cycling conditions consisted of an initial template denaturation of 94ºC for 3 min, followed by 30-cycles of denaturation at 94ºC for 45 sec, annealing at 50ºC for 60 sec, extension at 72ºC for 1.5 min, and a final extension at 72ºC for 10 min. Five-microliters of each PCR product loaded with gel loading dye were run on a 1% agarose gel in 1X Tris-acetate-EDTA (TAE) buffer stained with Sybr Safe DNA and visualized using an Alpha-Innotech instrument equipped with the FluorChem Q imaging software (Ex. λ/Em. λ of 475/537nm).

16S rRNA gene Amplicon Illumina Sequencing

10-50 ng of each metagenomic DNA template was first amplified using the sequencing primers designed to incorporate Illumina adapters and a sample barcode sequence, allowing directional sequencing covering the hypervariable region V3-V4. Primers used were as follows: 341F (AATGATACGGCGACCACCGAGATCTACACTATGGTAATTGT*CCTACGGGAGGCAGCAG*) and 806R (CAAGCAGAAGACGGCATACGAGATN NNNNNNNNNNNAGTCAGTCAGCC*GGACTACHVGGGTWTCTAAT*) (sequences of the primers are in italics, and N sequences corresponding to the barcodes). PCR mixtures contained 10-μl of diluted DNA template, 10-μl of HotMaster Taq DNA Polymerase Mix (5 Prime), and 1μl of each primer mix (10 μM). The cycling conditions consisted of an initial denaturation of 94°C for 3 min, followed by 30 cycles of denaturation at 94°C for 45 sec, annealing at 50 °C for 60 sec, extension at 72°C for 1.5 min, and a final extension at 72°C for 10 min.

PCR products were then purified using a magnetic bead capture kit Agencourt Ampure XP purification beads (Beckman Coulter, Brea, CA, USA). Amplicons from each library were quantified and pooled in equimolar concentrations. Pooled libraries were electrophoresed in a 2% agarose gel with gel loading dye and Sybr Safe DNA gel stain. Bands were visualized under under UV transillumination, the band at ~590 bp was excised and DNA was purified using the Minelute Gel Extraction kit (Qiagen). The purified DNA libraries pool was quantitated on an Agilent bioanalyzer DNA 1000 chips (Agilent, Santa Clara, CA, USA) using a Bioanalyzer to verify the DNA size fragment. The final concentration of the library was determined using a SYBR green quantitative PCR (qPCR) assay with primers specific to the Illumina adapters (Kapa Biosystems, Woburn, MA, USA) using a LightCycler 96 Real-Time PCR System Roche Diagnostics GmbH, Mannheim, Germany). The final amplicon pool was denatured at 4 nM before diluting to a final concentration of 12 pM. The libraries pool was then mixed with >5% PhiX Illumina control and were sequenced by 2 x 250 bp paired-end sequencing on the Miseq platform using MiSeq V2 reagent kit (Illumina, CA, USA), according to the manufacturer’s specifications and generating paired-end reads of 250b in length in each direction.

Statistical analysis

In addition, PERMANOVA tests were conducted to compare beta-diversity measures (i.e., Bray-Curtis) between sites (i.e., pancreatic duct, pancreatic tail, pancreatic head, etc.), and sample groups (i.e., disease versus non-diseased, etc.). Briefly, PERMANOVA is an extension of the traditional analysis of variance (ANOVA) to a square matrix of pairwise distances with significance testing performed by permutation [1].

Zero-inflated beta regression models represent a general class of mixture models where the response variable is assumed to have mixed continuous-discrete distribution with probability mass at zero. For our application, a logistic regression component to model OTU presence/absence (p0) and a beta regression component was used to model non-zero microbial abundance (μ). The rationale for selecting this model stems from two distinct characteristics of microbiome data: the preponderance of zero OTU counts across samples [2], commonly referred to as zero-inflation, and the fact that OTU relative abundance measurements are continuous and bounded between 0 and 1, and as a result, are reasonably well approximated with a beta distribution. Zero-inflated beta regression models were fit using the function “BEINF0”, as implemented in the R package “gamlss”.

Due to sample size limitations, associations between genus-level relative abundances and demographic/clinical variables were identified marginally by fitting zero-inflated beta regression models regressing on a single predictor. Models were fit only to genera with less than 90% of the counts being 0. As such, the total number of genera that were tested thus varied across the considered model. Models failing to converge due to data sparseness were considered not significant and were not carried forward for subsequent analyses. Associations were identified by conducting likelihood-ratio tests (LRT) and considered potentially meaningful when either the LRT p-value was less than 0.05 or the Akaike Information Criterion (AIC) of the alternative model was smaller when compared to the null model.

We conducted statistical analyses focused on identifying genera for which relative abundance differed significantly between RIH cancer patients and NDRI non-cancer patients across the set of pancreatic sites. Models were fit to the OTU data from the following tissue samples: pancreatic duct, pancreatic head, pancreatic tail, pancreatic tumor, pancreatic normal and duodenum. In order to account for within-subject correlation, a random intercept term for subject IDs was incorporated into the zero-inflated beta regression models. The utilized models also adjusted for age, sex, and log library size as fixed effects. Other multi-level categorical predictors such as sequencing run or body site were not included into the models in order to reduce sparseness and improve convergence behavior. Unfortunately, testing fixed effects in generalized linear mixed models via simple LRTs is known to be inefficient and unreliable for small to moderate sample sizes [3]. To address this issue, permutation based tests were utilized. First, the observed likelihood ratio statistic (LRS) comparing the full against the null model, excluding study ID as covariate, was calculated. The null distribution of the LRS was then estimated by permuting study ID labels across patients. P-values were derived from 500 permutations per genus and adjusting for multiple testing was achieved via the false discovery rate method.

We also considered zero-inflated beta regression to compare relative abundance of bacteria by ICD code to evaluate whether profiles differed across the different types of RIH patients. For the purpose of this analysis, ICD10 codes were grouped into three categories: pancreatic cancer (ICD10 codes C25.0-C25.9), periampullary cancer (ICD10 codes C24.0-C24.1), and other pancreatic conditions (ICD10 codes K86.0-K86.3)(Table 1). We hereafter refer to these categories as C25, C24, and K86, respectively. Two strategies were considered: First, 30 NDRI pancreatic-head samples were compared with 30 RIH tumor-samples and adjusted for age, BMI, sex and sequencing run. The same strategy was also applied to compare the effects of ICD codes in RIH samples from NDRI samples using data from duodenum and pancreatic duct tissue. In the second approach, we restricted the analysis to RIH patients to account for other clinical variables (e.g. prior chemotherapy or use of antibiotics in prior 6 months); covariates that were adjusted for were selected empirically by identifying variables that exhibited an association with at least 30% of the OTUs that were formally tested. Bonferroni corrections were made to adjust for multiple comparisons when interpreting relative mean abundances of genera across the ICD codes; p-values from the Wald-tests for the mean value comparisons for the ICD codes were considered meaningful if they were less than 0.00057 (0.05/88; given that a maximum of 88 genera were tested).

Supplemental Table 1. Number of samples per anatomical site from the Rhode Island Hospital [RIH] and the National Disease Research Interchange [NDRI].

|  |  |  |
| --- | --- | --- |
| **Source** | **Anatomical Site** | **N** |
| RIH | Bile Duct Swab | 19 |
| RIH | Pancreatic Stent | 1 |
| RIH | Duodenum | 16 |
| RIH | Ileum Swab | 1 |
| RIH | Jejunum Swab | 29 |
| RIH | Normal Pancreas | 6 |
| RIH | Pancreas Swab | 1 |
| RIH | Pancreas Tumor | 30 |
| RIH | Pancreatic Duct | 20 |
| RIH | Stomach Swab | 6 |
| RIH | Stool | 12 |
| NDRI | Duodenum | 32 |
| NDRI | Pancreas Head | 30 |
| NDRI | Pancreas Tail | 19 |
| NDRI | Pancreatic Duct | 32 |
| Total | | 254 |

**Supplemental Table 2.** Results (at the species level) from multivariable zero-inflated beta regression models comparing bacteria presence/absence

and relative abundance in tissue and swab samples from NDRI and RIH subjects\*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Species** | **Total Read**  **Counts** | **Non-zero**  **Samples** | **Estimated Mean Relative Abundance (μ)\*\*** | | | **Estimated Proportion of Presence (P1)** | | | **Global Perm Test\*** | | |
| **RIH** | **NDRI** | **Wald**  **p-value** | **RIH** | **NDRI** | **Wald**  **p-value** | **p-value** | **p-adjusted^** | **AIC difference** |
| *G.multispecies\_spp670\_3* | 6160 | 27 | 0.0004 | < 0.0001 | < 0.0001 | 0.1238 | 0.0050 | 0.0001 | < 0.0001 | < 0.0001 | 5.32 |
| *L. gasseri* | 526388 | 120 | 0.0090 | 0.0144 | 0.0328 | 0.2598 | 0.8260 | < 0.0001 | < 0.0001 | < 0.0001 | 21.78 |
| *L. salivarius* | 354859 | 111 | 0.0068 | 0.0202 | 0.0000 | 0.2767 | 0.6752 | < 0.0001 | < 0.0001 | < 0.0001 | 13.92 |
| *S. intermedius* | 1295 | 52 | 0.0003 | 0.0003 | 0.2429 | 0.4586 | 0.1190 | 0.0003 | < 0.0001 | < 0.0001 | 1.55 |
| *S. multispecies\_spp573\_2* | 16829 | 44 | 0.0021 | 0.0007 | 0.0170 | 0.1662 | 0.0110 | 0.0000 | < 0.0001 | < 0.0001 | 7.52 |
| *L. saburreum* | 327 | 25 | 0.0004 | 0.0012 | 0.0026 | 0.0725 | 0.0019 | 0.0001 | < 0.0001 | < 0.0001 | 6.24 |
| *A. vaginalis* | 47175 | 26 | 0.0009 | 0.0045 | 0.0154 | 0.0107 | 0.1349 | 0.0008 | < 0.0001 | < 0.0001 | 9.13 |
| *P. micra* | 35977 | 60 | 0.0087 | 0.0047 | 0.0532 | 0.4428 | 0.1040 | < 0.0001 | < 0.0001 | < 0.0001 | 8.04 |
| *F. nucleatum\_subsp.\_vincentii* | 4538 | 29 | 0.0041 | 0.0048 | 0.7534 | 0.1173 | 0.0150 | 0.0012 | < 0.0001 | < 0.0001 | 2.11 |
| *B. wadsworthia* | 102226 | 60 | 0.0070 | 0.0013 | 0.0000 | 0.2000 | 0.0197 | 0.0001 | < 0.0001 | < 0.0001 | 2.95 |
| *A .junii* | 2735 | 51 | 0.0007 | 0.0002 | 0.0000 | 0.3859 | 0.1411 | 0.0010 | < 0.0001 | < 0.0001 | 8.51 |
| *A. rimae* | 1160 | 27 | 0.0012 | 0.0028 | 0.0824 | 0.0525 | 0.0049 | 0.0038 | 0.0020 | 0.0197 | 0.06 |
| *G. multispecies\_spp669\_2* | 43586 | 95 | 0.0143 | 0.0039 | 0.0000 | 0.6426 | 0.3961 | 0.0035 | 0.0020 | 0.0197 | 5.90 |
| *S. gordonii* | 1470 | 42 | 0.0004 | 0.0004 | 0.9423 | 0.4310 | 0.0884 | 0.0001 | 0.0020 | 0.0197 | 2.61 |
| *S. lactarius* | 1482 | 54 | 0.0004 | 0.0003 | 0.0383 | 0.5216 | 0.1164 | 0.0000 | 0.0020 | 0.0197 | 1.40 |
| *C. disporicum* | 261404 | 88 | 0.0074 | 0.0082 | 0.6160 | 0.5326 | 0.2768 | 0.0055 | 0.0020 | 0.0197 | 4.93 |
| *D. pneumosintes* | 11534 | 38 | 0.0012 | 0.0012 | 0.9904 | 0.1523 | 0.0061 | 0.0000 | 0.0020 | 0.0197 | 1.71 |
| *F. multispecies\_spp923\_6* | 1886 | 24 | 0.0001 | < 0.0001 | 0.0027 | 0.0024 | 0.0001 | 0.0000 | 0.0020 | 0.0197 | 1.43 |
| *F. multispecies\_spp930\_3* | 3878 | 29 | 0.0003 | 0.0001 | 0.0083 | 0.0934 | 0.0058 | 0.0002 | 0.0020 | 0.0197 | 2.37 |
| *F. multispecies\_spp933\_3* | 75928 | 63 | 0.0079 | 0.0045 | 0.0532 | 0.5203 | 0.1627 | 0.0001 | 0.0020 | 0.0197 | 2.87 |
| *F .multispecies\_spp935\_4* | 1898 | 26 | 0.0006 | 0.0001 | 0.0018 | 0.0179 | 0.0003 | 0.0002 | 0.0020 | 0.0197 | 2.75 |
| *K. ascorbata\_nov\_87.30%* | 1277 | 25 | 0.0020 | 0.0010 | 0.1912 | 0.0851 | 0.0101 | 0.0055 | 0.0020 | 0.0197 | 3.87 |
| *G. parahaemolysans* | 10971 | 33 | 0.0050 | 0.0059 | 0.5264 | 0.2940 | 0.0456 | 0.0006 | 0.0040 | 0.0310 | 0.51 |
| *L. fermentum* | 39573 | 59 | 0.0035 | 0.0070 | 0.0377 | 0.0169 | 0.2266 | 0.0000 | 0.0040 | 0.0310 | 6.34 |
| *S .multispecies\_spp386\_18* | 184 | 38 | 0.0001 | 0.0001 | 0.0023 | 0.3784 | 0.0716 | 0.0007 | 0.0040 | 0.0310 | 1.16 |
| *S. multispecies\_spp597\_2* | 357 | 37 | 0.0001 | 0.0002 | 0.3053 | 0.3223 | 0.0029 | 0.0000 | 0.0040 | 0.0310 | 3.30 |
| *B. gnavus* | 112467 | 62 | 0.0019 | 0.0024 | 0.3974 | 0.1019 | 0.3278 | 0.0018 | 0.0040 | 0.0310 | 1.43 |
| *A. variabilis* | 970 | 50 | 0.0008 | 0.0004 | 0.0016 | 0.3482 | 0.1502 | 0.0042 | 0.0040 | 0.0310 | 6.60 |
| *L. multispecies\_spp767\_2* | 73192 | 57 | 0.0063 | 0.0062 | 0.9679 | 0.0401 | 0.2282 | 0.0001 | 0.0060 | 0.0434 | 5.44 |
| *S .multispecies\_spp756\_2* | 1332 | 29 | 0.0005 | 0.0009 | 0.0155 | 0.2675 | 0.0595 | 0.0015 | 0.0060 | 0.0434 | 1.92 |

*\*All models are adjusted for age, sex, BMI and log library size. Only bacteria (at species-level) associated with source of samples at p<0.05 after correcting for multiple comparisons are shown. Permutation testing accounts for within subject correlation via random intercept.*

*\*\*Among non-zero samples.*

*^Adjusted for multiple testing*

|  |
| --- |
| Supplemental Table 2 Full OTU |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Bacillales;f\_\_NA;g\_\_Gemella;s\_\_multispecies\_spp670\_3 |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Lactobacillaceae;g\_\_Lactobacillus;s\_\_gasseri |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Lactobacillaceae;g\_\_Lactobacillus;s\_\_salivarius |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Streptococcaceae;g\_\_Streptococcus;s\_\_intermedius |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Streptococcaceae;g\_\_Streptococcus;s\_\_multispecies\_spp573\_2 |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Clostridia;o\_\_Clostridiales;f\_\_Lachnospiraceae\_[XIV];g\_\_Lachnoanaerobaculum;s\_\_saburreum |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Tissierellia;o\_\_Tissierellales;f\_\_Peptoniphilaceae;g\_\_Anaerococcus;s\_\_vaginalis |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Tissierellia;o\_\_Tissierellales;f\_\_Peptoniphilaceae;g\_\_Parvimonas;s\_\_micra |
| k\_\_Bacteria;p\_\_Fusobacteria;c\_\_Fusobacteriia;o\_\_Fusobacteriales;f\_\_Fusobacteriaceae;g\_\_Fusobacterium;s\_\_nucleatum\_subsp.\_vincentii |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Deltaproteobacteria;o\_\_Desulfovibrionales;f\_\_Desulfovibrionaceae;g\_\_Bilophila;s\_\_wadsworthia |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Gammaproteobacteria;o\_\_Pseudomonadales;f\_\_Moraxellaceae;g\_\_Acinetobacter;s\_\_junii |
| k\_\_Bacteria;p\_\_Actinobacteria;c\_\_Coriobacteriia;o\_\_Coriobacteriales;f\_\_Coriobacteriaceae;g\_\_Atopobium;s\_\_rimae |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Bacillales;f\_\_NA;g\_\_Gemella;s\_\_multispecies\_spp669\_2 |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Streptococcaceae;g\_\_Streptococcus;s\_\_gordonii |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Streptococcaceae;g\_\_Streptococcus;s\_\_lactarius |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Clostridia;o\_\_Clostridiales;f\_\_Clostridiaceae;g\_\_Clostridium;s\_\_disporicum |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Clostridia;o\_\_Clostridiales;f\_\_Veillonellaceae;g\_\_Dialister;s\_\_pneumosintes |
| k\_\_Bacteria;p\_\_Fusobacteria;c\_\_Fusobacteriia;o\_\_Fusobacteriales;f\_\_Fusobacteriaceae;g\_\_Fusobacterium;s\_\_multispecies\_spp923\_6 |
| k\_\_Bacteria;p\_\_Fusobacteria;c\_\_Fusobacteriia;o\_\_Fusobacteriales;f\_\_Fusobacteriaceae;g\_\_Fusobacterium;s\_\_multispecies\_spp930\_3 |
| k\_\_Bacteria;p\_\_Fusobacteria;c\_\_Fusobacteriia;o\_\_Fusobacteriales;f\_\_Fusobacteriaceae;g\_\_Fusobacterium;s\_\_multispecies\_spp933\_3 |
| k\_\_Bacteria;p\_\_Fusobacteria;c\_\_Fusobacteriia;o\_\_Fusobacteriales;f\_\_Fusobacteriaceae;g\_\_Fusobacterium;s\_\_multispecies\_spp935\_4 |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Gammaproteobacteria;o\_\_Enterobacteriales;f\_\_Enterobacteriaceae;g\_\_Kluyvera;s\_\_ascorbata\_nov\_87.30% |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Bacillales;f\_\_NA;g\_\_Gemella;s\_\_parahaemolysans |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Lactobacillaceae;g\_\_Lactobacillus;s\_\_fermentum |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Streptococcaceae;g\_\_Streptococcus;s\_\_multispecies\_spp386\_18 |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Streptococcaceae;g\_\_Streptococcus;s\_\_multispecies\_spp597\_2 |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Clostridia;o\_\_Clostridiales;f\_\_Lachnospiraceae;g\_\_Blautia;s\_\_gnavus |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Gammaproteobacteria;o\_\_Pseudomonadales;f\_\_Moraxellaceae;g\_\_Acinetobacter;s\_\_variabilis |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Lactobacillaceae;g\_\_Lactobacillus;s\_\_multispecies\_spp767\_2 |

**Supplemental Table 3.** Results from zero-inflated beta regression models comparing bacteria presence/absence and relative abundance across subject disease ICD codes\*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Genus** | **Total read**  **counts**  **counts** | **Non-zero**  **samples** | **Estimated Mean Relative Abundance**  **(μ)\*\*** | | | | **Estimated Proportion of Presence**  **(P1)** | | | | **p-value** | **AIC difference** | **p-adjusted^** |
| **Control (N=29)** | **C24 (N=7)** | **C25 (N=16)** | **K86.2**  **(N=6)** | **Control (N=29)** | **C24 (N=7)** | **C25 (N=16)** | **K86.2**  **(N=6)** |
| *Porphyromonas* | 7008 | 15 | 0.0006 | 0.0903 | 0.0006 | 0.0062 | 0.3333 | 0.1429 | 0.0588 | 0.5000 | 0.0000 | 28.03 | 0.0000 |
| *Peptoclostridium* | 4137 | 14 | 0.0022 | 0.0010 | 0.0927 | 0.0025 | 0.3667 | 0.0000 | 0.0588 | 0.3333 | 0.0004 | 12.40 | 0.0273 |
| *Acinetobacter* | 7916 | 43 | 0.0238 | 0.0386 | 0.0733 | 0.0583 | 0.5333 | 1.0000 | 0.8824 | 0.8333 | 0.0007 | 11.20 | 0.0454 |
| *Kluyvera* | 2000 | 15 | 0.0015 | 0.0061 | 0.0087 | 0.0082 | 0.0667 | 0.2857 | 0.4706 | 0.5000 | 0.0030 | 7.83 | 0.1841 |
| *Lactobacillus* | 251585 | 40 | 0.1540 | 0.0801 | 0.1371 | 0.1007 | 0.8667 | 0.5714 | 0.2941 | 0.8333 | 0.0033 | 7.54 | 0.2072 |
| *Aggregatibacter* | 591 | 18 | 0.0001 | 0.0002 | 0.0007 | 0.0049 | 0.3333 | 0.5714 | 0.1765 | 0.1667 | 0.0034 | 7.48 | 0.2120 |
| *Ralstonia* | 153 | 15 | 0.0024 | 0.0012 | 0.0280 | 0.0047 | 0.2667 | 0.0000 | 0.2353 | 0.5000 | 0.0038 | 7.23 | 0.2349 |
| *Capnocytophaga* | 316 | 12 | 0.0001 | 0.0002 | 0.0010 | 0.0018 | 0.1667 | 0.2857 | 0.1176 | 0.5000 | 0.0061 | 6.06 | 0.3768 |
| *Enterococcus* | 28254 | 29 | 0.0385 | 0.0429 | 0.0508 | 0.0316 | 0.5333 | 1.0000 | 0.1765 | 0.5000 | 0.0065 | 5.91 | 0.4004 |
| *Clostridium* | 100517 | 29 | 0.0659 | 0.0519 | 0.0475 | 0.0521 | 0.5667 | 0.8571 | 0.1176 | 0.6667 | 0.0099 | 4.85 | 0.6112 |
| *Gemella* | 7769 | 24 | 0.0042 | 0.0107 | 0.0146 | 0.0077 | 0.2667 | 0.2857 | 0.5294 | 0.8333 | 0.0127 | 4.21 | 0.7852 |
| *Prevotella* | 53918 | 36 | 0.0241 | 0.0334 | 0.0972 | 0.0453 | 0.7000 | 0.7143 | 0.3529 | 0.6667 | 0.0225 | 2.72 | 1.0000 |
| *Pseudomonas* | 64128 | 31 | 0.0526 | 0.0684 | 0.1324 | 0.2021 | 0.3667 | 0.8571 | 0.6471 | 0.5000 | 0.0248 | 2.47 | 1.0000 |
| *Raoultella* | 31688 | 9 | 0.0212 | 0.0449 | 0.0400 | 0.0028 | 0.0333 | 0.5714 | 0.2353 | 0.0000 | 0.0255 | 2.40 | 1.0000 |
| *Slackia* | 14667 | 36 | 0.1066 | 0.0491 | 0.0858 | 0.1034 | 0.8000 | 0.4286 | 0.4706 | 0.1667 | 0.0259 | 2.36 | 1.0000 |
| *Selenomonas* | 438 | 13 | 0.0002 | 0.0003 | 0.0035 | 0.0005 | 0.2000 | 0.4286 | 0.1176 | 0.3333 | 0.0298 | 1.98 | 1.0000 |
| *Haemophilus* | 41752 | 34 | 0.0469 | 0.0496 | 0.0570 | 0.0745 | 0.6667 | 0.7143 | 0.2353 | 0.8333 | 0.0521 | 0.48 | 1.0000 |
| *Atopobium* | 5767 | 15 | 0.0015 | 0.0017 | 0.0068 | 0.0097 | 0.3667 | 0.2857 | 0.0588 | 0.1667 | 0.0602 | 0.08 | 1.0000 |
| *Neisseria* | 2253 | 11 | 0.0102 | 0.0066 | 0.0010 | 0.0158 | 0.2000 | 0.2857 | 0.0000 | 0.5000 | 0.0760 | -0.57 | 1.0000 |
| *Bilophila* | 10685 | 13 | 0.0131 | 0.1220 | 0.0316 | 0.0121 | 0.2000 | 0.4286 | 0.1765 | 0.1667 | 0.0807 | -0.74 | 1.0000 |
| *Streptococcus* | 354060 | 59 | 0.1089 | 0.2092 | 0.2472 | 0.1273 | 0.9667 | 1.0000 | 1.0000 | 1.0000 | 0.0837 | -0.85 | 1.0000 |
| *Leptotrichia* | 15419 | 21 | 0.0036 | 0.0058 | 0.0133 | 0.0140 | 0.3667 | 0.5714 | 0.1765 | 0.5000 | 0.0854 | -0.90 | 1.0000 |

*\* Only bacteria (at genus-level) associated with ICD code (overall) at p<0.10 prior to correcting for multiple comparisons are shown. Due to missing BMI on two individuals, numbers are based on 58 tissue samples (for comparability to Table 2 these samples were left out).*

*\*\*Among non-zero samples.*

*^Adjusted for multiple testing.*

|  |
| --- |
| Supplementary Table 2. Full OTU |
| k\_\_Bacteria;p\_\_Bacteroidetes;c\_\_Bacteroides;o\_\_Bacteroidales;f\_\_Porphyromonadaceae;g\_\_Porphyromonas |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Clostridia;o\_\_Clostridiales;f\_\_Peptostreptococcaceae;g\_\_Peptoclostridium |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Gammaproteobacteria;o\_\_Pseudomonadales;f\_\_Moraxellaceae;g\_\_Acinetobacter |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Gammaproteobacteria;o\_\_Enterobacteriales;f\_\_Enterobacteriaceae;g\_\_Kluyvera |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Lactobacillaceae;g\_\_Lactobacillus |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Gammaproteobacteria;o\_\_Pasteurellales;f\_\_Pasteurellaceae;g\_\_Aggregatibacter |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Betaproteobacteria;o\_\_Burkholderiales;f\_\_Burkholderiaceae;g\_\_Ralstonia |
| k\_\_Bacteria;p\_\_Bacteroidetes;c\_\_Flavobacteria;o\_\_Flavobacteriales;f\_\_Flavobacteriaceae;g\_\_Capnocytophaga |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Enterococcaceae;g\_\_Enterococcus |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Clostridia;o\_\_Clostridiales;f\_\_Clostridiaceae;g\_\_Clostridium |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Bacillales;f\_\_Gemellaceae;g\_\_Gemella |
| k\_\_Bacteria;p\_\_Bacteroidetes;c\_\_Bacteroides;o\_\_Bacteroidales;f\_\_Prevotellaceae;g\_\_Prevotella |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Gammaproteobacteria;o\_\_Pseudomonadales;f\_\_Pseudomonadaceae;g\_\_Pseudomonas |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Gammaproteobacteria;o\_\_Enterobacterales;f\_\_Enterobacteriaceae;g\_\_Raoultella |
| k\_\_Bacteria;p\_\_Actinobacteria;c\_\_Coriobacteriia;o\_\_Eggerthellales;f\_\_Eggerthellaceae;g\_\_Slackia |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Clostridia;o\_\_Clostridiales;f\_\_Veillonellaceae;g\_\_Selenomonas |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Gammaproteobacteria;o\_\_Pasteurellales;f\_\_Pasteurellaceae;g\_\_Haemophilus |
| k\_\_Bacteria;p\_\_Actinobacteria;c\_\_Coriobacteriia;o\_\_Coriobacteriales;f\_\_Atopobiaceae;g\_\_Atopobium |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Betaproteobacteria;o\_\_Neisseriales;f\_\_Neisseriaceae;g\_\_Neisseria |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Deltaproteobacteria;o\_\_Desulfovibrionales;f\_\_Desulfovibrionaceae;g\_\_Bilophila |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Streptococcaceae;g\_\_Streptococcus |
| k\_\_Bacteria;p\_\_Fusobacteria;c\_\_Fusobacteriia;o\_\_Fusobacteriales;f\_\_Leptotrichiaceae;g\_\_Leptotrichia |

Supplemental Table 4. Results from multivariable zero-inflated beta regression models comparing bacteria presence/absence and relative abundance across subject disease ICD codes for RIH samples (excluding NDRI samples)\*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Genus** | **Total read**  **counts** | **Non-zero**  **samples** | **Estimated Mean Relative Abundance**  **(μ)\*\*** | | | **Estimated Proportion of Presence**  **(P1)** | | | **p-value** | **AIC difference** | **p-adjusted^** |
| **C24**  **(N=7)** | **C25**  **(N=17)** | **K86.2**  **(N=6)** | **C24**  **(N=7)** | **C25**  **(N=17)** | **K86.2**  **(N=6)** |
| *Porphyromonas* | 6553 | 5 | 0.0901 | 0.0004 | 0.0060 | 0.1429 | 0.0588 | 0.5000 | 0.0001 | 16.83 | 0.0031 |
| *Atopobium* | 984 | 4 | 0.0005 | 0.0052 | 0.0081 | 0.2857 | 0.0588 | 0.1667 | 0.0001 | 15.20 | 0.0066 |
| *Enterococcus* | 2370 | 13 | 0.0247 | 0.0301 | 0.0174 | 1.0000 | 0.1765 | 0.5000 | 0.0016 | 9.41 | 0.0916 |
| *Dialister* | 5111 | 6 | 0.0702 | 0.0109 | 0.0010 | 0.1429 | 0.1765 | 0.3333 | 0.0044 | 7.14 | 0.2523 |
| *clostridium* | 895 | 12 | 0.0098 | 0.0086 | 0.0099 | 0.8571 | 0.1176 | 0.6667 | 0.0053 | 6.72 | 0.3036 |
| *Stomatobaculum* | 961 | 8 | 0.0006 | 0.0074 | 0.0045 | 0.2857 | 0.1765 | 0.5000 | 0.0083 | 5.71 | 0.4717 |
| *Neisseria* | 1982 | 5 | 0.0025 | 0.0004 | 0.0074 | 0.2857 | 0.0000 | 0.5000 | 0.0186 | 3.84 | 1.0000 |
| *Lactobacillus* | 837 | 14 | 0.0018 | 0.0084 | 0.0031 | 0.5714 | 0.2941 | 0.8333 | 0.0190 | 3.79 | 1.0000 |
| *Ralstonia* | 100 | 7 | 0.0022 | 0.0292 | 0.0054 | 0.0000 | 0.2353 | 0.5000 | 0.0204 | 3.62 | 1.0000 |
| *Aggregatibacter* | 569 | 8 | 0.0003 | 0.0008 | 0.0051 | 0.5714 | 0.1765 | 0.1667 | 0.0260 | 3.05 | 1.0000 |
| *Mogibacterium* | 55 | 4 | 0.0001 | 0.0015 | 0.0003 | 0.2857 | 0.0588 | 0.1667 | 0.0269 | 2.97 | 1.0000 |
| *Capnocytophaga* | 310 | 7 | 0.0002 | 0.0011 | 0.0019 | 0.2857 | 0.1176 | 0.5000 | 0.0278 | 2.89 | 1.0000 |
| *Propionibacterium* | 5 | 5 | 0.0001 | 0.0004 | 0.0001 | 0.2857 | 0.0588 | 0.3333 | 0.0289 | 2.80 | 1.0000 |
| *Granulicatella* | 1649 | 14 | 0.0014 | 0.0023 | 0.0023 | 0.7143 | 0.2353 | 0.8333 | 0.0414 | 1.95 | 1.0000 |
| *Megasphaera* | 1765 | 6 | 0.0025 | 0.0203 | 0.0035 | 0.2857 | 0.0588 | 0.5000 | 0.0457 | 1.70 | 1.0000 |

*\*Only bacteria (at genus-level) associated with ICD code (overall) at p<0.10 before correcting for multiple comparisons are shown; given small numbers, these models are marginal models for the ICD codes without other covariates. Only Porphyromonas remained statistically significant after adjusting for previous chemotherapy and presence of stent.*

*\*\*Among non-zero samples.*

*^Adjusted for multiple testing.*

|  |
| --- |
| **Supplementary Table 3**. Full OTU |
| k\_\_Bacteria;p\_\_Bacteroidetes;c\_\_Bacteroides;o\_\_Bacteroidales;f\_\_Porphyromonadaceae;g\_\_Porphyromonas |
| k\_\_Bacteria;p\_\_Actinobacteria;c\_\_Coriobacteriia;o\_\_Coriobacteriales;f\_\_Atopobiaceae;g\_\_Atopobium |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Enterococcaceae;g\_\_Enterococcus |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Clostridia;o\_\_Clostridiales;f\_\_Veillonellaceae;g\_\_Dialister |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Clostridia;o\_\_Clostridiales;f\_\_Clostridiaceae;g\_\_Clostridium |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Clostridia;o\_\_Clostridiales;f\_\_Lachnospiraceae;g\_\_Stomatobaculum |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Betaproteobacteria;o\_\_Neisseriales;f\_\_Neisseriaceae;g\_\_Neisseria |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Lactobacillaceae;g\_\_Lactobacillus |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Betaproteobacteria;o\_\_Burkholderiales;f\_\_Burkholderiaceae;g\_\_Ralstonia |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Gammaproteobacteria;o\_\_Pasteurellales;f\_\_Pasteurellaceae;g\_\_Aggregatibacter |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Clostridia;o\_\_Clostridiales;f\_\_Peptostreptococcaceae\_[XI];g\_\_Mogibacterium |
| k\_\_Bacteria;p\_\_Bacteroidetes;c\_\_Flavobacteria;o\_\_Flavobacteriales;f\_\_Flavobacteriaceae;g\_\_Capnocytophaga |
| k\_\_Bacteria;p\_\_Actinobacteria;c\_\_Actinobacteria;o\_\_Actinomycetales;f\_\_Propionibacteriaceae;g\_\_Propionibacterium |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Lactobacillales;f\_\_Carnobacteriaceae;g\_\_Granulicatella |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Clostridia;o\_\_Clostridiales;f\_\_Veillonellaceae;g\_\_Megasphaera |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Gammaproteobacteria;o\_\_Pasteurellales;f\_\_Pasteurellaceae;g\_\_Haemophilus |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Clostridia;o\_\_Clostridiales;f\_\_Veillonellaceae;g\_\_Selenomonas |
| k\_\_Bacteria;p\_\_Actinobacteria;c\_\_Coriobacteriia;o\_\_Eggerthellales;f\_\_Eggerthellaceae;g\_\_Slackia |
| k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli;o\_\_Bacillales;f\_\_Staphylococcaceae;g\_\_Staphylococcus |
| k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Gammaproteobacteria;o\_\_Enterobacteriales;f\_\_Enterobacteriaceae;g\_\_Enterobacter |

**References for supplemental material**

1. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecology 2001;26(1):32-46.

2. Chen EZ, Li H. A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. Bioinformatics 2016;32(17):2611-7.

3. Bolker BM, Brooks ME, Clark CJ*, et al.* Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol 2009;24(3):127-35.